

Stochastik

Inhalt

Vorbemerkung	3
I. Maßtheoretische Grundlagen	4
I. 1. Mengen, Strukturen, Abbildungen	5
I. 2. σ -Algebren	8
I. 3. Maß und Wahrscheinlichkeit	17
I. 4. Produktmaße	27
I. 5. Messbare Abbildungen	31
I. 6. Borel'sche σ -Algebra und Lebesgue'sches Maß	40
I. 7. Das Lebesgue-Integral	49
II. Grundprinzipien stochastischer Modellierung	66
II.1. Elemente der Kombinatorik	66
II.2. Zufallsvariablen und ihre Verteilung, stochastische Unabhängigkeit.....	70
II.3. Verteilungsfunktionen und Momente	83
II.4. Erzeugende Funktionen	118
II.5. Grundlegende Verteilungen	122
II.6. Das Gesetz der Großen Zahlen	134
II.7. Der Zentrale Grenzwertsatz	142
II.8. Bedingte Verteilungen	152
II.9. Die mehrdimensionale Normalverteilung	166
III. Grundprinzipien der Statistik	171
III.1. Elementare Schätzverfahren	174
III.2. Elementare Testverfahren	186
III.3. Lineare und nicht-lineare Regression	193
III.4. Lage-Skalen-Familien, Q-Q-Plots	198
IV. Stochastische Prozesse	201
IV.1. Markoff-Ketten	201
IV.2. Markoff-Prozesse	211
Verzeichnis der Definitionen, Sätze und Lemmata	225
Literatur	227
Danksagung	228

Vorbemerkung

Der Beginn der rigorosen Entwicklung desjenigen Teils der Mathematik, den man heute *Stochastik* nennt (Wort griechischen Ursprungs, sinngemäß etwa: "Kunst des geschickten Vermutens"), wird allgemein mit dem berühmten Briefwechsel zwischen Blaise Pascal (1623 – 1662) und Pierre de Fermat (1601 – 1665) aus dem Jahre 1654 in Verbindung gebracht, dem folgendes Glückspielproblem des französischen Adligen George Brossin Antoine Gombaud, Chevalier de Méré (1607 – 1684) zugrunde liegt:

Warum ist es unvoreilhaft, zum Erreichen einer Doppelsechs mit zwei Würfeln 24 Würfe zu tun, als zum Erreichen einer Sechs mit einem Würfel 4 Würfe zu tun, obwohl das Verhältnis 24 zu 36 (was die Anzahl der Ergebnisse bei 2 Würfeln ist) dasselbe ist wie 4 zu 6 (was die Anzahl der Ergebnisse bei einem Würfel ist)?

Neben diesem Problem war es aber vor allem das sogenannte *Teilungsproblem* des Luca Pacioli (1445 – 1517), das sich auf die "gerechte" Aufteilung des Spieleinsatzes bei vorzeitiger Beendigung eines Spiels bezog, welches Pascal und Fermat in ihrem Schriftwechsel und insbesondere Pascal selbst in seinem "Traité du triangle arithmétique" ausführlicher erörterten. Auch Christiaan Huygens (1629 – 1695) hörte zwei Jahre später in Paris von diesem Briefwechsel, da er aber keinen Zugang zu den dort behandelten Methoden hatte, begann er unabhängig von Pascal und Fermat mit Arbeiten an dem fundamentalen Begriff des "Erwartungswerts", aus dem der auf niederländisch verfasste Text "Tractaet handelende van Reeckening in Speelen van Geluck" hervorging. Sein akademischer Lehrer Frans van Schooten (1615 – 1660) besorgte eine lateinische Übersetzung und fügte diese im Jahre 1657 als "Tractatus de Ratiociniis in Aleae Ludo" seinem eigenen Werk "Exercitationum Mathematicarum Libri Quinque" als Anhang an. Der Huygens'sche Text war von herausragender Bedeutung für die nachfolgende Entwicklung der Wahrscheinlichkeitsrechnung vor allem durch Jakob Bernoulli (1654 – 1705); so ist dessen erster Teil der "Ars conjectandi" (1713 posthum veröffentlicht) im wesentlichen eine kommentierte und erweiterte Wiederauflage der Huygens'schen Abhandlung. Bernoulli hat wohl als erster ein auf mathematischen Prinzipien beruhendes, schwaches "Gesetz der Großen Zahlen" formuliert (und damit die Stochastik vom Ruch des Glückspiels befreit), in seiner frühen Fassung als eine Art Grenzwertaussage für die Annäherung relativer Häufigkeiten bei mehrfacher Versuchswiederholung an die zugrundeliegende "Wahrscheinlichkeit" des interessierenden Ereignisses. Ein korrekter Beweis des heute als "Starkes Gesetz der Großen Zahlen" bekannten Satzes gelang allerdings nach einem fehlerhaften Versuch durch Emile Borel (1871 – 1956) erst Felix Hausdorff (1868 – 1942) in seinem 1914 erschienenen Buch "Grundzüge der Mengenlehre". Eine rigide Axiomatisierung der Wahrscheinlichkeitsrechnung verdanken wir aber vor allem den Bemühungen von David Hilbert (1862 – 1943) und seinen Schülern ab etwa 1900; das heute als allgemein verbindlich akzeptierte *maßtheoretische Fundament der Stochastik*, das insbesondere auch auf der abstrakten Integrationstheorie von Henri-Léon Lebesgue (1875 – 1941) beruht, geht auf diese Bemühungen zurück und gipfelt in der durch Andrej Nikolajewitsch Kolmogoroff (1903 – 1987) im Jahre 1933 in sehr allgemeinem Rahmen entwickelten, heute ihm zu Ehren so benannten Axiomatik. Eine sehr ausführliche Darstellung der geschichtlichen Entwicklung der Stochastik findet man z.B. in der Monographie von SCHNEIDER (1988), vgl. auch BARTH UND HALLER (1998) oder KRENGEL (2003). In ELSTRODT (1996) findet man darüber hinaus viele historische Zitate und Anmerkungen zur Entwicklung der abstrakten Maß- und Integrationstheorie.

I. Maßtheoretische Grundlagen

Moderne Anwendungen der Stochastik kommen heutzutage in nahezu allen wissenschaftlichen Disziplinen vor. In der Physik spielt der Wahrscheinlichkeitsbegriff spätestens seit Albert Einsteins (1879 – 1955) fundamentaler Arbeit über Diffusionsprozesse aus dem Jahre 1905 eine zentrale Rolle, ebenso in der später entwickelten Quantentheorie und relativistischen Mechanik. Louis Bachelier (1870 – 1946) legte bereits im Jahr 1900 mit seiner Arbeit "Théorie de la spéculation", die auf Diffusionsprozesse explizit Bezug nimmt, das Fundament für die heutige moderne Finanzmathematik, die sich als jüngster und berühmtester Anwendungszweig der Stochastik vor allem seit den bahnbrechenden Arbeiten der späteren Nobelpreisträger Fischer Black, Robert Merton und Myron Scholes ab 1972 stürmisch entwickelt hat. Bereits seit 1669 spielt die Stochastik in der Versicherungsmathematik, befördert vor allem durch den schon genannten Christiaan Huygens, eine grundlegende Rolle. 1725 verfasst Abraham de Moivre (1667 – 1754) das erste Lehrbuch der Versicherungsmathematik mit dem Titel *Annuities upon Lives*. Auch Gotfried Wilhelm Leibniz (1646 – 1716) hat sich mit der Versicherungsmathematik befasst; ihm zu Ehren veranstaltet die Hamburger Feuerkasse, die 1667 als erstes öffentlich-rechtliches Versicherungsunternehmen der Welt gegründet wurde, jährlich das sogenannte *Leibniz-Forum*.

Weitere aktuelle Anwendungsgebiete sind u.a. die Signalverarbeitung (auch im Audio- und Videobereich oder bei der Rekonstruktion gestörter Bilder), die Geostatistik (Bergbau, Lagerstättenkunde) und die Stochastische Geometrie (mit Anwendungen in Biologie und Ökologie), Stochastische Modellierung in der Informatik sowie der große Bereich der Mathematischen Statistik mit Anwendungen u.a. in der Ökonomie und der Medizin.

I. Maßtheoretische Grundlagen

Leider herrscht bei vielen "reinen" Mathematikern, aber auch Pädagogen, die Mathematik an der Schule unterrichten, auch heute noch der Eindruck vor, als ob die wesentlichen Gegenstände der Stochastik Spielkarten, Lottokugeln oder Würfel seien. Die vorigen Ausführungen belegen jedoch, dass die Stochastik seit weit über 100 Jahren ein sehr ernstzunehmendes Teilgebiet der abstrakten Mathematik ist, welches heute wesentlich auf der Analysis, insbesondere der Funktionalanalysis (etwa im Bereich der Stochastischen Prozesse), der Topologie und der abstrakten Maß- und Integrationstheorie aufbaut. Dies zeigt auch die lange Reihe berühmter Mathematiker, die sich im Laufe ihres Lebens auf die eine oder andere Weise mit der Stochastik beschäftigt haben. Das in der modernen Finanzmathematik benötigte *Stochastische Integral*, bei dem Funktionen als Integranden auftreten, die in keinem endlichen Intervall von beschränkter Variation bzw. in jedem Punkt stetig, aber in keinem Punkt differenzierbar sind, wurde sogar erst in der Mitte des letzten Jahrhunderts durch den Japaner Kiyosi Itô (geb. 1915) neu entwickelt. Zumindest rudimentäre maßtheoretische Kenntnisse sind daher zum Verständnis stochastischer Theoriebildung, aber auch für eine korrekte Modellierung "konkreter" Anwendungsfälle unerlässlich. Aus diesem Grund ist das erste Kapitel dieses Textes der Erarbeitung einiger dieser Grundlagen gewidmet. Recht ausführliche und anspruchsvolle Darstellungen findet man in BAUER (1992), ELSTRODT (1996), KLENKE (2008) oder SCHÜRGER (1998).

Ein Großteil dieses Skriptes orientiert sich an der Monographie *Stochastik für Informatiker* (MATHAR UND PFEIFER (1990)) sowie deren Vorläufern, die während unserer Zeit als Wissenschaftliche Assistenten an der RWTH Aachen entstanden sind. Eine Monographie, die inhaltlich unserem Verständnis der maßtheoretischen Grundlagen der Stochastik weitgehend entspricht, ist LEADBETTER, CAMBANIS UND PIPIRAS (2014).

I. 1. Mengen, Strukturen, Abbildungen

Ein wesentlicher Teil der mathematischen Stochastik wird mit Hilfe mengentheoretischer Begriffe formuliert. Deshalb soll zunächst die zugrunde liegende Terminologie geklärt werden. Viele der hier angesprochenen Themen finden sich in größerer Tiefe auch in DEISER (2010).

Ausgangspunkt der meisten Betrachtungen ist eine nicht-leere Menge Ω , die die möglichen Ergebnisse zufälliger Vorgänge beschreibt, sowie geeignete Systeme von Teilmengen hiervon. Insbesondere interessieren uns folgende Begriffe bzw. Operationen mit (Teil-)Mengen:

\in, \notin :	als Element (nicht) enthalten
\wedge, \vee, \neg :	logisches "und" / "oder" / "nicht"
\exists, \forall :	es existiert / für alle
$\subset, \subseteq, \supset, \supseteq$:	(echte) Teilmenge / Obermenge von
$\mathfrak{P}(\Omega) = \{A \mid A \subseteq \Omega\}$:	Potenzmenge von Ω
$A^c = \{\omega \in \Omega \mid \omega \notin A\}$:	Komplement der Menge A
$A \setminus B = \{\omega \in \Omega \mid \omega \in A \wedge \omega \notin B\}$:	Differenz der Mengen A und B
$A \cap B = \{\omega \in \Omega \mid \omega \in A \wedge \omega \in B\}$:	Durchschnitt der Mengen A und B
$A \cup B = \{\omega \in \Omega \mid \omega \in A \vee \omega \in B\}$:	Vereinigung der Mengen A und B
$A \oplus B = \{\omega \in \Omega \mid \omega \in A \vee \omega \in B\}, A \cap B = \emptyset$:	Vereinigung der <i>disjunkten</i> Mengen A und B
$A \times B = \{(\omega_1, \omega_2) \mid \omega_1 \in A \wedge \omega_2 \in B\}$:	kartesisches Produkt der Mengen A und B

Häufig werden auch Verknüpfungen mehrerer, ggf. sogar unendlich vieler Mengen $\{A_i\}_{i \in I} \subseteq \mathfrak{P}(\Omega)$ mit geeigneten Indexmengen I betrachtet. Dazu gehören:

$\bigcup_{i \in I} A_i = \{\omega \in \Omega \mid \exists i \in I: \omega \in A_i\}$:	Vereinigung der Mengen $\{A_i\}_{i \in I}$
$\bigoplus_{i \in I} A_i = \bigcup_{i \in I} A_i$:	Vereinigung der <i>paarweise disjunkten</i> Mengen $\{A_i\}_{i \in I}$
$\bigcap_{i \in I} A_i = \{\omega \in \Omega \mid \forall i \in I: \omega \in A_i\}$:	Durchschnitt der Mengen $\{A_i\}_{i \in I}$
$\prod_{i \in I} A_i = \left\{ f: I \rightarrow \bigcup_{i \in I} A_i \mid \forall i \in I: f(i) \in A_i \right\}$:	kartesisches Produkt der Mengen $\{A_i\}_{i \in I}$

Dabei haben wir von der Schreibweise $f: I \rightarrow M$ für Abbildungen f mit Definitionsbereich I und Werten in M Gebrauch gemacht.

I. Maßtheoretische Grundlagen

Für *abzählbare* Indexmengen wie etwa $I = \mathbb{N}$ schreibt man alternativ auch $\bigcup_{i=1}^{\infty} A_i$ für die Vereinigung, $\bigcap_{i=1}^{\infty} A_i$ für den Durchschnitt und $\prod_{i=1}^{\infty} A_i = \left\{ \{\omega_i\}_{i \in \mathbb{N}} \mid \forall i \in I : \omega_i \in A_i \right\}$ ¹ für das kartesische Produkt u.s.w.; entsprechend bei *endlichen* Indexmengen.

Wichtig im Zusammenhang mit dem kartesischen Produkt über beliebige Indexmengen ist das nachfolgende

Auswahlaxiom: Es seien sämtliche Mengen $\{A_i\}_{i \in I}$ sowie die Indexmenge I selbst nicht-leer.

Dann ist auch das kartesische Produkt $\prod_{i \in I} A_i = \left\{ f : I \rightarrow \bigcup_{i \in I} A_i \mid \forall i \in I : f(i) \in A_i \right\}$ nicht-leer.

Das Auswahlaxiom ist unabhängig von den klassischen Axiomen der Zermelo-Fraenkel'schen Mengenlehre und spielt für viele Existenzsätze in der Mathematik eine fundamentale Rolle. Wir werden deshalb wie meist üblich die Gültigkeit dieses Axioms voraussetzen. (Eine anschauliche Interpretation dieses Axioms besagt, dass man aus allen Mengen $\{A_i\}_{i \in I}$ "gleichzeitig" ein Element $\omega_i = f(i) \in A_i$ auswählen kann.)

Die folgenden, leicht zu beweisenden Beziehungen sind neben anderen von elementarer Wichtigkeit; wir geben sie daher hier ohne Beweis an:

Lemma 1. Es gilt für beliebige, nicht-leere Indexmengen I, J :

$$\begin{aligned} \left(\bigcup_{i \in I} A_i \right)^c &= \bigcap_{i \in I} A_i^c \\ &\text{(Regeln von de Morgan)} \\ \left(\bigcap_{i \in I} A_i \right)^c &= \bigcup_{i \in I} A_i^c \\ \bigcap_{j \in J} \prod_{i \in I} A_{ij} &= \prod_{i \in I} \bigcap_{j \in J} A_{ij} \\ \left(\bigcup_{i \in I} A_i \right) \cap \left(\bigcup_{j \in J} B_j \right) &= \bigcup_{i \in I} \bigcup_{j \in J} (A_i \cap B_j) \\ \left(\bigcup_{i \in I} A_i \right) \times \left(\bigcup_{j \in J} B_j \right) &= \bigcup_{i \in I} \bigcup_{j \in J} (A_i \times B_j). \end{aligned}$$

Gelegentlich werden wir auch von dem Begriff der *Äquivalenzrelation* Gebrauch machen, der folgendermaßen definiert ist.

¹ D.h. bei abzählbaren Indexmengen orientiert man sich hier an der üblichen Folgen- bzw. Tupel-Notation.

I. Maßtheoretische Grundlagen

Definition 1. Eine Teilmenge $R \subseteq \Omega \times \Omega$ heißt *Äquivalenzrelation* über Ω , wenn gilt:

$$\forall x \in \Omega: (x, x) \in R \quad (\text{Reflexivität})$$

$$\forall x, y \in \Omega: (x, y) \in R \Rightarrow (y, x) \in R \quad (\text{Symmetrie})$$

$$\forall x, y, z \in \Omega: (x, y) \in R \wedge (y, z) \in R \Rightarrow (x, z) \in R \quad (\text{Transitivität})$$

Bemerkung: Anstatt $(x, y) \in R$ oder $x R y$ schreibt man meist $x \sim_R y$ oder einfacher nur $x \sim y$, wenn die Menge R bekannt ist oder es keine Verwechslungsmöglichkeiten gibt.

Eine Äquivalenzrelation über Ω führt in natürlicher Weise zu einer *Klasseneinteilung*, indem man die zu jedem $x \in \Omega$ gehörige Äquivalenzklasse

$$R_x := \{y \in \Omega \mid y \sim_R x\}$$

betrachtet. Es gilt nämlich:

Lemma 2. Ist R eine Äquivalenzrelation über der Menge Ω , so gilt:

$$\forall x, y \in \Omega: R_x \cap R_y \neq \emptyset \Rightarrow R_x = R_y,$$

d.h. alle Äquivalenzklassen R_x, R_y sind entweder disjunkt oder fallen zusammen.

Beweis: Sei $R_x \cap R_y \neq \emptyset$ für $x, y \in \Omega$. Dann existiert ein $z \in R_x \cap R_y$, d.h. es gilt $z \sim_R x$ und $z \sim_R y$, also wegen der Transitivität auch $x \sim_R y$. Für jedes Element $w \in R_x$ gilt folglich $w \sim_R x \sim_R y$, also $w \in R_y$ und somit $R_x \subseteq R_y$. Auf dieselbe Weise zeigt man $R_y \subseteq R_x$, d.h. es gilt $R_x = R_y$, wie behauptet. ■

Wegen $x \in R_x$ (Reflexivität) liegt nun jedes Element $x \in \Omega$ in genau einer Äquivalenzklasse. Damit existiert nach Lemma 2 eine eindeutig bestimmte disjunkte Zerlegung

$$\Omega = \bigoplus_{i \in I} \Omega_i$$

in die paarweise verschiedenen Äquivalenzklassen $\Omega_i, i \in I$ mit einer geeigneten, nicht-leeren Indexmenge I .

Ist umgekehrt eine solche disjunkte Zerlegung vorgegeben, so induziert diese eine Äquivalenzrelation R über Ω , indem man

$$x \sim_R y :\Leftrightarrow \exists i \in I: x \in \Omega_i \wedge y \in \Omega_i$$

setzt.

I. Maßtheoretische Grundlagen

Man ist manchmal auch an einem sog. *Repräsentantensystem* einer Äquivalenzrelation interessiert, d.h. an einer Menge M , die aus jeder Äquivalenzklasse genau ein Element enthält. Man beachte aber dabei, dass für eine solche Auswahl i.Allg. das *Auswahlaxiom* herangezogen werden muss, weil man hierfür eine Abbildung $f: I \rightarrow \Omega$ benötigt mit der Eigenschaft $f(i) \in \Omega_i, i \in I$. Solche Repräsentantensysteme werden später bei der Modellbildung für geeignete "Ereignissysteme" eine wesentliche Rolle spielen.

Definition 2. Zwei Mengen A und B heißen *gleichmächtig*, wenn es eine bijektive Abbildung $f: A \rightarrow B$ gibt. Ist eine Menge A gleichmächtig zu der (endlichen) Menge $B = \{1, \dots, n\}$ mit $n \in \mathbb{N}$, so bezeichnen wir mit $\#(A) = n$ die Anzahl der Elemente von A . Ist A zu keiner endlichen Menge gleichmächtig, so wählen wir entsprechend als Anzahl der Elemente $\#(A) = \infty$.

Man kann mit Hilfe des Mächtigkeitsbegriffs leicht eine Äquivalenzrelation $\sim_{\#}$ auf der Menge $\mathfrak{P}(\Omega) \times \mathfrak{P}(\Omega)$ definieren, indem man setzt:

$$\forall A, B \in \mathfrak{P}(\Omega): A \sim_{\#} B \Leftrightarrow A \text{ ist gleichmächtig zu } B.$$

Gilt $A \sim_{\#} I \subseteq \mathbb{N}$, so nennen wir die Menge A *abzählbar*, anderenfalls *überabzählbar*.

Ähnlich wie das Auswahlaxiom spielt auch das folgende Axiom allgemein eine wichtige Rolle in der Mathematik.

Kontinuumshypothese: Für jede Menge M mit $\mathbb{N} \subseteq M \subseteq \mathbb{R}$ gilt: $M \sim_{\#} \mathbb{N}$ oder $M \sim_{\#} \mathbb{R}$, d.h. es gibt bezüglich der Mächtigkeit keine Menge "zwischen" \mathbb{N} und \mathbb{R} .

Auch dieses Axiom ist unabhängig von den klassischen Axiomen der Zermelo-Fraenkel'schen Mengenlehre; seine Gültigkeit wird wie beim Auswahlaxiom in der Regel vorausgesetzt.

I. 2. σ -Algebren

Zur Formalisierung des Begriffes eines (zufälligen) "Ereignisses" verwenden wir – wie schon erwähnt – die Sprache der Mengenlehre. Wir gehen dabei davon aus, dass sich alle relevanten Ereignisse als Teilmengen einer festen, nicht-leeren Menge Ω beschreiben lassen. Dass hierfür leider nicht immer die gesamte Potenzmenge $\mathfrak{P}(\Omega)$ geeignet ist, werden wir später sehen.

Definition 3 (σ -Algebra). Ein Mengensystem $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ heißt σ -Algebra (über Ω), wenn die folgenden drei definierenden Eigenschaften erfüllt sind:

$$\begin{aligned} \Omega &\in \mathcal{A} && \text{(Normiertheit)} \\ A \in \mathcal{A} &\Rightarrow A^c \in \mathcal{A} && \text{(Komplementstabilität)} \\ \{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A} &\Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A} && \text{(abzählbare Vereinigungsstabilität).} \end{aligned}$$

Das Paar (Ω, \mathcal{A}) heißt *Messraum*; die Elemente von \mathcal{A} werden oft auch als (\mathcal{A}) -*messbare* Mengen bezeichnet.

I. Maßtheoretische Grundlagen

Aufgrund der de Morgan'schen Regeln (Lemma 1) können alternativ auch die folgenden, äquivalenten Eigenschaften zur Definition einer σ -Algebra herangezogen werden:

$$\emptyset \in \mathcal{A} \quad (\text{für die Normiertheit})$$

oder auch

$$\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A} \Rightarrow \bigcap_{n=1}^{\infty} A_n \in \mathcal{A} \quad (\text{abzählbare Durchschnittsstabilität; für die Vereinigungsstabilität}).$$

Eine weitere, nicht-triviale Charakterisierung einer σ -Algebra gibt

Satz 1. Ein Mengensystem $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ ist genau dann eine σ -Algebra über Ω , wenn die folgenden vier (definierenden) Eigenschaften erfüllt sind:

$$\begin{aligned} \Omega &\in \mathcal{A} \\ A \in \mathcal{A} &\Rightarrow A^c \in \mathcal{A} \\ A, B \in \mathcal{A} &\Rightarrow A \cap B \in \mathcal{A} \\ \{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A} \text{ paarweise disjunkt} &\Rightarrow \bigoplus_{n=1}^{\infty} A_n \in \mathcal{A}. \end{aligned}$$

Beweis: Es reicht, zu zeigen, dass die letzten beiden Eigenschaften zur abzählbaren Vereinigungsstabilität äquivalent sind.

Sei also $\mathcal{A} \subseteq \mathfrak{P}(\Omega)$ eine σ -Algebra über Ω . Dann ist \mathcal{A} auch abzählbar durchschnittsstabil; insbesondere gilt mit $A_1 := A$, $A_2 := B$, $A_n := \Omega$ für $n \geq 2$:

$$A \cap B = \bigcap_{n=1}^{\infty} A_n \in \mathcal{A}.$$

Die vierte Eigenschaft ist trivialerweise erfüllt, da sie ja sogar für beliebige (nicht notwendig paarweise disjunkte) Mengenfolgen $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ gilt.

Es mögen nun umgekehrt die obigen vier Eigenschaften gelten; $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ sei des weiteren eine beliebige Mengenfolge. Induktiv folgt zunächst, dass mit je zwei auch der Durchschnitt *endlich vieler* Mengen aus \mathcal{A} zu \mathcal{A} gehört. Definiere

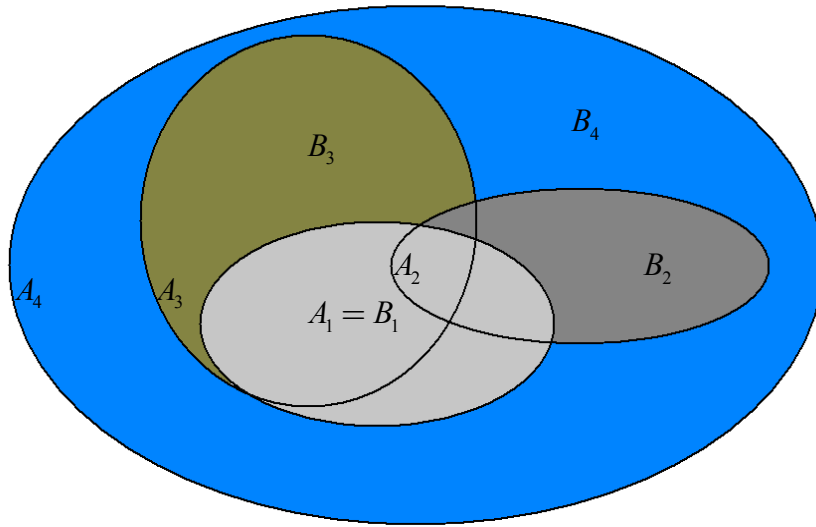
$$\begin{aligned} B_1 &:= A_1 \quad \text{und} \\ B_n &:= A_n \cap \bigcap_{i=1}^{n-1} A_i^c \quad \text{für } n \geq 2. \end{aligned}$$

Dann gehören alle Mengen B_n zu \mathcal{A} , die Mengen sind paarweise disjunkt, und es ist

$$\bigcup_{n=1}^{\infty} A_n = \bigoplus_{n=1}^{\infty} B_n \in \mathcal{A}$$

I. Maßtheoretische Grundlagen

(sog. *Zwiebelschalen-Argument*; die Mengen B_n enthalten von A_n nämlich genau das, was nicht schon in den "früheren" Mengen A_1, \dots, A_{n-1} enthalten ist, vgl. die nachfolgende Skizze). ■



Triviale Beispiele für σ -Algebren sind u.a.: $\mathcal{A} = \mathfrak{P}(\Omega)$ oder $\mathcal{A} = \{\emptyset, \Omega\}$.

Es ist häufig wichtig, möglichst "kleine" σ -Algebren zur Verfügung zu haben, die gewisse vorgegebene Teilmengen von Ω als Elemente enthalten. Die Konstruktion solcher σ -Algebren beruht auf dem nachfolgenden

Lemma 3. Es sei Ω eine nicht-leere Menge und \mathfrak{A} ein System von σ -Algebren über Ω . Dann ist auch $\mathcal{B} = \bigcap_{\mathcal{A} \in \mathfrak{A}} \mathcal{A}$ eine σ -Algebra über Ω . Ist $\mathcal{E} \subseteq \mathfrak{P}(\Omega)$ ein beliebiges System von Teilmengen von Ω , so existiert eine bezüglich Inklusion eindeutig bestimmte kleinste σ -Algebra, die \mathcal{E} als Teilmenge enthält, nämlich $\sigma(\mathcal{E}) := \bigcap_{\mathcal{A} \in \mathfrak{A}_{\mathcal{E}}} \mathcal{A}$, wobei $\mathfrak{A}_{\mathcal{E}}$ das System *aller* σ -Algebren über Ω bezeichnet, die \mathcal{E} als Teilmenge enthalten. \mathcal{E} heißt auch *Erzeuger der σ -Algebra $\sigma(\mathcal{E})$* .

Beweis: Zeige zunächst: $\mathcal{B} = \bigcap_{\mathcal{A} \in \mathfrak{A}} \mathcal{A}$ ist eine σ -Algebra über Ω :

Es ist $\emptyset \in \mathcal{B}$, da $\emptyset \in \mathcal{A}$ gilt für alle $\mathcal{A} \in \mathfrak{A}$ (Normiertheit).

Ist $A \in \mathcal{B}$, so ist $A \in \mathcal{A}$ für alle $\mathcal{A} \in \mathfrak{A}$, demnach auch $A^c \in \mathcal{A}$ für alle $\mathcal{A} \in \mathfrak{A}$, also $A^c \in \mathcal{B}$ (Komplementstabilität).

Ist $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{B}$, so ist $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ für alle $\mathcal{A} \in \mathfrak{A}$, demnach auch $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$ für alle $\mathcal{A} \in \mathfrak{A}$,

also $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}$ (abzählbare Vereinigungsstabilität).

Damit ist $\mathcal{B} = \bigcap_{\mathcal{A} \in \mathfrak{A}} \mathcal{A}$ eine σ -Algebra über Ω , wie behauptet.

I. Maßtheoretische Grundlagen

Insbesondere ist damit $\sigma(\mathcal{E})$ eine σ -Algebra über Ω , die in jeder anderen σ -Algebra \mathcal{A} über Ω , die \mathcal{E} umfaßt, als Teilmenge enthalten und damit bzgl. Inklusion minimal ist. ■

Für die obigen beiden Beispiele gilt etwa:

$$\mathcal{A} = \mathfrak{P}(\Omega) = \sigma(\{\{\omega\} \mid \omega \in \Omega\}) \text{ (falls } \Omega \text{ abzählbar ist!)} \text{ oder } \mathcal{A} = \{\emptyset, \Omega\} = \sigma(\{\emptyset\}) = \sigma(\{\Omega\}).$$

Weitere strukturelle Eigenschaften von σ -Algebren können elegant mit Hilfe der folgenden Begriffsbildung untersucht werden.

Definition 4 (Atom). Es sei (Ω, \mathcal{A}) ein Messraum. Eine nicht-leere Menge $A \in \mathcal{A}$ heißt *Atom* von \mathcal{A} , wenn A minimales Element bezüglich Inklusion ist in $\mathcal{A} \setminus \{\emptyset\}$, d.h. wenn gilt

$$B \in \mathcal{A}, B \neq \emptyset, B \subseteq A \Rightarrow B = A.$$

Die Menge \mathcal{T} aller Atome einer σ -Algebra \mathcal{A} wird als *Atomsystem* bezeichnet. Ein Atomsystem heißt *erschöpfend*, wenn $\mathcal{A} = \sigma(\mathcal{T})$ gilt; es heißt *vollständig*, wenn es zu jedem $\omega \in \Omega$ ein Atom $T_\omega \in \mathcal{T}$ gibt mit $\omega \in T_\omega$.

Für die obigen beiden Beispiele gilt etwa

$$\mathcal{T} = \{\{\omega\} \mid \omega \in \Omega\} \text{ für } \mathcal{A} = \mathfrak{P}(\Omega) \text{ (falls } \Omega \text{ abzählbar ist)} \text{ oder } \mathcal{T} = \{\Omega\} \text{ für } \mathcal{A} = \{\emptyset, \Omega\}.$$

Hier sind beide Atomsysteme zugleich erschöpfend und vollständig. Ein wichtiges Beispiel eines vollständigen, aber nicht erschöpfenden Atomsystems werden später im Zusammenhang mit der sog. *Borel'schen σ -Algebra* kennenlernen. Es gibt auch σ -Algebren, die keine Atome besitzen; ein solches Beispiel werden wir im Zusammenhang mit den sog. *Produkt- σ -Algebren* später kennenlernen.

Lemma 4. Es sei \mathcal{T} Atomsystem in einem Messraum (Ω, \mathcal{A}) . Dann gilt:

$$\forall S, T \in \mathcal{T} : S \cap T \neq \emptyset \Rightarrow S = T,$$

d.h. je zwei Atome einer σ -Algebra sind entweder disjunkt oder identisch. Ist das Atomsystem vollständig, so wird durch

$$\omega \sim_{\mathcal{T}} \eta \Leftrightarrow \exists T \in \mathcal{T} : \omega \in T \wedge \eta \in T$$

eine Äquivalenzrelation auf Ω definiert.

I. Maßtheoretische Grundlagen

Beweis: Seien S, T Atome mit $W := S \cap T \neq \emptyset$. Es ist offenbar $W \subseteq S$ und $W \subseteq T$ mit $W \in \mathcal{A}$, so dass wegen der Minimalität der Atome folgt: $S = W = T$, was zu zeigen war. Im Falle der Vollständigkeit bilden die Atome in \mathcal{T} eine disjunkte Zerlegung von Ω , weil zu jedem $\omega \in \Omega$ ein Atom $T_\omega \in \mathcal{T}$ existiert mit $\omega \in T_\omega$, so dass

$$\Omega = \bigcup_{\omega \in \Omega} \{\omega\} \subseteq \bigcup_{\omega \in \Omega} T_\omega \subseteq \bigcup_{T \in \mathcal{T}} T \subseteq \Omega, \text{ also } \Omega = \bigoplus_{T \in \mathcal{T}} T$$

gilt, da in \mathcal{T} als Menge nur paarweise verschiedene, also, nach obigem, paarweise disjunkte Atome gezählt werden. Gemäß den Ausführungen zu Lemma 2 wird also durch das Atomsystem die genannte Äquivalenzrelation erzeugt. ■

Ist die Grundmenge Ω abzählbar, vereinfacht sich die Struktur von σ -Algebren erheblich. Dies liegt u.a. an dem folgenden, fundamentalen Resultat:

Satz 2 (Struktursatz für σ -Algebren I). Es sei (Ω, \mathcal{A}) ein Messraum mit *abzählbarer* Grundmenge Ω . Dann gilt:

- a) \mathcal{A} besitzt ein erschöpfendes, vollständiges, *abzählbares* Atomsystem $\mathcal{T} = \{T_i \mid i \in I\}$ mit einer endlichen oder unendlichen Indexmenge $I \subseteq \mathbb{N}$.
- b) Jedes Element A von \mathcal{A} besitzt eine eindeutige Darstellung

$$A = \bigoplus_{i \in I_A} T_i \text{ mit einer eindeutigen Indexmenge } I_A \subseteq I;$$

allgemeiner gilt (mit der Konvention $\bigcup_{i \in \emptyset} T_i = \emptyset$):

$$\mathcal{A} = \left\{ \bigoplus_{i \in J} T_i \mid J \subseteq I \right\}.$$

Ist insbesondere Ω *endlich* mit $\#(\mathcal{T}) = \#(I) = n \in \mathbb{N}$, so besitzt \mathcal{A} genau 2^n Elemente.

Beweis:

- a) Wir zeigen zunächst, dass es zu jedem $\omega \in \Omega$ ein Atom T_ω gibt mit $\omega \in T_\omega$. Dazu definieren wir

$$T_\omega := \bigcap \{A \in \mathcal{A} \mid \omega \in A\}.$$

Es bleibt lediglich zu zeigen, dass $T_\omega \in \mathcal{A}$ ist²; die Aussage ergibt sich dann wegen $\omega \in T_\omega$ sowie der konstruktionsbedingten Minimalität bzgl. Inklusion. (Denn wäre T_ω kein Atom, so gäbe es eine echte Teilmenge $S \subset T_\omega$ mit $\omega \notin S$; dann wäre aber $\omega \in T_\omega \setminus S \subset T_\omega$: Widerspruch!)

Zum Nachweis von $T_\omega \in \mathcal{A}$ betrachten wir das System

² Man beachte, dass die Potenzmenge einer abzählbar-unendlichen Menge bereits überabzählbar ist, so dass es ggf. auch überabzählbar viele Mengen $A \in \mathcal{A}$ mit $\omega \in A$ gibt.

I. Maßtheoretische Grundlagen

$$\mathcal{C}_\omega := \{A^c \in \mathcal{A} \mid \omega \in A\} = \{B \in \mathcal{A} \mid \omega \notin B\}.$$

Da Ω abzählbar ist, können wir Ω darstellen als $\Omega = \{\omega\} \oplus \{\omega_k \mid k \in K\}$ mit einer geeigneten Indexmenge $K \subseteq \mathbb{N}$ (endlich oder unendlich). Betrachte nun

$$T_\omega^c = \bigcup \{A^c \in \mathcal{A} \mid \omega \in A\} = \bigcup_{B \in \mathcal{C}_\omega} B \subseteq \Omega.$$

Es existiert daher eine Teilmenge $L \subseteq K$ mit

$$T_\omega^c = \bigcup_{B \in \mathcal{C}_\omega} B = \{\omega_\ell \mid \ell \in L\}.$$

Insbesondere existiert damit zu jedem $\ell \in L$ eine Menge $B_\ell \in \mathcal{C}_\omega \subseteq \mathcal{A}$ mit $\omega_\ell \in B_\ell$. Es folgt

$$T_\omega^c = \bigcup_{B \in \mathcal{C}_\omega} B = \{\omega_\ell \mid \ell \in L\} \subseteq \bigcup_{\ell \in L} B_\ell \subseteq \bigcup_{B \in \mathcal{C}_\omega} B, \text{ also } T_\omega^c = \bigcup_{\ell \in L} B_\ell \in \mathcal{A}$$

und somit auch $T_\omega \in \mathcal{A}$, wie zu zeigen war.

Aus der Abzählbarkeit von Ω und der gerade gezeigten Eigenschaft, dass zu jedem $\omega \in \Omega$ ein Atom T_ω existiert mit $\omega \in T_\omega$, folgt nun:

$$\Omega = \bigcup_{\omega \in \Omega} \{\omega\} \subseteq \bigcup_{\omega \in \Omega} T_\omega \subseteq \Omega, \text{ also } \Omega = \bigcup_{\omega \in \Omega} T_\omega.$$

Da nach Lemma 4 nur gelten kann: $T_\omega = T_\eta$ oder $T_\omega \cap T_\eta = \emptyset$ für alle $\omega, \eta \in \Omega$, ist also $\mathcal{T} := \{T_\omega \mid \omega \in \Omega\}$ ein vollständiges, abzählbares Atomsystem mit

$$A = \bigcup_{\omega \in A} \{\omega\} \subseteq \bigcup_{\omega \in A} T_\omega \subseteq \bigcup_{\omega \in A} A = A, \text{ also } A = \bigcup_{\omega \in A} T_\omega \text{ für alle } A \in \mathcal{A},$$

weil aufgrund der Minimalität von Atomen für jedes $\omega \in A \in \mathcal{A}$ gilt: $\omega \in T_\omega \subseteq A$. Damit ist \mathcal{T} auch erschöpfend. Ferner existiert nach Lemma 4 eine höchstens abzählbare Indexmenge $I \subseteq \mathbb{N}$, so dass nach geeigneter Nummerierung (unter Weglassung aller Duplikate von Atomen) gilt:

$$\mathcal{T} = \{T_\omega \mid \omega \in \Omega\} = \{T_{\omega_i} \mid i \in I\}.$$

(Zur Vereinfachung der Notation bleiben wir im folgenden bei der in der Satzaussage benutzten Bezeichnungsweise $\mathcal{T} = \{T_i \mid i \in I\}$.)

- b) Mit der gerade gezeigten Beziehung $A = \bigcup_{\omega \in A} T_\omega$ für alle $A \in \mathcal{A}$ ergibt sich die erste Behauptung, wenn man wieder Duplikate von Atomen T_ω aus der Vereinigung entfernt. Weil umge-

I. Maßtheoretische Grundlagen

kehrt auch jede Vereinigung $\bigoplus_{i \in J} T_i$ mit einer beliebigen Indexmenge $J \subseteq I$ zu \mathcal{A} gehört, folgt auch die zweite Aussage.

Da im Falle der Endlichkeit die Potenzmenge $\mathfrak{P}(I)$ genau 2^n Elemente enthält, ist der Satz damit vollständig bewiesen. ■

Der Beweis von Satz 2 zeigt, dass der Aussagenteil b) auch im Fall *beliebiger* Grundmengen richtig bleibt. Wir formulieren das entsprechende Resultat der Vollständigkeit halber noch einmal gesondert.

Satz 3 (Struktursatz für σ -Algebren II). Es sei (Ω, \mathcal{A}) ein *beliebiger* Messraum und $\mathcal{T} = \{T_i \mid i \in I\}$ ein abzählbares Atomsystem mit einer Indexmenge $I \subseteq \mathbb{N}$. Dann gilt: \mathcal{T} ist vollständig genau dann, wenn \mathcal{T} erschöpfend ist; \mathcal{A} kann dann dargestellt werden als

$$\mathcal{A} = \left\{ \bigoplus_{i \in J} T_i \mid J \subseteq I \right\}.$$

Beweis: " \Rightarrow ": Zu jedem $\omega \in \Omega$ existiert ein Index $i_\omega \in I$ mit $\omega \in T_{i_\omega}$. Also folgt für jede Menge $A \in \mathcal{A}$:

$$A = \bigcup_{\omega \in A} \{\omega\} \subseteq \bigcup_{\omega \in A} T_{i_\omega} \subseteq \bigcup_{\omega \in A} A = A, \text{ also } A = \bigcup_{\omega \in A} T_{i_\omega} = \bigoplus_{\omega \in I_A} T_{i_\omega},$$

wobei I_A gerade diejenige Menge $\omega \in A$ enthält, deren Atome T_{i_ω} paarweise disjunkt sind, d.h. jede Menge $A \in \mathcal{A}$ ist als disjunkte Vereinigung höchstens abzählbar vieler Atome darstellbar.

Also ist \mathcal{T} erschöpfend, und es gilt $\mathcal{A} = \left\{ \bigoplus_{i \in J} T_i \mid J \subseteq I \right\}$.

" \Leftarrow ": Es ist $T_0 := \left\{ \bigoplus_{i \in I} T_i \right\}^c \in \sigma(\mathcal{T}) = \mathcal{A}$. Ferner ist $\mathcal{A}_0 := \left\{ \bigoplus_{i \in J} T_i \mid J \subseteq I \cup \{0\} \right\} \subseteq \mathcal{A}$ eine σ -Algebra über Ω , die $\mathcal{T} = \{T_i \mid i \in I\}$ enthält, d.h. es gilt $\mathcal{A} = \mathcal{A}_0$. Nach Voraussetzung ist aber T_0 kein Atom, woraus $T_0 = \emptyset$ und damit die Behauptung folgt, womit der Satz bewiesen ist. ■

Satz 3 kann oft direkt zur Konstruktion von σ -Algebren bei beliebigen Grundmengen Ω herangezogen werden. Hierzu betrachten wir zwei Beispiele:

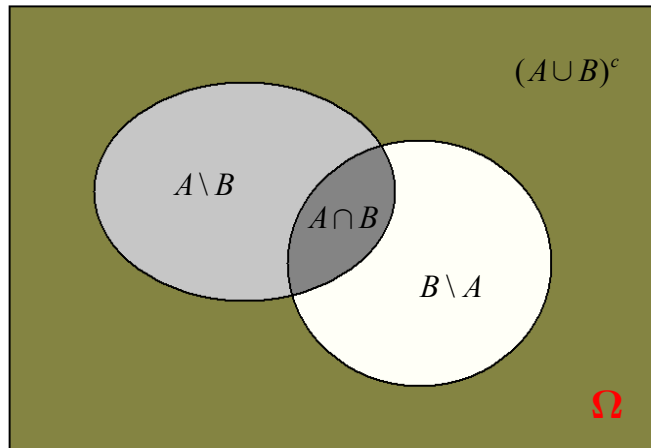
1. Sei $\emptyset \subset A \subset \Omega$. Die σ -Algebra $\mathcal{A} = \sigma(\{A\})$ besitzt offenbar das Atomsystem $\mathcal{T} = \{A, A^c\}$, so dass \mathcal{A} die folgenden $4 = 2^2$ Elemente besitzt:

$$\mathcal{A} = \sigma(\{A\}) = \{\emptyset, A, A^c, \Omega\}.$$

I. Maßtheoretische Grundlagen

2. Sei $A, B \subset \Omega$ mit $A \not\subset B$, $B \not\subset A$, $A \neq B$, $A \cap B \neq \emptyset$, $A \cup B \neq \Omega$. Die σ -Algebra $\mathcal{A} = \sigma(\{A, B\})$ besitzt dann das Atomsystem $\mathcal{T} = \{A \setminus B, B \setminus A, A \cap B, (A \cup B)^c\}$ (vgl. nachfolgende Skizze), so dass \mathcal{A} die folgenden $16 = 2^4$ Elemente besitzt:

$$\begin{aligned} \mathcal{A} = \sigma(\{A, B\}) = \{ & \emptyset, A \setminus B, B \setminus A, A \cap B, A, B, A \cup B, (A \cup B)^c, (A \setminus B) \cup (B \setminus A), \\ & (A \setminus B) \cup (A \cup B)^c, (B \setminus A) \cup (A \cup B)^c, (A \cap B) \cup (A \cup B)^c, A \cup (A \cup B)^c, \\ & B \cup (A \cup B)^c, (A \setminus B) \cup (B \setminus A) \cup (A \cup B)^c = (A \cap B)^c, \Omega \} \end{aligned}$$



Eine Struktur, die vor allem bei mehrfach wiederholten Experimenten wichtig ist, ist die sogenannte *Produkt- σ -Algebra*.

Definition 5 (Produkt- σ -Algebra). Es seien $(\Omega_i, \mathcal{A}_i)$, $i \in I$ Messräume mit einer *endlichen* Indexmenge I . Das Produkt $\mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i$ der σ -Algebren \mathcal{A}_i , $i \in I$ (Produkt- σ -Algebra) über dem kartesischen Produkt $\Omega = \prod_{i \in I} \Omega_i$ ist definiert durch

$$\bigotimes_{i \in I} \mathcal{A}_i := \sigma(\mathcal{E}) \quad \text{mit} \quad \mathcal{E} := \left\{ \prod_{i \in I} A_i \mid \forall i \in I : A_i \in \mathcal{A}_i \right\}.$$

Das Paar $\left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{A}_i \right)$ heißt entsprechend *Produkt-Messraum*.

Man beachte hier, dass der Erzeuger \mathcal{E} i.Allg. selbst keine σ -Algebra ist, wie schon das obige, unter 1. angegebene einfache Beispiel $\mathcal{A}_1 = \mathcal{A}_2 = \{\emptyset, A, A^c, \Omega\}$ mit $\Omega_1 = \Omega_2 = \Omega$, $\emptyset \subset A \subset \Omega$ zeigt: so gilt etwa

$$B = (A \times A) \cup (A^c \times A^c) \in \mathcal{A}_1 \otimes \mathcal{A}_1,$$

jedoch ist B offenkundig kein Element von \mathcal{E} .

I. Maßtheoretische Grundlagen

Im Fall abzählbarer Grundmengen Ω_i lässt sich die Produkt- σ -Algebra aber wieder einfach über die Atome der einzelnen σ -Algebren charakterisieren. Dies ergibt sich etwa aus

Satz 4 (Struktursatz für σ -Algebren III). Es seien $(\Omega_i, \mathcal{A}_i)$, $i \in I$ beliebige Messräume mit einer endlichen Indexmenge I , wobei vorausgesetzt sei, dass sämtliche σ -Algebren \mathcal{A}_i vollständige, abzählbare Atomsysteme \mathcal{T}_i , $i \in I$ besitzen. Dann ist

$$\mathcal{T} := \left\{ \times_{i \in I} T_i \mid \forall i \in I : T_i \in \mathcal{T}_i \right\}$$

das eindeutig bestimmte, erschöpfende, vollständige, abzählbare Atomsystem für die Produkt- σ -Algebra $\mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i$.

Beweis: Sei \mathcal{E} wie in Definition 5. Dann ist offensichtlich $\mathcal{T} = \left\{ \times_{i \in I} T_i \mid \forall i \in I : T_i \in \mathcal{T}_i \right\} \subseteq \mathcal{E}$. Ferner sind alle Elemente von \mathcal{T} paarweise disjunkt (vgl. Lemma 1). Da jede Menge in \mathcal{E} darstellbar ist als $\times_{i \in I} A_i = \times_{i \in I} \bigcup_{j \in J_i} T_{ij}$ (mit geeigneten Atomen $T_{ij} \in \mathcal{T}_i$) und diese wiederum analog Lemma 1 als abzählbare Vereinigung von endlichen kartesischen Produkten aus Atomen geschrieben werden kann, gilt auch $\mathcal{E} \subseteq \sigma(\mathcal{T})$ und damit $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{T}) \subseteq \sigma(\mathcal{E})$, also $\mathcal{A} = \sigma(\mathcal{E}) = \sigma(\mathcal{T})$. Da \mathcal{T} abzählbar ist und die Vereinigung aller Elemente von \mathcal{T} ganz Ω ergibt, kann jede Menge aus $\mathcal{A} = \sigma(\mathcal{T})$ dargestellt werden als abzählbare disjunkte Vereinigung von Mengen der Form $\times_{i \in I} T_i$ mit $T_i \in \mathcal{T}_i$. Wegen der Minimalität von Atomen ist also notwendig jedes Atom T von $\mathcal{A} = \sigma(\mathcal{T})$ von der Form $T = \times_{i \in I} T_i$ mit $T_i \in \mathcal{T}_i$. Damit ist \mathcal{T} Atomsystem für die Produkt- σ -Algebra $\mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i$. Aufgrund der Vollständigkeit gibt es zu jedem $\omega_i \in \Omega_i$ ein Atom $T_{\omega_i} \in \mathcal{T}_i$ mit $\omega_i \in T_{\omega_i}$, so dass für $\omega = (\omega_i)_{i \in I} \in \times_{i \in I} \Omega_i$ folgt: $\omega \in \times_{i \in I} T_{\omega_i} \in \mathcal{T}$, d.h. \mathcal{T} ist auch vollständig. ■

Für das Beispiel $\mathcal{A}_1 = \mathcal{A}_2 = \{\emptyset, A, A^c, \Omega\}$ erhält man etwa $\mathcal{T}_1 = \mathcal{T}_2 = \{A, A^c\}$ und damit $\mathcal{T} = \{A \times A, A \times A^c, A^c \times A, A^c \times A^c\}$, so dass die Produkt- σ -Algebra $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ hier aus $16 = 2^4$ Elementen besteht.

Gelegentlich ist es nötig, den Grundraum Ω eines Messraums (Ω, \mathcal{A}) auf eine kleinere Teilmenge $\Xi \subset \Omega$ einzuschränken. Man kann dann die ursprüngliche σ -Algebra \mathcal{A} auch entsprechend reduzieren.

Lemma 5 (Spur- σ -Algebra). Es sei (Ω, \mathcal{A}) ein beliebiger Messraum und $\Xi \subset \Omega$ eine nicht-leere Teilmenge. Dann wird durch

$$\mathcal{A}_\Xi := \{\Xi \cap A \mid A \in \mathcal{A}\}$$

I. Maßtheoretische Grundlagen

eine σ -Algebra über Ξ definiert. Diese heißt Spur- σ -Algebra von \mathcal{A} in Ξ . Falls bereits $\Xi \in \mathcal{A}$, so gilt einfacher

$$\mathcal{A}_\Xi := \{\Xi \cap A \mid A \in \mathcal{A}\} = \{B \in \mathcal{A} \mid B \subseteq \Xi\}.$$

Beweis: Es ist $\Xi = \Xi \cap \Omega \in \mathcal{A}_\Xi$. Für $A \in \mathcal{A}$ ist $\Xi \setminus (\Xi \cap A) = \Xi \setminus A = \Xi \cap A^c$, also mit $\Xi \cap A$ auch das Komplement (bzgl. Ξ) in \mathcal{A}_Ξ als Element enthalten. Für $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ ist schließlich $\Xi \cap \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} (\Xi \cap A_n) \in \mathcal{A}_\Xi$, also \mathcal{A}_Ξ eine σ -Algebra über Ξ . Die zweite Aussage folgt aus der Tatsache, dass mit $\Xi \in \mathcal{A}$ auch $\Xi \cap A \in \mathcal{A}$ gilt für alle $A \in \mathcal{A}$. ■

I. 3. Maß und Wahrscheinlichkeit

Der Begriff der "Wahrscheinlichkeit" ist seit dem Beginn des 20. Jahrhunderts axiomatisch eng mit dem eines "Maßes" verbunden, der seinerseits auf das seit Jahrtausenden bekannte Problem der Bestimmung von Länge, Fläche oder Volumen zurückgeht. Erinnerung sei hier exemplarisch an die klassische griechische Mathematik mit dem Versuch, die Fläche eines Kreises durch Approximation mit regelmäßigen Polyedern zu bestimmen. Um so erstaunlicher ist die Tatsache, dass eine sehr reichhaltige Theorie des "Messens" mit einer vergleichsweise geringen Axiomatik auskommt.

Definition 6 (Maß und Maßraum). Es sei \mathcal{A} eine σ -Algebra über einer nicht-leeren Menge Ω . Eine Abbildung $\mu: \mathcal{A} \rightarrow \bar{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ heißt *Maß*, wenn folgende drei definierenden Eigenschaften gegeben sind:

$$\begin{aligned} \mu(\emptyset) &= 0 && \text{(Normiertheit)} \\ \forall A \in \mathcal{A}: \mu(A) &\geq 0 && \text{(Nichtnegativität)} \\ \forall \{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A} \text{ paarweise disjunkt: } \mu\left(\bigoplus_{n=1}^{\infty} A_n\right) &= \sum_{n=1}^{\infty} \mu(A_n) && \text{(\sigma-Additivität)}. \end{aligned}$$

Das Tripel $(\Omega, \mathcal{A}, \mu)$ heißt *Maßraum*. Dabei mögen in $\bar{\mathbb{R}}$ folgende erweiterte Rechenregeln gelten:

$$\forall x \in \bar{\mathbb{R}}: x + \infty = \infty + x = \infty.$$

In manchen Fällen betrachtet man auch Abbildungen μ auf einer nicht-leeren Teilmenge $\mathcal{S} \subseteq \mathcal{A}$ mit entsprechenden Eigenschaften, sofern alle beteiligten Mengen in \mathcal{S} liegen (insbesondere die disjunkte Vereinigung). Man spricht dann von einem *Prämaß* auf \mathcal{S} .

Gilt für das Maß $\mu(\Omega) = 1$, so heißt μ *Wahrscheinlichkeitsmaß* (bzw. synonym: *Wahrscheinlichkeitsverteilung*); es wird in der Regel mit dem Symbol P bezeichnet (lat.: *probabilitas*, engl.: *pro-*

I. Maßtheoretische Grundlagen

bability, frz.: *probabilité*). Jedes nicht-triviale endliche Maß μ (d.h. mit der Eigenschaft $\mu(\Omega) > 0$) kann durch Normierung zu einem Wahrscheinlichkeitsmaß P gemacht werden vermöge

$$P(A) := \frac{\mu(A)}{\mu(\Omega)}, \quad A \in \mathcal{A}.$$

Unter dem trivialen Maß μ auf \mathcal{A} verstehen wir das Maß mit der Eigenschaft $\forall A \in \mathcal{A} : \mu(A) = 0$. Es wird auch als *Nullmaß* bezeichnet; allgemeiner bezeichnen wir Mengen $A \in \mathcal{A}$ mit der Eigenschaft $\mu(A) = 0$ als (μ) -*Nullmengen*. Beim trivialen Maß sind also alle messbaren Mengen Nullmengen. Ist $A = \{\omega \in \Omega \mid \omega \text{ besitzt die Eigenschaft } \mathfrak{E}\} \in \mathcal{A}$ ein Ereignis mit $\mu(A^c) = 0$, so sagt man, die Eigenschaft \mathfrak{E} bestehe μ -fast überall. Ist (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, so sagt man auch, die Eigenschaft \mathfrak{E} bestehe P -fast sicher. Ist ferner $\Xi \in \mathcal{A}$ eine nicht-leere Teilmenge, so heißt das durch

$$\nu(A) := \mu(\Xi \cap A) \quad \text{für alle } A \in \mathcal{A}$$

definierte Maß das *Spurmaß* von μ bezüglich Ξ .

Auf das schon angesprochene Problem der Längen-, Flächen- und Volumenmessung werden wir in diesem Zusammenhang in einem gesonderten Kapitel später zurückkommen.

Ein einfaches Beispiel für ein Maß ist die im Zusammenhang mit Definition 2 angesprochene Anzahl der Elemente einer Menge. Es gilt nämlich:

Lemma 6. Das Tripel $(\Omega, \mathfrak{P}(\Omega), \#)$ ist für jede nicht-leere Menge Ω stets ein Maßraum.

Beweis: Trivialerweise ist $\#(\emptyset) = 0$, weil die leere Menge kein Element enthält. Die σ -Additivität ergibt sich aus Definition 2: Sind die $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A} = \mathfrak{P}(\Omega)$ paarweise disjunkt, so ist entweder $\#(A_n) = 0$ ($\Leftrightarrow A_n = \emptyset$) oder $\#(A_n) \geq 1$ für alle $n \in \mathbb{N}$. Sei $I = \{n \in \mathbb{N} \mid \#(A_n) \geq 1\}$. Für $\#(I) < \infty$ ist

$$\#\left(\bigoplus_{n=1}^{\infty} A_n\right) = \#\left(\bigoplus_{n \in I} A_n\right) = \sum_{n \in I} \#(A_n) = \sum_{n=1}^{\infty} \#(A_n).$$

Für $\#(I) = \infty$ gilt $\#\left(\bigoplus_{n=1}^{\infty} A_n\right) \geq \#(I) = \infty$ und $\sum_{n=1}^{\infty} \#(A_n) \geq \sum_{n=1}^{\infty} 1 = \#(I) = \infty$, also ebenfalls

$$\#\left(\bigoplus_{n=1}^{\infty} A_n\right) = \infty = \sum_{n=1}^{\infty} \#(A_n). \quad \blacksquare$$

Man nennt die Anzahlabbildung " $\#$ " in diesem Zusammenhang auch das *abzählende Maß* oder kurz *Zählmaß*.

I. Maßtheoretische Grundlagen

Definition 7 (Gleichverteilung). Ist Ω endlich, so heißt das durch die obige Normierung erhaltene Wahrscheinlichkeitsmaß P mit

$$P(A) := \frac{\#(A)}{\#(\Omega)}, \quad A \in \mathfrak{P}(\Omega)$$

die *diskrete Gleichverteilung* über Ω (auch: Laplace-Verteilung).

Hierauf kommen wir später noch einmal ausführlicher bei der allgemeinen Diskussion diskreter stochastischer Modelle zurück.

Die Gleichverteilung ist historisch eines der ersten stochastischen Modelle überhaupt und motiviert die Bedeutung der *Kombinatorik*, in deren Rahmen häufig die Bestimmung der Anzahl der Elemente "komplizierter" Mengen A erfolgt.

Aus den definierenden Eigenschaften eines Maßes ergeben sich sofort eine Reihe von sehr nützlichen "Rechenregeln", die im folgenden näher vorgestellt werden.

Satz 5. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Dann gilt für alle Ereignisse $A, B \in \mathcal{A}$, $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$:

$$\mu(A \oplus B) = \mu(A) + \mu(B) \quad \text{für } A \cap B = \emptyset \quad (\text{endliche Additivität})$$

$$A \subseteq B \Rightarrow \mu(A) \leq \mu(B) \quad (\text{Monotonie})$$

$$A \subseteq B, \mu(A) < \infty \Rightarrow \mu(B \setminus A) = \mu(B) - \mu(A) \quad (\text{Subtraktivität})$$

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$$

bzw. allgemeiner, unter der einschränkenden Voraussetzung $\mu(A_i) < \infty$, $1 \leq i \leq n$:

$$\mu\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < k_2 < \dots < k_i \leq n} \mu\left(\bigcap_{j=1}^i A_{k_j}\right) \quad (\text{Siebformel von Sylvester-Poincaré})$$

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots, \bigcup_{n \in \mathbb{N}} A_n = A \Rightarrow \mu(A) = \lim_{n \rightarrow \infty} \mu(A_n) \quad (\text{Stetigkeit von unten})$$

$$A_1 \supseteq A_2 \supseteq \dots, \mu(A_1) < \infty, \bigcap_{n \in \mathbb{N}} A_n = A \Rightarrow \mu(A) = \lim_{n \rightarrow \infty} \mu(A_n) \quad (\text{Stetigkeit von oben})$$

$$\mu\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mu(A_i), \quad n \in \mathbb{N} \cup \{\infty\} \quad ((\sigma\text{-})\text{Sub-Additivität})$$

Beweis:

Endliche Additivität: wähle $A_1 := A$, $A_2 := B$, $A_n := \emptyset$ für $n \geq 3$, dann ist

$$\mu(A \cup B) = \mu\left(\bigoplus_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n) = \mu(A) + \mu(B) + \sum_{n=3}^{\infty} \mu(\emptyset) = \mu(A) + \mu(B),$$

I. Maßtheoretische Grundlagen

wie behauptet.

Monotonie: wähle $C := B \setminus A = B \cap A^c \in \mathcal{A}$, dann ist $B = A \oplus C$ mit

$$\mu(A) \leq \mu(A) + \mu(C) = \mu(A \oplus C) = \mu(B),$$

wie behauptet.

Subtraktivität: aus dem Vorhergehenden folgt sofort:

$$\mu(B \setminus A) = \mu(C) = \mu(B) - \mu(A),$$

wie behauptet.

Siebformel: o.B.d.A. können wir für die erste Aussage $\mu(A) < \infty$ und $\mu(B) < \infty$ voraussetzen (anderenfalls steht wegen $\mu(A \cup B) \geq \max\{\mu(A), \mu(B)\}$ auf beiden Seiten der Gleichung der Wert ∞). Es folgt:

$$\begin{aligned} \mu(A \cup B) + \mu(A \cap B) &= \mu((A \setminus [A \cap B]) \oplus (A \cap B) \oplus (B \setminus [A \cap B])) + \mu(A \cap B) \\ &= \mu(A) - \mu(A \cap B) + \mu(A \cap B) + \mu(B) - \mu(A \cap B) + \mu(A \cap B) = \mu(A) + \mu(B). \end{aligned}$$

Diese erste Aussage können wir nun zum induktiven Beweis der allgemeinen Siebformel heranziehen, die für $n = 1$ offensichtlich richtig ist. Sei die Siebformel also für ein $n \in \mathbb{N}$ gültig. Wir erhalten dann mit der vorherigen Aussage:

$$\begin{aligned} \mu\left(\bigcup_{i=1}^{n+1} A_i\right) &= \mu\left(\bigcup_{i=1}^n A_i\right) + \mu(A_{n+1}) - \mu\left(\bigcup_{i=1}^n (A_i \cap A_{n+1})\right) \\ &= \left\{ \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < k_2 < \dots < k_i \leq n} \mu\left(\bigcap_{j=1}^i A_{k_j}\right) \right\} + \mu(A_{n+1}) - \sum_{i=1}^n (-1)^{i-1} \sum_{1 \leq k_1 < k_2 < \dots < k_i \leq n} \mu\left(\bigcap_{j=1}^i (A_{k_j} \cap A_{n+1})\right) \\ &= \sum_{i=1}^{n+1} (-1)^{i-1} \sum_{1 \leq k_1 < k_2 < \dots < k_i \leq n+1} \mu\left(\bigcap_{j=1}^i A_{k_j}\right), \end{aligned}$$

also die entsprechende Siebformel für $n + 1$.

Stetigkeit von unten: wähle $B_1 := A_1$, $B_n := A_n \setminus \bigcup_{i=1}^{n-1} A_i = A_n \setminus A_{n-1}$ für $n \geq 2$, so ist $A_n = \bigoplus_{i=1}^n B_i$ für alle

$n \in \mathbb{N}$ mit $\bigcup_{i=1}^{\infty} A_i = \bigoplus_{i=1}^{\infty} B_i$, also

$$\mu(A) = \mu\left(\bigoplus_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(B_i) = \lim_{n \rightarrow \infty} \mu\left(\bigoplus_{i=1}^n B_i\right) = \lim_{n \rightarrow \infty} \mu(A_n),$$

wie behauptet.

I. Maßtheoretische Grundlagen

Stetigkeit von oben: wähle $B_n := A_1 \setminus A_n$ für $n \in \mathbb{N}$, dann gilt $\emptyset = B_1 \subseteq B_2 \subseteq B_3 \subseteq \dots \subseteq B := \bigcup_{i=1}^{\infty} B_i$

mit $B = \bigcup_{i=1}^{\infty} (A_1 \setminus A_i) = A_1 \cap \left(\bigcap_{i=1}^{\infty} A_i \right)^c = A_1 \setminus A$ bzw. $A_n = A_1 \setminus B_n$, $A = A_1 \setminus B$, also folgt mit der Stetigkeit von unten und der Subtraktivität:

$$\mu(A) = \mu(A_1 \setminus B) = \mu(A_1) - \mu(B) = \mu(A_1) - \lim_{n \rightarrow \infty} \mu(B_n) = \lim_{n \rightarrow \infty} \mu(A_1 \setminus B_n) = \lim_{n \rightarrow \infty} \mu(A_n),$$

wie behauptet.

σ -Subadditivität: dies zeigen wir mit vollständiger Induktion aus der einfachen Siebformel:

Die Behauptung stimmt trivialerweise für $n = 1$. Sie gelte nun für ein beliebiges $n \in \mathbb{N}$. Dann ist

$$\mu\left(\bigcup_{i=1}^{n+1} A_i\right) = \mu\left(A_{n+1} \cup \bigcup_{i=1}^n A_i\right) \leq \mu(A_{n+1}) + \mu\left(\bigcup_{i=1}^n A_i\right) \leq \mu(A_{n+1}) + \sum_{i=1}^n \mu(A_i) = \sum_{i=1}^{n+1} \mu(A_i),$$

Womit der Induktionsschritt gezeigt ist. Die Aussage gilt also (zunächst) für alle $n \in \mathbb{N}$. Sie gilt aber auch für $n = \infty$, denn es ist

$$\mu\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n \mu(A_i) \leq \sum_{i=1}^{\infty} \mu(A_i) \quad \text{für alle } n \in \mathbb{N}$$

mit einer monoton wachsenden Mengenfolge $\left\{ \bigcup_{i=1}^n A_i \right\}_{n \in \mathbb{N}}$. Falls $\sum_{i=1}^{\infty} \mu(A_i) = \infty$ gilt, ist weiter

nichts zu zeigen. Ansonsten ist auch die Folge $\left\{ \mu\left(\bigcup_{i=1}^n A_i\right) \right\}_{n \in \mathbb{N}}$ monoton wachsend; da sie nach oben beschränkt ist, ist sie also konvergent mit

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i),$$

wegen der Stetigkeit von μ von unten, wie behauptet. ■

Bisher haben wir Maße nur "abstrakt" studiert, d.h. grundlegende mathematische Eigenschaften abgeleitet, die sich aus der formalen Definition 6 ergeben. Es stellt sich nun natürlich die Frage, wie man Maße "konkret" angeben kann, wie etwa im Fall der diskreten Gleichverteilung. Dabei ist es nützlich, zu wissen, welche Mindestangaben für eine eindeutige Bestimmung des Maßes erforderlich sind. Dies ist besonders einfach im Fall der Existenz von Atomsystemen und führt allgemeiner auf die Frage der Fortsetzbarkeit von Maßen.

I. Maßtheoretische Grundlagen

Satz 6 (kleiner Maßfortsetzungssatz). Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $\mathcal{T} = \{T_i \mid i \in I\}$ ein vollständiges, abzählbares Atomsystem für \mathcal{A} mit einer abzählbaren Indexmenge I . Dann ist μ bereits eindeutig durch die Werte $\mu(T_i)$, $i \in I$ bestimmt. Ist umgekehrt $\{\mu_i\}_{i \in I}$ eine Folge nicht-negativer erweitert-reeller Zahlen, so wird durch

$$\mu(T_i) := \mu_i, \quad i \in I$$

in eindeutiger Weise ein Maß μ auf \mathcal{A} definiert. Gilt zusätzlich $\sum_{i \in I} \mu_i = 1$, so ist μ ein Wahrscheinlichkeitsmaß auf \mathcal{A} . Falls Ω selbst höchstens abzählbar ist, kann μ sogar (i.Allg. allerdings nicht mehr eindeutig) auf die maximale σ -Algebra $\mathfrak{B}(\Omega)$ fortgesetzt werden.

Beweis: Nach Satz 3 kann jede Menge $A \in \mathcal{A}$ in eindeutiger Weise dargestellt werden als $A = \bigoplus_{i \in J} T_i$ mit einer geeigneten Indexmenge $J \subseteq I$. Damit ist

$$\mu(A) = \mu\left(\bigoplus_{i \in J} T_i\right) = \sum_{i \in J} \mu(T_i),$$

d.h. μ ist eindeutig durch die Werte $\mu(T_i)$, $i \in I$ bestimmt. Umgekehrt wird bei gegebener Werten $\{\mu_i\}_{i \in I}$ durch

$$\mu(A) = \mu\left(\bigoplus_{i \in J} T_i\right) := \sum_{i \in J} \mu_i, \quad A \in \mathcal{A}$$

ein Maß μ in eindeutiger Weise auf \mathcal{A} definiert, das im Falle von $\mu(\Omega) = \mu\left(\bigoplus_{i \in I} T_i\right) = \sum_{i \in I} \mu_i = 1$ auch ein Wahrscheinlichkeitsmaß auf \mathcal{A} ist.

Falls Ω abzählbar ist, ist auch jedes der Atome T_i , $i \in I$ als Teilmenge von Ω selbst abzählbar, etwa

$$T_i = \{t_{ij} \mid j \in K_i\} \quad \text{mit einer abzählbaren Indexmenge } K_i.$$

Dann ist $\Omega = \{t_{ij} \mid i \in I, j \in K_i\}$. Es bezeichne $\mu_i := \mu(T_i)$, $i \in I$. Wähle für jedes $i \in I$ eine (ggf. endliche) Folge $\{\mu_{ij}\}_{j \in K_i}$ nicht-negativer erweitert-reeller Zahlen mit $\mu_i = \sum_{j \in K_i} \mu_{ij}$, $i \in I$ (eine solche Auswahl ist i.Allg. nicht eindeutig). Durch die Zuordnung

$$\tilde{\mu}(\{t_{ij}\}) := \mu_{ij}, \quad i \in I, j \in K_i$$

wird nun nach obigem ein Maß $\tilde{\mu}$ auf $\mathfrak{B}(\Omega)$ definiert mit

$$\tilde{\mu}(T_i) = \sum_{j \in K_i} \tilde{\mu}(\{t_{ij}\}) = \sum_{j \in K_i} \mu_{ij} = \mu_i = \mu(T_i), \quad i \in I,$$

I. Maßtheoretische Grundlagen

d.h. $\tilde{\mu}$ stimmt mit μ auf allen Atomen T_i , $i \in I$ und damit auf \mathcal{A} überein und ist somit eine Fortsetzung von μ auf $\mathfrak{P}(\Omega)$. ■

Satz 6 besagt wegen Satz 2 im Wesentlichen, dass in jedem diskreten Maßraum $(\Omega, \mathcal{A}, \mu)$ (d.h. mit höchstens abzählbarer Grundmenge Ω) das Maß μ stets auf die gesamte Potenzmenge $\mathfrak{P}(\Omega)$ fortgesetzt werden kann. Insofern können von vornherein auch alle diskreten stochastischen Modelle, d.h. diskreten Maßräume (Ω, \mathcal{A}, P) mit einem Wahrscheinlichkeitsmaß P , o.B.d.A. als maximal, d.h. mit $\mathcal{A} = \mathfrak{P}(\Omega)$, angenommen werden. In aller Regel werden wir dies deshalb im folgenden auch tun.

Leider gibt es wichtige Maßräume $(\Omega, \mathcal{A}, \mu)$, die *nicht* diskret sind, und bei denen eine sinnvolle Fortsetzung von μ auf die gesamte Potenzmenge $\mathfrak{P}(\Omega)$ auch nicht gelingt. Dies ist etwa bei den Borel'schen Maßräumen der Fall, bei denen das Maß μ die anschaulichen Begriffe "Länge", "Fläche" oder "Volumen" widerspiegelt. Darauf kommen wir in einem gesonderten Kapitel später noch einmal zurück.

Ein für solche Situationen angemessener Fortsetzungssatz ist erheblich komplizierter und soll daher hier auch nur mit seiner Beweisidee vorgestellt werden. Vorher benötigen wir aber noch einen neuen Strukturbegriff.

Definition 8 (Semi-Ring). Es sei Ω eine nicht-leere Menge. Ein Mengensystem $\mathcal{S} \subseteq \mathfrak{P}(\Omega)$ heißt *Semi-Ring* (über Ω), wenn die folgenden drei definierenden Eigenschaften erfüllt sind:

$$\begin{aligned} \emptyset &\in \mathcal{S} \\ S, T \in \mathcal{S} &\Rightarrow S \cap T \in \mathcal{S} \\ S, T \in \mathcal{S} &\Rightarrow \exists n \in \mathbb{N}, S_1, \dots, S_n \in \mathcal{S} \text{ paarweise disjunkt mit } S \setminus T = \bigoplus_{i=1}^n S_i. \end{aligned}$$

Die dritte obige Bedingung besagt also, dass in einem Semi-Ring die Differenz zweier Mengen stets als endliche disjunkte Vereinigung anderer Mengen des Semi-Rings dargestellt werden kann.

Das folgende Lemma zeigt, dass in einem Semi-Ring auch jede *abzählbare* Vereinigung von Mengen als disjunkte Vereinigung geschrieben werden kann.

Lemma 7. Es sei $\mathcal{S} \subseteq \mathfrak{P}(\Omega)$ ein Semi-Ring über einer nicht-leeren Menge Ω und $S_1, \dots, S_n \in \mathcal{S}$ für ein $n \in \mathbb{N}$. Dann existieren endlich viele Mengen $T_1, \dots, T_m \in \mathcal{S}$ mit $m \in \mathbb{N}$ derart, dass

$$\bigcup_{i=1}^n S_i = \bigoplus_{j=1}^m T_j$$

gilt. Ist allgemeiner $\{S_n\}_{n \in \mathbb{N}} \subseteq \mathcal{S}$ eine Folge von Mengen, so existiert eine weitere Folge $\{T_n\}_{n \in \mathbb{N}} \subseteq \mathcal{S}$ paarweise disjunkter Mengen mit

I. Maßtheoretische Grundlagen

$$\bigcup_{n=1}^{\infty} S_n = \bigoplus_{n=1}^{\infty} T_n.$$

Beweis: Wir benötigen zunächst das folgende Hilfsresultat:

Sind $A, B_1, \dots, B_n \in \mathcal{S}$ mit $n \in \mathbb{N}$, so gibt es ein $r_n \in \mathbb{N}$ und paarweise disjunkte Mengen $C_1, \dots, C_{r_n} \in \mathcal{S}$ mit $A \setminus \bigcup_{i=1}^n B_i = \bigoplus_{i=1}^{r_n} C_i$. Dies zeigen wir mit vollständiger Induktion:

Für $n=1$ entspricht diese Aussage der Definition eines Semi-Rings. Sei nun die Aussage wahr für ein $n \in \mathbb{N}$, dann ist

$$A \setminus \bigcup_{i=1}^{n+1} B_i = \left(A \setminus \bigcup_{i=1}^n B_i \right) \setminus B_{n+1} = \left(\bigoplus_{i=1}^{r_n} C_i \right) \setminus B_{n+1} = \bigoplus_{i=1}^{r_n} (C_i \setminus B_{n+1}).$$

In erneuter Anwendung der Definition erhalten wir nun für jedes $i \in \{1, 2, \dots, r_n\}$ jeweils paarweise disjunkte Mengen $D_i^1, \dots, D_i^{m_i} \in \mathcal{S}$ mit $C_i \setminus B_{n+1} = \bigoplus_{\ell_i=1}^{m_i} D_i^{\ell_i}$, so dass

$$A \setminus \bigcup_{i=1}^{n+1} B_i = \bigoplus_{i=1}^{r_n} (C_i \setminus B_{n+1}) = \bigoplus_{i=1}^{r_n} \bigoplus_{\ell_i=1}^{m_i} D_i^{\ell_i}.$$

Durch Zusammenfassen der Terme und Ummummerierung ergibt sich jetzt erneut eine Darstellung für $A \setminus \bigcup_{i=1}^{n+1} B_i$ als Vereinigung disjunkter Mengen des Semi-Rings, womit das Hilfsresultat bewiesen ist. Sei nun

$$V_n := \bigcup_{i=1}^n S_i, \quad n \in \mathbb{N}.$$

Dann existiert eine monoton wachsende Folge $\{k_n\}_{n \in \mathbb{N}} \subseteq \mathbb{N}$ von Indizes und eine Folge $\{T_n\}_{n \in \mathbb{N}} \subseteq \mathcal{S}$ paarweise disjunkter Mengen, so dass

$$V_n = \bigoplus_{i=1}^{k_n} T_i, \quad n \in \mathbb{N}.$$

Dies zeigen wir wieder mit vollständiger Induktion: Der Fall $n=1$ ist trivial. Gelte also die letzte Behauptung für ein $n \in \mathbb{N}$. Dann folgt

$$V_{n+1} := \bigcup_{i=1}^{n+1} S_i = V_n \oplus (S_{n+1} \setminus V_n) = V_n \oplus \left(S_{n+1} \setminus \bigcup_{i=1}^n S_i \right), \quad n \in \mathbb{N}.$$

Nach dem obigen Hilfsresultat besitzt $S_{n+1} \setminus \bigcup_{i=1}^n S_i$ eine Darstellung als Vereinigung paarweise disjunkter Mengen des Semi-Rings, etwa

I. Maßtheoretische Grundlagen

$$S_{n+1} \setminus \bigcup_{i=1}^n S_i = \bigoplus_{i=1}^r \tilde{T}_i$$

mit $r \in \mathbb{N}$ und geeigneten Mengen $\tilde{T}_1, \dots, \tilde{T}_r \in \mathcal{S}$. Mit $k_{n+1} = k_n + r$ und $T_{k_n+i} := \tilde{T}_i$, $i = 1, \dots, r$ folgt also

$$V_{n+1} = \bigoplus_{i=1}^{k_{n+1}} T_i.$$

Schließlich ergibt sich noch

$$\bigcup_{n=1}^{\infty} S_n = \bigcup_{n=1}^{\infty} V_n = \bigcup_{n=1}^{\infty} \bigoplus_{i=1}^n T_i = \bigoplus_{n=1}^{\infty} T_n,$$

womit das Lemma bewiesen ist. ■

Definition 9. Es sei (Ω, \mathcal{A}) ein beliebiger Messraum. Eine Folge $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ heißt Ω *abzählbar überdeckend*, wenn gilt:

$$\Omega = \bigcup_{n=1}^{\infty} A_n.$$

Eine Mengenfunktion ν auf einem beliebigen Teilsystem $\mathcal{U} \subseteq \mathcal{A}$ heißt (dort) σ -*endlich*, wenn es eine Ω abzählbar überdeckende Folge $\{U_n\}_{n \in \mathbb{N}} \subseteq \mathcal{U}$ gibt mit der Eigenschaft

$$\forall n \in \mathbb{N}: \nu(U_n) < \infty.$$

Satz 7 (großer Maßfortsetzungssatz). Es sei (Ω, \mathcal{A}) ein beliebiger Messraum und $\mathcal{S} \subseteq \mathfrak{P}(\Omega)$ ein \mathcal{A} erzeugender Semi-Ring über Ω , d.h. es gelte $\mathcal{A} = \sigma(\mathcal{S})$. Ferner sei $\mu: \mathcal{S} \rightarrow \overline{\mathbb{R}}$ ein Prämaß, d.h. μ ist normiert, nicht-negativ und σ -additiv auf \mathcal{S} . Dann gibt es ein (nicht notwendig eindeutig bestimmtes) Maß $\tilde{\mu}$ auf $\mathcal{A} = \sigma(\mathcal{S})$ mit $\tilde{\mu}(S) = \mu(S)$ für alle $S \in \mathcal{S}$, d.h. $\tilde{\mu}$ ist Fortsetzung von μ auf \mathcal{A} . Die Fortsetzung $\tilde{\mu}$ von μ auf \mathcal{A} ist eindeutig, wenn μ auf \mathcal{S} σ -endlich ist.

Bevor wir die **Beweisidee** skizzieren, wollen wir kurz zeigen, dass Satz 7 den Satz 6 impliziert. Dazu ist nur zu beachten, dass das System $\mathcal{S} := \{\emptyset\} \cup \mathcal{T}$ trivialerweise ein Semi-Ring ist, weil die Atome in \mathcal{T} paarweise disjunkt sind. Wegen $\Omega = \bigoplus_{i \in I} T_i$ ist \mathcal{S} darüber hinaus auch Ω abzählbar überdeckend. Mit der Zuordnung $\mu(T_i) := \mu_i$, $i \in I$ ist μ ferner σ -additiv auf \mathcal{S} , da hier die dritte Bedingung in Satz 7, $\bigoplus_{n=1}^{\infty} S_n \in \mathcal{S}$, nur erfüllt ist, wenn alle bis auf eine der Mengen S_n leer sind.

Wir kommen jetzt zur Beweisidee von Satz 7. Sie geht in der hier vorgestellten Form zurück auf den in Berlin geborenen griechischen Ingenieur und Mathematiker Constantin Carathéodory (1873 – 1950; er war in der Zeit von 1913 bis 1918 Nachfolger von Felix Klein in Göttingen und nach mehreren Zwischenstationen zuletzt Professor an der Universität München). Der Beweis, der auch

I. Maßtheoretische Grundlagen

heute noch allgemein so verwendet wird, stammt aus seinen „Vorlesungen über reelle Funktionen“ (1918); siehe auch BAUER (1992).

In der Situation von Satz 7 können wir μ zunächst problemlos auf den von \mathcal{S} erzeugten Ring \mathcal{R} fortsetzen, das ist das kleinste Teilmengensystem von Ω , welches \mathcal{S} enthält, durchschnittstabil (äquivalent: vereinigungsstabil) und differenzstabil ist³. Dieser Ring besteht nämlich gerade aus allen Mengen, die durch endliche Vereinigungsbildung von Mengen des Semi-Rings \mathcal{S} entstehen (Ring-Eigenschaften einfach nachrechnen). Nach Lemma 7 kann eine solche Menge aber immer auch als disjunkte Vereinigung geschrieben werden, so dass gilt:

$$R \in \mathcal{R} \Leftrightarrow \exists n \in \mathbb{N}, S_1, \dots, S_n \in \mathcal{S} \text{ paarweise disjunkt mit } R = \bigoplus_{i=1}^n S_i.$$

Es ist dann nahe liegend, die Fortsetzung⁴ von μ auf \mathcal{R} für solche R folgendermaßen zu definieren:

$$\mu(R) = \mu\left(\bigoplus_{i=1}^n S_i\right) := \sum_{i=1}^n \mu(S_i).$$

Diese Fortsetzung ist wohldefiniert, denn sei etwa $R = \bigoplus_{i=1}^n S_i = \bigoplus_{j=1}^k T_j$ mit $S_1, \dots, S_n, T_1, \dots, T_k \in \mathcal{S}$, so gilt

$$\begin{aligned} \sum_{i=1}^n \mu(S_i) &= \sum_{i=1}^n \mu(R \cap S_i) = \sum_{i=1}^n \mu\left(\bigoplus_{j=1}^k \{T_j \cap S_i\}\right) = \sum_{i=1}^n \sum_{j=1}^k \mu(S_i \cap T_j) \\ &= \sum_{j=1}^k \mu\left(\bigoplus_{i=1}^n \{S_i \cap T_j\}\right) = \sum_{j=1}^k \mu(R \cap T_j) = \sum_{j=1}^k \mu(T_j). \end{aligned}$$

Ähnlich wie oben bezeichne jetzt für jede Teilmenge $Q \subseteq \Omega$ das (Überdeckungs-)System $\mathcal{U}(Q)$ die Menge aller Folgen $\{R_n\}_{n \in \mathbb{N}}$ von Mengen aus dem Ring \mathcal{R} mit $Q \subseteq \bigcup_{n=1}^{\infty} R_n$. Die durch

$$\mu^*(Q) := \begin{cases} \inf \left\{ \sum_{n=1}^{\infty} \mu(R_n) \mid \{R_n\}_{n \in \mathbb{N}} \in \mathcal{U}(Q) \right\}, & \text{falls } \mathcal{U}(Q) \neq \emptyset \\ \infty, & \text{sonst} \end{cases}$$

auf $\mathfrak{P}(\Omega)$ definierte Mengenfunktion heißt das zu μ gehörige *äußere* Maß (der Menge Q). Dieses ist i. Allg. nicht σ -additiv (auf $\mathfrak{P}(\Omega)$), also dort kein Maß im eigentlichen Sinne. Es gilt aber

$$\mu^*(R) = \mu(R) \quad \text{für alle } R \in \mathcal{R},$$

wie man sofort sieht, weil ja in jedem Fall die Folge $\{R_n\}_{n \in \mathbb{N}}$ mit $R_1 := R$ und $R_n := \emptyset$ für $n > 1$ in $\mathcal{U}(R)$ liegt und damit wegen der Prämaß-Eigenschaft von μ das den Wert von $\mu^*(R)$ bestimmende Infimum auch für diese Folge angenommen wird.

Der wesentliche Beweisschritt ist nun der, dass man zeigen kann, dass das System

$$\mathcal{A}^* := \left\{ A \subseteq \Omega \mid \mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \setminus A) \text{ für alle } Q \subseteq \Omega \right\}$$

³ Ein Ring enthält definitionsgemäß auch immer die leere Menge, die hier aber schon in \mathcal{S} enthalten ist.

⁴ Der Einfachheit halber bezeichnen wir auch diese Fortsetzung von μ mit demselben Symbol.

I. Maßtheoretische Grundlagen

eine σ -Algebra bildet, die die von \mathcal{R} (bzw. äquivalent die von \mathcal{S}) erzeugte σ -Algebra $\mathcal{A} = \sigma(\mathcal{R}) = \sigma(\mathcal{S})$ umfasst, und dass nun die Einschränkung $\tilde{\mu}$ von μ^* auf \mathcal{A}^* (und damit insbesondere auch auf \mathcal{A}) tatsächlich ein Maß ist, welches also eine gewünschte Fortsetzung von dem ursprünglich nur auf \mathcal{S} definierten μ ist⁵.

Die für die Eindeutigkeit gemachten Zusatzannahmen in Satz 7 sind dabei unverzichtbar, denn auf dem trivialen (Semi-)Ring $\mathcal{S} = \mathcal{R} = \{\emptyset\}$ gilt stets $\mu(S) = 0$ für alle $S \in \mathcal{S}$, wohingegen die Werte $\mu(\Omega)$ beliebig nicht-negativ gewählt werden können (beachte: $\sigma(\mathcal{S}) = \sigma(\mathcal{R}) = \{\emptyset, \Omega\}$).

Die Kernaussagen der Sätze 6 und 7 kann man insgesamt folgendermaßen zusammenfassen:

Gibt es in einem beliebigen Messraum (Ω, \mathcal{A}) einen \mathcal{A} erzeugenden Semi-Ring \mathcal{S} , auf dem eine normierte, nicht-negative, σ -additive und σ -endliche Mengenfunktion μ definiert ist, so kann μ in eindeutiger Weise zu einem Maß auf \mathcal{A} fortgesetzt werden.

Dieser Satz ist fundamental für die gesamte Maßtheorie und wird vor allem zur Konstruktion von Produktmaßen, die in korrespondierenden Wahrscheinlichkeitsmodellen, die vor allem die für statistische Anwendungen besonders wichtige *stochastische Unabhängigkeit* beschreiben, benötigt. Wir wollen uns deshalb im nächsten Abschnitt mit solchen Maßen genauer beschäftigen.

I. 4. Produktmaße

Zu dem Begriff des Produktmaßes gelangt man anschaulich schon dann, wenn man den Längenbegriff, der dem später behandelten Lebesgue-Maß für den euklidischen Raum \mathbb{R}^1 entspricht, auf höhere Dimensionen ausdehnen will (Flächeninhalt im \mathbb{R}^2 , Volumen im \mathbb{R}^3). Bereits in der Grundschule lernen wir, dass die Fläche eines Rechtecks zu berechnen ist als das Produkt aus "Länge" und "Breite", also aus zwei jeweils für sich genommen eindimensionalen Größen. Das Konzept des Produktmaßes greift genau dieses Prinzip in verallgemeinerter Form auf.

Satz 8. Es seien $(\Omega_i, \mathcal{A}_i, \mu_i)$, $i \in I$ Maßräume mit einer endlichen Indexmenge I . $\mathcal{A} = \bigotimes_{i \in I} \mathcal{A}_i$ bezeichne wie üblich die Produkt- σ -Algebra der σ -Algebren \mathcal{A}_i , $i \in I$ über dem kartesischen Produkt $\Omega = \prod_{i \in I} \Omega_i$. Die Maße μ_i , $i \in I$ seien sämtlich σ -endlich. Dann gelten folgende Aussagen:

- a) Das System $\mathcal{S} := \left\{ \prod_{i \in I} A_i \mid \forall i \in I: A_i \in \mathcal{A}_i \right\}$ bildet einen Ω abzählbar überdeckenden Semi-Ring.
- b) Es existiert ein eindeutig bestimmtes Maß μ auf \mathcal{A} mit der Eigenschaft:

$$\forall A_i \in \mathcal{A}_i, i \in I: \mu\left(\prod_{i \in I} A_i\right) = \prod_{i \in I} \mu_i(A_i).$$

Beweis: Wir zeigen zunächst, dass $\mathcal{S} = \left\{ \prod_{i \in I} A_i \mid \forall i \in I: A_i \in \mathcal{A}_i \right\}$ einen Semi-Ring bildet, indem wir die drei Eigenschaften aus Definition 8 nachweisen:

⁵ Da diese größere σ -Algebra \mathcal{A}^* aber von μ^* bzw. μ bzw. deren Nullmengen abhängt, formuliert man den Fortsetzungssatz in der Regel nur für $\mathcal{A} = \sigma(\mathcal{R}) = \sigma(\mathcal{S})$.

I. Maßtheoretische Grundlagen

- Es ist $\emptyset = \prod_{i \in I} \emptyset$, also gilt $\emptyset \in \mathcal{S}$.
- Seien $S, T \in \mathcal{S}$ mit $S = \prod_{i \in I} A_i$, $T = \prod_{i \in I} B_i$. Dann ist nach Lemma 1 auch $S \cap T = \prod_{i \in I} (A_i \cap B_i) \in \mathcal{S}$.
- Seien $S, T \in \mathcal{S}$ mit $S = \prod_{i \in I} A_i$, $T = \prod_{i \in I} B_i$. Wir definieren $C_{i1} := B_i$, $C_{i2} := B_i^c$, $i \in I$ sowie eine Indexmenge $K := \left\{ (j_k)_{k \in I} \in \{1, 2\}^I \right\} \setminus \mathbf{1}$, wobei hier $\mathbf{1}$ dasjenige Element aus $\{1, 2\}^I$ bezeichne, das aus der identischen Wiederholung der Eins besteht. (K enthält also genau $2^{\#(I)} - 1$ Elemente.) Es folgt nun:

$$T^c = \left(\prod_{i \in I} B_i \right)^c = \bigoplus_{(j_k)_{k \in I} \in K} \prod_{i \in I} C_{i, j_i},$$

d.h. T^c besteht genau aus der paarweise disjunkten Vereinigung aller kartesischen Produkte, die aus den Mengen B_i oder deren Komplementen gebildet werden können, bis auf die Menge $\prod_{i \in I} B_i$ selbst. Damit ist aber die Mengendifferenz $S \setminus T$ wegen

$$\begin{aligned} S \setminus T &= \left(\prod_{i \in I} A_i \right) \setminus \left(\prod_{i \in I} B_i \right) = \left(\prod_{i \in I} A_i \right) \cap \left(\bigoplus_{(j_k)_{k \in I} \in K} \prod_{i \in I} C_{i, j_i} \right) = \bigoplus_{(j_k)_{k \in I} \in K} \left\{ \left(\prod_{i \in I} A_i \right) \cap \left(\prod_{i \in I} C_{i, j_i} \right) \right\} \\ &= \bigoplus_{(j_k)_{k \in I} \in K} \underbrace{\prod_{i \in I} (A_i \cap C_{i, j_i})}_{\in \mathcal{S}} \end{aligned}$$

darstellbar als disjunkte Vereinigung von Mengen aus \mathcal{S} , da jede der involvierten σ -Algebren \mathcal{A}_i , $i \in I$ komplement- und schnitts stabil ist.

Also ist \mathcal{S} ein Semi-Ring über Ω , wie behauptet. \mathcal{S} ist trivialerweise Ω abzählbar überdeckend, weil bereits $\Omega = \prod_{i \in I} \Omega_i \in \mathcal{S}$ gilt und damit die Folge $\{S_n\}_{n \in \mathbb{N}} \subseteq \mathcal{S}$ mit $S_n := \Omega$ für alle $n \in \mathbb{N}$ das Gewünschte leistet. Wir definieren nun eine Mengenfunktion μ auf \mathcal{S} vermöge

$$\forall A_i \in \mathcal{A}_i, i \in I: \mu \left(\prod_{i \in I} A_i \right) := \prod_{i \in I} \mu_i(A_i).$$

Diese besitzt trivialerweise die Eigenschaften $\mu(\emptyset) = \mu \left(\prod_{i \in I} \emptyset \right) = \prod_{i \in I} \mu_i(\emptyset) = 0$ und $\mu(S) \geq 0$ für alle $S \in \mathcal{S}$. Mit Mitteln, die uns hier nicht in voller Allgemeinheit zur Verfügung stehen, lässt sich weiter zeigen, dass die σ -Additivität jedes einzelnen Maßes μ_i , $i \in I$ auch die σ -Additivität von μ auf \mathcal{S} impliziert.⁶ Ferner ist μ im Sinne von Definition 9 auch σ -endlich, denn nach Voraussetzung ist jedes einzelne Maß μ_i σ -endlich, d.h. für jedes $i \in I$ existiert eine Ω_i abzählbar überdeckende Folge $\{A_{in}\}_{n \in \mathbb{N}} \subseteq \mathcal{A}_i$ mit der Eigenschaft $\mu_i(A_{in}) < \infty$ für alle $n \in \mathbb{N}$. Dann ist aber auch die Mengenfolge $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ mit $A_n := \prod_{i \in I} A_{in}$ für alle $n \in \mathbb{N}$ die Gesamtmenge $\Omega = \prod_{i \in I} \Omega_i$ abzählbar überdeckend mit $\mu(A_n) = \mu \left(\prod_{i \in I} A_{in} \right) = \prod_{i \in I} \mu_i(A_{in}) < \infty$, womit aufgrund von Satz 7 alles gezeigt ist. ■

⁶ Dies werden wir im Spezialfall von Borel-Maßen später mit Methoden der Topologie explizit zeigen.

I. Maßtheoretische Grundlagen

Definition 10 (Produktmaß). Das unter den Voraussetzungen von Satz 8 eindeutig bestimmte Maß μ heißt *Produktmaß* der Maße $\mu_i, i \in I$ und wird mit dem Symbol

$$\mu = \bigotimes_{i \in I} \mu_i$$

bezeichnet.

Ein einfaches Beispiel für ein "natürliches" Produktmaß ist das abzählende Maß $\#$ auf dem endlichen kartesischen Produkt abzählbarer Mengen; damit ist auf dem endlichen kartesischen Produkt *endlicher* Mengen auch die diskrete Gleichverteilung aus Definition 7 ein Produktmaß.

Überhaupt vereinfacht sich die Struktur von Produktmaßen im Hinblick auf die Sätze 4 und 6 bei diskreten Grundmengen ganz erheblich. Sind die Maßräume $(\Omega_i, \mathcal{A}_i, \mu_i), i \in I$ in Satz 8 nämlich diskret (d.h. alle Grundmengen $\Omega_i, i \in I$ sind höchstens abzählbar), so existiert ja nach den Sätzen 2 und 4 je ein eindeutig bestimmtes, erschöpfendes, vollständiges, abzählbares Atomsystem $\mathcal{T}_i = \{T_{ij} \mid j \in J_i\}$ für jede der σ -Algebren $\mathcal{A}_i, i \in I$ und damit nach Satz 6 auch ein entsprechendes Atomsystem $\mathcal{T} := \left\{ \times_{i \in I} T_i \mid \forall i \in I: T_i \in \mathcal{T}_i \right\} = \{S_k \mid k \in K\}$ für die Produkt- σ -Algebra \mathcal{A} , wobei für die Indexmenge K gilt: $K \sim_{\#} \times_{i \in I} J_i$. Damit kann man das Produktmaß μ relativ leicht explizit über die Werte der Maße μ_i für die jeweiligen Atome angeben, wenn man die folgenden Beziehungen beachtet:

$$\begin{aligned} \mu\left(\bigoplus_{k \in M} S_k\right) &= \sum_{k \in M} \mu(S_k) \text{ für alle } M \subseteq K \text{ mit} \\ \mu(S_k) &= \mu\left(\times_{i \in I} T_i\right) = \prod_{i \in I} \mu_i(T_i) \text{ für eine Darstellung } S_k = \times_{i \in I} T_i \in \mathcal{T}. \end{aligned}$$

Für den Fall, dass die Maße μ_i bereits auf ganz $\mathfrak{P}(\Omega_i)$ gegeben sind (was ja nach geeigneter Fortsetzung im diskreten Rahmen stets möglich ist), ist das Produktmaß einfacher direkt gegeben durch

$$\mu(A) = \sum_{(\omega_i)_{i \in I} \in A} \prod_{i \in I} \mu_i(\{\omega_i\}), \quad A \in \mathfrak{P}(\Omega).$$

Es gibt natürlich auch Maße auf Produkt- σ -Algebren, die keine Produktmaße sind. Dazu wählen wir beispielhaft $\Omega_1 = \Omega_2 = \{1, 2\}$ mit den jeweiligen Potenzmengen als zugehörigen σ -Algebren \mathcal{A}_1 und \mathcal{A}_2 (dann ist die Potenzmenge von $\Omega_1 \times \Omega_2$ die Produkt- σ -Algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$) und setzen

$$\mu(\{(\omega_1, \omega_2)\}) := \omega_1 + \omega_2 \text{ für alle } \omega = (\omega_1, \omega_2) \in \Omega = \Omega_1 \times \Omega_2.$$

Angenommen, es gäbe eine Darstellung $\mu = \mu_1 \otimes \mu_2$ mit zwei (notwendig endlichen) Maßen μ_1 auf \mathcal{A}_1 und μ_2 auf \mathcal{A}_2 . Zur Abkürzung setzen wir $x_i := \mu_1(\{i\})$ und $y_j := \mu_2(\{j\})$ für $i \in \Omega_1, j \in \Omega_2$. Nach Definition des Maßes μ sind die x_i bzw. y_j strikt positiv, und es ergibt sich das Gleichungssystem

$$x_i \cdot y_j = i + j \text{ für } i, j \in \{1, 2\}.$$

I. Maßtheoretische Grundlagen

Dieses Gleichungssystem ist aber nicht lösbar, denn es gilt speziell

$$\begin{aligned}x_1 \cdot y_1 = 2, \quad x_2 \cdot y_1 = 3 &\Rightarrow \frac{x_1}{x_2} = \frac{2}{3} \\x_1 \cdot y_2 = 3, \quad x_2 \cdot y_2 = 4 &\Rightarrow \frac{x_1}{x_2} = \frac{3}{4} \quad \text{Widerspruch!}\end{aligned}$$

In diesem Zusammenhang ist noch die allgemeine Frage interessant, ob für den Fall, dass $\mu = \mu_1 \otimes \mu_2$ tatsächlich ein Produktmaß auf einer beliebigen Produkt- σ -Algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ ist, die einzelnen „Faktoren“ μ_1 auf \mathcal{A}_1 und μ_2 auf \mathcal{A}_2 eindeutig bestimmt sind. Im Allgemeinen ist das nicht richtig, denn für jedes reelle $c > 0$ führen die Maße $\nu_1 = c \cdot \mu_1$ und $\nu_2 = \frac{1}{c} \cdot \mu_2$ zum gleichen Produktmaß $\nu_1 \otimes \nu_2 = \mu_1 \otimes \mu_2$. Nur im Fall von Wahrscheinlichkeitsmaßen liegt Eindeutigkeit vor, weil dann zwingend $c = 1$ gelten muss (wegen $c = 1/c$).

Liegt ein nicht-triviales Produktmaß $\mu = \mu_1 \otimes \mu_2$ auf der Produkt- σ -Algebra $\mathcal{A}_1 \otimes \mathcal{A}_2$ vor, lassen sich geeignete (dann ebenfalls nicht-triviale) „Faktoren“ μ_1 auf \mathcal{A}_1 und μ_2 auf \mathcal{A}_2 bestimmen durch die Beziehungen

$$\mu(A_1 \times \Omega_2) = \mu_1(A_1) \cdot \mu_2(\Omega_2) \quad \text{und} \quad \mu(\Omega_1 \times A_2) = \mu_1(\Omega_1) \cdot \mu_2(A_2) \quad \text{für alle } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2.$$

Unter Beachtung der vorangehenden Bemerkung erhält man also alle Faktoren durch die Wahl

$$\mu_1(A_1) = c_1 \cdot \mu(A_1 \times \Omega_2) \quad \text{und} \quad \mu_2(A_2) = c_2 \cdot \mu(\Omega_1 \times A_2) \quad \text{für alle } A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$$

zurück, solange die Bedingung

$$c_1 \cdot c_2 \cdot \mu(\Omega_1 \times \Omega_2) = 1$$

bestehen bleibt (wegen $\mu(\Omega_1 \times \Omega_2) = \mu_1(\Omega_1) \cdot \mu_2(\Omega_2) = c_1 \cdot c_2 \cdot \{\mu(\Omega_1 \times \Omega_2)\}^2$). Für Wahrscheinlichkeitsmaße ist diese Bedingung gemäß den obigen Ausführungen immer erfüllt. Ähnlich kann man auch im höherdimensionalen Fall argumentieren.

Für gewisse Anwendungen in der Stochastik reichen Produktmaße mit endlich vielen Komponenten oft nicht aus, vor allem dann, wenn Grenzwertsätze wie das eingangs schon erwähnte Gesetz der großen Zahlen oder Stochastische Prozesse betrachtet werden. Aufgrund der speziellen Normiertheit von Wahrscheinlichkeitsmaßen kann man Produkt-Wahrscheinlichkeitsmaße aber sogar noch für beliebige Indexmengen konsistent definieren (vgl. z.B. ELSTRODT (1996) und BAUER (1992)). Dafür benötigt man allerdings auch eine geeignete Erweiterung des Begriffs der Produkt- σ -Algebra für beliebig viele Komponenten. Ohne einen geeigneten Abbildungsbegriff, der im nachfolgenden Abschnitt behandelt wird, ist dies jedoch nicht einfach zu leisten.

I. 5. Messbare Abbildungen

In der Analysis und Linearen Algebra spielen geeignete Abbildungen zwischen den dort zu Grunde liegenden Strukturen (z.B. metrische Räume, Vektorräume) eine wichtige Rolle, wobei Begriffe wie Stetigkeit oder Linearität von Bedeutung sind. In der Maßtheorie sind die entsprechenden Objekte die *messbaren Abbildungen*, wobei sich hier ganz ähnliche Eigenschaften wie in den genannten Teilgebieten der Mathematik zeigen.

Definition 11 (messbare Abbildung). Es seien $(\Omega_i, \mathcal{A}_i)$, $i=1,2$ Messräume. Eine Abbildung $T: \Omega_1 \rightarrow \Omega_2$ heißt $(\mathcal{A}_1$ - $\mathcal{A}_2)$ -*messbar* (oder, wenn keine Verwechslungen möglich sind, kurz auch nur *messbar*), falls die Urbildabbildung T^{-1} die Eigenschaft

$$\forall A_2 \in \mathcal{A}_2 : T^{-1}(A_2) := \{\omega_1 \in \Omega_1 \mid T(\omega_1) \in A_2\} \in \mathcal{A}_1$$

besitzt.

Man kann diese Eigenschaft kurz und prägnant auch so ausdrücken: eine Abbildung T ist genau dann messbar, wenn die Urbilder messbarer Mengen messbar sind. Damit ähnelt diese Definition der Definition stetiger Abbildungen zwischen topologischen Räumen⁷, die völlig analog besagt: eine Abbildung T ist genau dann stetig, wenn die Urbilder offener Mengen offen sind. Der enge Zusammenhang zwischen Topologie und Maßtheorie wird besonders deutlich in dem anschließenden Abschnitt über die Borel'sche σ -Algebra (und das Lebesgue-Maß).

Die in Definition 11 genannten Anforderungen an die Messbarkeit sind nicht so substantiell, wie es auf den ersten Blick erscheinen mag. Zu vorgegebenen Grundmengen Ω_1 und Ω_2 mit einer beliebigen Abbildung $T: \Omega_1 \rightarrow \Omega_2$ existieren nämlich immer geeignete σ -Algebren \mathcal{A}_1 über Ω_1 bzw. \mathcal{A}_2 über Ω_2 , so dass T messbar ist, wenn nur auf Ω_2 bzw. auf Ω_1 eine σ -Algebra vorgegeben ist. Dies liegt an folgendem

Lemma 8. Es seien Ω_1 und Ω_2 nicht-leere Mengen und $T: \Omega_1 \rightarrow \Omega_2$ eine beliebige Abbildung.

a) Es sei \mathcal{A}_2 eine σ -Algebra über Ω_2 . Dann ist das System

$$\mathcal{A}_1 := T^{-1}(\mathcal{A}_2) := \{T^{-1}(A_2) \mid A_2 \in \mathcal{A}_2\}$$

aller Urbilder messbarer Mengen eine σ -Algebra über Ω_1 , so dass T $(\mathcal{A}_1$ - $\mathcal{A}_2)$ -messbar ist.

b) Es sei \mathcal{A}_1 eine σ -Algebra über Ω_1 . Dann ist das System

$$\mathcal{A}_2 := \{A_2 \in \mathfrak{P}(\Omega_2) \mid T^{-1}(A_2) \in \mathcal{A}_1\}$$

⁷ Eine Topologie über einer nicht-leeren Menge ist durch das System der sog. *offenen Mengen* gegeben, die ähnliche Eigenschaften wie messbare Mengen besitzen. Eine Topologie enthält stets die leere sowie die Grundmenge und ist abgeschlossen gegenüber endlicher Durchschnitts- und beliebiger (auch überabzählbarer!) Vereinigungsbildung. Die Komplemente offener Mengen heißen *abgeschlossen* und sind i.d.R. nicht mehr offen.

I. Maßtheoretische Grundlagen

eine σ -Algebra über Ω_2 , so dass T (\mathcal{A}_1 - \mathcal{A}_2)-messbar ist.

Beweis:

a) Die Eigenschaft einer σ -Algebra ergibt sich aus folgenden drei Überlegungen:

- Es ist $T^{-1}(\emptyset) = \emptyset$, also gilt $\emptyset \in T^{-1}(\mathcal{A}_2)$.
- Für $A_1 \in T^{-1}(\mathcal{A}_2)$ mit der Darstellung $A_1 = T^{-1}(A_2)$, $A_2 \in \mathcal{A}_2$ ist

$$A_1^c = (T^{-1}(A_2))^c = T^{-1}(A_2^c) \in T^{-1}(\mathcal{A}_2),$$

d.h. $T^{-1}(\mathcal{A}_2)$ ist komplementstabil.

- Für $\{A_{1n}\}_{n \in \mathbb{N}} \subseteq T^{-1}(\mathcal{A}_2)$ mit den Darstellungen $A_{1n} = T^{-1}(A_{2n})$, $A_{2n} \in \mathcal{A}_2$ ist

$$\bigcup_{n=1}^{\infty} A_{1n} = \bigcup_{n=1}^{\infty} T^{-1}(A_{2n}) = T^{-1}\left(\bigcup_{n=1}^{\infty} A_{2n}\right) \in T^{-1}(\mathcal{A}_2),$$

d.h. $T^{-1}(\mathcal{A}_2)$ ist abzählbar vereinigungsstabil.

Damit ist $\mathcal{A}_1 = T^{-1}(\mathcal{A}_2)$ eine σ -Algebra über Ω_1 mit der gewünschten Eigenschaft.

b) Die Eigenschaft einer σ -Algebra ergibt sich wieder aus folgenden drei Überlegungen:

- Es ist $T^{-1}(\emptyset) = \emptyset \in \mathcal{A}_1$, also gilt $\emptyset \in \mathcal{A}_2$.
- Für $A_2 \in \mathcal{A}_2$ ist $T^{-1}(A_2) \in \mathcal{A}_1$, also gilt auch $T^{-1}(A_2^c) = (T^{-1}(A_2))^c \in \mathcal{A}_1$, so dass auch $A_2^c \in \mathcal{A}_2$; demnach ist \mathcal{A}_2 komplementstabil.
- Für $\{A_{2n}\}_{n \in \mathbb{N}} \subseteq \mathcal{A}_2$ ist $T^{-1}(A_{2n}) \in \mathcal{A}_1$, also gilt auch

$$T^{-1}\left(\bigcup_{n=1}^{\infty} A_{2n}\right) = \bigcup_{n=1}^{\infty} T^{-1}(A_{2n}) \in \mathcal{A}_1,$$

d.h. \mathcal{A}_2 ist abzählbar vereinigungsstabil.

Damit ist \mathcal{A}_2 eine σ -Algebra über Ω_2 mit der gewünschten Eigenschaft. ■

Weil jede σ -Algebra über Ω_1 , bezüglich der T messbar ist, $T^{-1}(\mathcal{A}_2)$ als Teilsystem enthalten muss, ist $T^{-1}(\mathcal{A}_2)$ also zugleich die *kleinste* σ -Algebra mit dieser Eigenschaft. In Anlehnung an die Erzeugung von σ -Algebren (vgl. Lemma 3) nennen wir deshalb $T^{-1}(\mathcal{A}_2)$ auch *die von T erzeugte σ -Algebra über Ω_1* .

I. Maßtheoretische Grundlagen

Umgekehrt ist jede σ -Algebra über Ω_2 , bezüglich der T messbar ist, auch selbst Teilsystem von $\mathcal{A}_2 = \{A_2 \in \mathfrak{P}(\Omega_2) \mid T^{-1}(A_2) \in \mathcal{A}_1\}$, so dass \mathcal{A}_2 die *größte* σ -Algebra mit dieser Eigenschaft ist.

In der speziellen Situation von Messräumen mit geeigneten Atomsystemen erhalten wir noch das folgende nützliche Ergebnis.

Lemma 9. Es seien Ω_1 und Ω_2 nicht-leere Mengen, $T: \Omega_1 \rightarrow \Omega_2$ eine beliebige Abbildung und \mathcal{A}_2 eine σ -Algebra über Ω_2 mit einem vollständigen (bzw. erschöpfenden⁸), abzählbaren Atomsystem $\mathcal{T}_2 = \{T_{2i} \mid i \in I_2\}$ mit abzählbarer Indexmenge I_2 . Dann gilt:

$$\mathcal{T}_1 := T^{-1}(\mathcal{T}_2) \setminus \{\emptyset\} = \{T^{-1}(T_{2i}) \mid i \in I_2\} \setminus \{\emptyset\}$$

ist ein vollständiges (bzw. erschöpfendes), abzählbares Atomsystem für die σ -Algebra $\mathcal{A}_1 = T^{-1}(\mathcal{A}_2)$.

Beweis: Jede Menge $A_2 \in \mathcal{A}_2$ ist nach Satz 3 in eindeutiger Weise darstellbar als $A_2 = \bigoplus_{i \in J} T_{2i}$ mit geeigneter Indexmenge $J \subseteq I_2$. Damit ist $T^{-1}(A_2) = \bigoplus_{i \in J} T^{-1}(T_{2i})$, also jede Menge aus $T^{-1}(\mathcal{A}_2)$ darstellbar als disjunkte Vereinigung von Elementen aus $T^{-1}(\mathcal{T}_2)$. Allerdings kann für gewisse $i \in I$ das Urbild $T^{-1}(T_{2i})$ leer sein (was eintreten kann, wenn T nicht surjektiv ist). Also muss ggf. die leere Menge noch aus $T^{-1}(\mathcal{T}_2)$ entfernt werden. Damit ist $\mathcal{T}_1 = T^{-1}(\mathcal{T}_2) \setminus \{\emptyset\}$ ein abzählbares, erschöpfendes (bzw. vollständiges) Atomsystem für \mathcal{A}_1 , weil nach obiger Darstellung die kleinsten nicht-leeren Mengen in \mathcal{A}_1 von der Form $T^{-1}(T_{2i})$ mit $i \in I_2$ sind. ■

Lemma 10. Es seien $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2$ Messräume; \mathcal{E}_2 sei ein (beliebiger) Erzeuger von \mathcal{A}_2 . Eine Abbildung $T: \Omega_1 \rightarrow \Omega_2$ ist genau dann $(\mathcal{A}_1, \mathcal{A}_2)$ -messbar, wenn gilt:

$$T^{-1}(\mathcal{E}_2) = \{T^{-1}(E_2) \mid E_2 \in \mathcal{E}_2\} \subseteq \mathcal{A}_1.$$

Beweis: Aufgrund der Definition der Messbarkeit von Abbildungen bleibt nur zu zeigen, dass die angegebene Bedingung hinreichend für die Messbarkeit von T ist. Wir betrachten dazu das Mengensystem $\mathcal{B}_2 = \{A_2 \in \mathfrak{P}(\Omega_2) \mid T^{-1}(A_2) \in \mathcal{A}_1\}$, welches nach Voraussetzung den Erzeuger \mathcal{E}_2 enthält und nach Lemma 8 b) eine σ -Algebra über Ω_2 ist. Es folgt

$$\mathcal{A}_2 = \sigma(\mathcal{E}_2) \subseteq \sigma(\mathcal{B}_2) = \mathcal{B}_2, \text{ also } T^{-1}(\mathcal{A}_2) \subseteq T^{-1}(\mathcal{B}_2) \subseteq \mathcal{A}_1,$$

woraus die Behauptung folgt. ■

⁸ Dies ist nach Satz 3 hier äquivalent.

I. Maßtheoretische Grundlagen

Zur Vereinfachung der Schreibweise wollen wir im Folgenden die Messbarkeit der Abbildung T durch die Notation

$$T : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$$

Zum Ausdruck bringen.

In vielen Anwendungen müssen auch Verkettungen von (messbaren) Abbildungen betrachtet werden. Dass die so zusammengesetzten Abbildungen messbar bleiben, zeigt

Lemma 11. Es seien $(\Omega_i, \mathcal{A}_i)$, $i = 1, 2, 3$ Messräume sowie T_1 und T_2 messbare Abbildungen mit

$$T_1 : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2) \quad \text{und} \quad T_2 : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_3, \mathcal{A}_3).$$

Dann ist die Komposition $T := T_2 \circ T_1 : \Omega_1 \rightarrow \Omega_3$ $(\mathcal{A}_1, \mathcal{A}_3)$ -messbar.

Beweis: Es gilt $T^{-1}(A_3) = T_1^{-1}\left(\underbrace{T_2^{-1}(A_3)}_{\in \mathcal{A}_2}\right) \in \mathcal{A}_1$ für alle Mengen $A_3 \in \mathcal{A}_3$, womit die Behauptung bewiesen ist. ■

Gelegentlich sind auch mehrere Abbildungen $T_i : \Omega \rightarrow \Omega_i$ für $i \in I$ auf derselben Grundmenge Ω mit u.U. verschiedenen Bildmengen Ω_i , $i \in I$ zu betrachten. In dieser Situation möchte man eine geeignete σ -Algebra über Ω zur Verfügung haben, bezüglich der alle Abbildung gemeinsam messbar sind. Da nach Lemma 8 die Systeme $T_i^{-1}(\mathcal{A}_i)$, $i \in I$ alle jeweils σ -Algebren über Ω sind, ist die kleinste σ -Algebra über Ω , die dies leistet, gegeben durch

$$\sigma(T_i; i \in I) := \sigma\left(\bigcup_{i \in I} T_i^{-1}(\mathcal{A}_i)\right).$$

Diese σ -Algebra heißt auch die *von den Abbildungen T_i , $i \in I$ erzeugte σ -Algebra* über Ω . Man beachte dabei, dass i. Allg. die Vereinigung von σ -Algebren nicht wieder eine σ -Algebra ergibt.

Ein ähnliches Ergebnis enthält

Lemma 12. Es seien $T_i : \Omega \rightarrow \Omega_i$ Abbildungen mit endlicher Indexmenge $I = \{1, \dots, n\}$ für $n \in \mathbb{N}$. Ferner sei \mathcal{A} eine σ -Algebra über Ω , und \mathcal{A}_i sei eine σ -Algebra über Ω_i für alle $i \in I$. Genau dann ist der Vektor (T_1, \dots, T_n) messbar bezüglich der Produkt- σ -Algebra $\bigotimes_{i \in I} \mathcal{A}_i$, wenn alle T_i $(\mathcal{A}, \mathcal{A}_i)$ -messbar sind.

I. Maßtheoretische Grundlagen

Beweis: Wir stellen zunächst fest, dass

$$(T_1, \dots, T_n)^{-1} \left(\bigotimes_{i=1}^n A_i \right) = \bigcap_{i=1}^n T_i^{-1}(A_i) \quad \text{für alle } A_i \in \mathcal{A}_i, i = 1, \dots, n$$

gilt.

" \Rightarrow ": Es sei $j \in I$ fest. Für die Wahl $A_i = \Omega_i, i \neq j$ und $A_j \in \mathcal{A}_j$ beliebig ergibt sich wegen $T_k^{-1}(\Omega_k) = \Omega$ für alle $k \in I$ aus der obigen Beziehung sofort

$$T_j^{-1}(A_j) = \bigcap_{i=1}^n T_i^{-1}(A_i) = (T_1, \dots, T_n)^{-1} \left(\bigotimes_{i=1}^n A_i \right) \in \mathcal{A},$$

woraus die Messbarkeit von T_j folgt.

" \Leftarrow ": Für alle $A_i \in \mathcal{A}_i, i = 1, \dots, n$ ist $\bigcap_{i=1}^n T_i^{-1}(A_i) \in \mathcal{A}$ aufgrund der Messbarkeit jedes T_i , also ist auch

$$(T_1, \dots, T_n)^{-1} \left(\bigotimes_{i=1}^n A_i \right) = \bigcap_{i=1}^n T_i^{-1}(A_i) \in \mathcal{A}$$

und damit (T_1, \dots, T_n) messbar nach Lemma 10, weil $\mathcal{S} := \left\{ \bigotimes_{i \in I} A_i \mid \forall i \in I : A_i \in \mathcal{A}_i \right\}$ ein Erzeuger von $\bigotimes_{i \in I} \mathcal{A}_i$ ist. ■

Einen interessanten Sachverhalt, der für die später zu behandelnden bedingten Wahrscheinlichkeiten wichtig ist, gibt der folgende

Satz 9 (Faktorisierungssatz I). Es seien $T : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ und $S : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_3, \mathcal{A}_3)$ messbare Abbildungen derart, dass $T(\Omega_1)$ abzählbar ist. Die σ -Algebra \mathcal{A}_3 enthalte alle einelementigen Teilmengen $\{\omega_3\}$ mit $\omega_3 \in \Omega_3$ als Elemente. Genau dann ist S bezüglich der von T erzeugten σ -Algebra $T^{-1}(\mathcal{A}_2)$ messbar, wenn es eine Abbildung $g : (\Omega_2, \mathcal{A}_2) \rightarrow (\Omega_3, \mathcal{A}_3)$ gibt mit der Eigenschaft

$$S = g \circ T.$$

Beweis: " \Leftarrow ": Ergibt sich unmittelbar aus Lemma 11.

" \Rightarrow ": Die gewünschte Abbildung g kann aus Gründen der Wohldefiniertheit nun nur existieren, wenn gilt

$$\forall \omega_1, \eta_1 \in \Omega_1 : T(\omega_1) = T(\eta_1) \Rightarrow S(\omega_1) = S(\eta_1),$$

d.h. S muss auf den Urbildmengen $T^{-1}(\{\omega_2\})$ mit $\omega_2 \in T(\Omega_1) \subseteq \Omega_2$ konstant sein.

Mit T ist auch S abzählbar-wertig, d.h. es ist $S(\Omega_1) \in \mathcal{A}_3$ als Vereinigung höchstens abzählbar vieler einelementiger Teilmengen von Ω_3 . Es gibt also eine Folge $\{\omega_{3k}\}_{k \in K}$ paarweise verschiedener

I. Maßtheoretische Grundlagen

Elemente von Ω_3 mit einer höchstens abzählbaren Indexmenge K sowie eine Folge paarweise disjunkter Mengen $\{A_{1k}\}_{k \in K}$ mit $A_{1k} \in T^{-1}(\mathcal{A}_2)$ für alle $k \in K$, so dass

$$\forall k \in K, \omega_1 \in A_{1k} : S(\omega_1) = \omega_{3k}$$

gilt. Ferner existiert zu jedem $A_{1k} \in T^{-1}(\mathcal{A}_2)$ eine Menge $A_{2k} \in \mathcal{A}_2$ mit $A_{1k} = T^{-1}(A_{2k})$, $k \in K$. Wir definieren jetzt g durch

$$\forall k \in K, \omega_2 \in A_{2k} : g(\omega_2) := \omega_{3k}.$$

Auf der ggf. nicht erfassten Menge $\tilde{A}_2 := \left(\bigcup_{k \in K} A_{2k} \right)^c \in \mathcal{A}_2$ können wir g z.B. mit beliebigem Wert $\tilde{\omega}_3 \in \Omega_3$ konstant wählen. Dann ist g konstruktionsbedingt $(\mathcal{A}_2 - \mathcal{A}_3)$ -messbar; denn sei

$$W_3 := \{\omega_{3k} \mid k \in K\} \cup \{\tilde{\omega}_3\} \in \mathcal{A}_3, \text{ der Wertebereich von } g,$$

und $A_3 \in \mathcal{A}_3$ beliebig, so ist $g^{-1}(A_3) = g^{-1}(A_3 \cap W_3) = \bigcup_{\omega_3 \in A_3 \cap W_3} g^{-1}(\{\omega_3\}) \in \mathcal{A}_2$, da für jedes $\omega_3 \in A_3 \cap W_3$ das Urbild $g^{-1}(\{\omega_3\})$ mit einer der Mengen A_{2k} , $k \in K$ oder mit \tilde{A}_2 zusammenfällt, die sämtlich in \mathcal{A}_2 liegen. Für $\omega_1 \in A_{1k}$, $k \in K$ ist nun $T(\omega_1) \in A_{2k}$, also $g \circ T(\omega_1) = \omega_{3k} = S(\omega_1)$. Damit erfüllt g die gewünschten Anforderungen. ■

Eine sehr wesentliche Anwendung von messbaren Abbildungen ergibt sich bei der Transformation von Maßen, die zur Standardmodellierung stochastischer Vorgänge gehört.

Satz 10. Es sei $T : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ eine messbare Abbildung. Dann wird für jedes Maß μ_1 auf \mathcal{A}_1 ein Maß μ_2 auf \mathcal{A}_2 definiert vermöge

$$\forall A_2 \in \mathcal{A}_2 : \mu_2(A_2) := \mu_1(T^{-1}(A_2)).$$

Beweis: Es ist $\mu_2(\emptyset) = \mu_1(T^{-1}(\emptyset)) = \mu_1(\emptyset)$ und trivialerweise $\mu_2 \geq 0$ auf \mathcal{A}_2 . Die σ -Additivität ergibt sich aus der Beziehung

$$\mu_2\left(\bigoplus_{n=1}^{\infty} A_{2n}\right) = \mu_1\left(T^{-1}\left(\bigoplus_{n=1}^{\infty} A_{2n}\right)\right) = \mu_1\left(\bigoplus_{n=1}^{\infty} T^{-1}(A_{2n})\right) = \sum_{n=1}^{\infty} \mu_1(T^{-1}(A_{2n})) = \sum_{n=1}^{\infty} \mu_2(A_{2n})$$

für alle Folgen paarweise disjunkter Mengen $\{A_{2n}\}_{n \in \mathbb{N}} \subseteq \mathcal{A}_2$.

I. Maßtheoretische Grundlagen

Definition 12 (Bildmaß). Es sei $T : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ eine messbare Abbildung und μ_1 ein Maß auf \mathcal{A}_1 . Das nach Satz 10 eindeutig bestimmte Maß μ_2 auf \mathcal{A}_2 heißt *Bildmaß von μ_1 unter T* . Es wird mit $\mu_2 = \mu_1^T$ bezeichnet.

Für die Modellbildung in der Stochastik ist das Bildmaß von zentraler Bedeutung. Ist nämlich (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $X : (\Omega, \mathcal{A}) \rightarrow (\mathfrak{X}, \mathcal{B})$ eine messbare Abbildung, so nennen wir diese in der Stochastik üblicherweise eine **Zufallsvariable** (hauptsächlich im Fall von $\mathfrak{X} \subseteq \mathbb{R}^1$), ansonsten einen **Zufallsvektor** (falls $\mathfrak{X} \subseteq \mathbb{R}^d$ mit $d > 1$) bzw. ein **Zufallselement** (etwa wenn \mathfrak{X} ein allgemeinerer topologischer Raum ist). Der zugrundeliegende Messraum (Ω, \mathcal{A}) ist dabei nicht von explizitem Interesse, vielmehr spielt die Kenntnis des Bild-Messraumes $(\mathfrak{X}, \mathcal{B})$ und des darauf induzierten Bildmaßes, die so genannte **Verteilung von X** , die zentrale Rolle.

Der Grund, warum in der Stochastik der Bildraum \mathfrak{X} so eng mit den reellen Zahlen verknüpft ist, liegt darin, dass die Realisierungen von Zufallsexperimenten, die man auch "Daten" oder "Stichproben" nennt, praktisch immer in entsprechenden Skalen gemessen werden (z.B. Anzahlen, Längen, Gewichte usw.) und mit den Ergebnissen des Zufallsexperiments meist auch gerechnet wird (z.B. Bildung der Summe oder des arithmetischen Mittels). Bei komplexeren Modellen, wie sie etwa in der Bildverarbeitung auftreten, sind die Ergebnisse des Zufallsexperiments in der Regel nicht mehr bloße Zahlen, sondern z.B. geometrische Strukturen, die selbst wieder Elemente eines geeigneten topologischen Raumes bilden.

Wir kommen zum Abschluss dieses Abschnitts noch einmal auf das Problem der Existenz von Produkt- σ -Algebren und Produkt-Maßen mit beliebig vielen Komponenten zurück.

Seien also $(\Omega_i, \mathcal{A}_i)$, $i \in I$ Messräume mit einer beliebigen, nicht-leeren Indexmenge I . Wir betrachten die so genannten *Projektionen* $\pi_j : \prod_{i \in I} \Omega_i \rightarrow \Omega_j$ für $j \in I$, die in vereinfachter Schreibweise definiert sind über $\pi_j \left((\omega_i)_{i \in I} \right) := \omega_j$ für $(\omega_i)_{i \in I} \in \prod_{i \in I} \Omega_i$. Die allgemeine Definition einer Produkt- σ -Algebra lautet:

Definition 13 (allgemeine Produkt- σ -Algebra). Es seien $(\Omega_i, \mathcal{A}_i)$, $i \in I$ Messräume mit einer beliebigen, nicht-leeren Indexmenge I . Dann heißt

$$\otimes_{i \in I} \mathcal{A}_i := \sigma \left(\pi_j ; j \in I \right) = \sigma \left(\bigcup_{j \in I} \pi_j^{-1} (\mathcal{A}_j) \right)$$

die Produkt- σ -Algebra der σ -Algebren \mathcal{A}_i , $i \in I$ (über $\prod_{i \in I} \Omega_i$).

Für endliche Indexmengen fällt diese Definition über die Bildung von Tupeln natürlich wieder mit der früheren Definition 5 zusammen.

I. Maßtheoretische Grundlagen

Wichtig zu bemerken ist in diesem Zusammenhang, dass ähnlich wie in Satz 8 das System⁹

$$\mathcal{S} := \bigcup_{\substack{\emptyset \subset J \subset I \\ J \text{ endlich}}} \left\{ \prod_{j \in J} A_j \times \prod_{j \in I \setminus J} \Omega_j \mid \forall j \in J : A_j \in \mathcal{A}_j \right\}$$

einen Semi-Ring über $\prod_{i \in I} \Omega_i$ bildet, der die Produkt- σ -Algebra $\bigotimes_{i \in I} \mathcal{A}_i$ erzeugt. Eine wesentliche Strukturaussage, die wir hier nicht beweisen wollen (siehe z.B. GÄNSSLER UND STUTE (1977), Satz 1.3.11 oder KLENKE (2006), Kapitel 14.1), enthält

Satz 11 (Struktursatz für Produkt- σ -Algebren). In dem Produkt-Messraum $\left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{A}_i \right)$ gilt: Zu jedem $A \in \bigotimes_{i \in I} \mathcal{A}_i$ existiert eine abzählbare Menge $J \subseteq I$ und eine Menge $B \in \bigotimes_{j \in J} \mathcal{A}_j$ derart, dass gilt: $A = B \times \prod_{i \in I \setminus J} \Omega_i$, d.h. "A hängt nur von abzählbar vielen Koordinaten ab".

Die Existenz und Eindeutigkeit eines entsprechenden Produktmaßes für normierte Maße garantiert der folgende Satz (siehe z.B. GÄNSSLER UND STUTE (1977), Satz 1.9.7):

Satz 12 (Produktmaßsatz). In dem Produkt-Messraum $\left(\prod_{i \in I} \Omega_i, \bigotimes_{i \in I} \mathcal{A}_i \right)$ mit Wahrscheinlichkeitsmaßen P_i auf \mathcal{A}_i , $i \in I$ gilt: Es existiert genau ein Wahrscheinlichkeitsmaß P auf $\bigotimes_{i \in I} \mathcal{A}_i$ mit der Eigenschaft

$$P(A) = P\left(B \times \prod_{i \in I \setminus J} \Omega_i \right) = \prod_{j \in J} P_j(B_j) \text{ für } A = B \times \prod_{i \in I \setminus J} \Omega_i, B = \prod_{j \in J} B_j$$

mit $B_j \in \mathcal{A}_j$, $j \in J \subseteq I$, J abzählbar.

Diese allgemeinen Produktmaße liefern für die Stochastik das "Herzstück" der Modellierung, denn sie geben den formalen Rahmen dafür ab, was wir *stochastische Unabhängigkeit* nennen. Das bereits in der historischen Vorbemerkung mehrfach genannte *Gesetz der Großen Zahlen* sowie der damit verwandte *Zentrale Grenzwertsatz* beruhen ganz wesentlich auf diesem Begriff; erst diese mathematischen Lehrsätze ermöglichen eine sinnvolle Verbindung zwischen "Theorie" und "Praxis", indem sie Möglichkeiten aufzeigen, wie aus "Daten" oder "Stichproben" Rückschlüsse auf die zu Grunde liegenden Wahrscheinlichkeiten gezogen werden können. Ohne den Begriff der stochastischen Unabhängigkeit gäbe es keine schließende Statistik als wissenschaftliche Disziplin und damit letztlich auch keine empirische Überprüfbarkeit naturwissenschaftlicher Erkenntnisse.

⁹ Da bei überabzählbaren Indexmengen I i.d.R. keine konstruktive Wohlordnung möglich ist, ist das kartesische Produkt $\prod_{j \in J} A_j \times \prod_{j \in I \setminus J} \Omega_j$ nicht im Sinne einer Reihenfolge zu verstehen; vgl. dazu Seite 5 unten.

I. Maßtheoretische Grundlagen

Bemerkung: Produkt- σ -Algebren mit überabzählbaren Indexmengen I besitzen i. Allg. keine Atome. Wählen wir beispielsweise $\Omega_i := \{0,1\}$, $\mathcal{A}_i := \mathfrak{P}(\Omega_i)$ und $I = \mathbb{R}$, so besteht $\bigotimes_{i \in I} \mathcal{A}_i$ nach Satz 11 genau aus solchen Mengen von Abbildungen $f : \mathbb{R} \rightarrow \{0,1\}$, die an höchstens abzählbar vielen Stellen "Vorgaben" bzgl. der Funktionswerte machen. So ist beispielsweise

$$A := \left\{ f : \mathbb{R} \rightarrow \{0,1\} \mid f(x) = 0 \text{ für alle } x \in \mathbb{Q} \right\} \in \bigotimes_{i \in I} \mathcal{A}_i,$$

aber für die einelementige Menge $B := \left\{ f : \mathbb{R} \rightarrow \{0,1\} \mid f(x) = 0 \text{ für alle } x \in \mathbb{R} \right\}$ gilt $B \notin \bigotimes_{i \in I} \mathcal{A}_i$. Die Produkt- σ -Algebra $\bigotimes_{i \in I} \mathcal{A}_i$ besitzt nun keine Atome; dies zeigen wir durch ein Widerspruchsargument: angenommen, es gäbe ein Atom $T \in \bigotimes_{i \in I} \mathcal{A}_i$. Dann existieren nach obigem eine höchstens abzählbare Menge $J \subset \mathbb{R}$ und eine Menge $B \in \bigotimes_{j \in J} \mathcal{A}_j$ mit

$$T = B \times \prod_{i \in I \setminus J} \Omega_i.$$

Wegen $J \neq \mathbb{R}$ existiert ein $k \in \mathbb{R} \setminus J$; wir setzen $K := J \oplus \{k\}$ und $C_k := \{0\}$. Die nicht-leere Menge

$$S := B \times C_k \times \prod_{i \in I \setminus K} \Omega_i$$

liegt dann ebenfalls in $\bigotimes_{i \in I} \mathcal{A}_i$, mit $S \subsetneq T$: Widerspruch zur Atom-Eigenschaft von T !

Insbesondere folgt hieraus, dass alle einelementigen Mengen $\{g\}$ mit $g \in \prod_{i \in I} \Omega_i$ keine Atome von $\bigotimes_{i \in I} \mathcal{A}_i$ sind. Für das Folgende wollen wir mit $\mathbf{0}$ die Nullabbildung auf \mathbb{R} bezeichnen. Wählt man jetzt speziell die von der Produkt- σ -Algebra $\bigotimes_{i \in I} \mathcal{A}_i$ und der Nullabbildung gemeinsam erzeugte σ -Algebra $\mathcal{B} := \sigma\left(\bigotimes_{i \in I} \mathcal{A}_i \cup \{\mathbf{0}\}\right)$, so besitzt diese das Atomsystem $\mathcal{T} = \{\{\mathbf{0}\}\}$, welches weder erschöpfend noch vollständig im Sinne von Definition 4 ist.

Die vorangehende Bemerkung zeigt damit auch, dass Satz 4 i. Allg. für *überabzählbare* Indexmengen nicht gilt; allerdings bleibt Satz 4 für *abzählbare* Indexmengen richtig, wie man sich anhand von Satz 11 leicht überlegen kann.

I. 6. Borel'sche σ -Algebra und Lebesgue'sches Maß

In diesem Abschnitt wenden wir uns der maßtheoretischen Behandlung der Begriffe "Länge", "Fläche" und "Volumen" sowie deren mehrdimensionaler Verallgemeinerungen zu. Der zu Grunde liegende Raum ist hier $\Omega = \mathbb{R}^d$ mit der Dimension $d \in \mathbb{N}$.

Definition 14 (Standard-Intervall / Borel'sche σ -Algebra). Für $\mathbf{a} = (a_1, \dots, a_d), \mathbf{b} = (b_1, \dots, b_d) \in \mathbb{R}^d$ heißt die Menge

$$I_{\mathbf{a}, \mathbf{b}} := \prod_{i=1}^d (a_i, b_i]$$

Standard-Intervall im \mathbb{R}^d . Dabei sei vereinbarungsgemäß $(a, b] = \emptyset$ für $b \leq a, a, b \in \mathbb{R}$. Die von $\mathcal{E}^d := \{I_{\mathbf{a}, \mathbf{b}} \mid \mathbf{a}, \mathbf{b} \in \mathbb{R}^d\}$ erzeugte σ -Algebra über \mathbb{R}^d heißt *Borel'sche σ -Algebra* der Dimension d und wird üblicherweise mit dem Symbol \mathcal{B}^d bezeichnet.

Die vielleicht zunächst eigenartig anmutende Definition des Standard-Intervalls entstand aus dem Bemühen, "größere" Intervalle als disjunkte Vereinigung "kleinerer" Intervalle darstellen zu können. Tatsächlich gilt folgender Sachverhalt:

Lemma 13. Der Erzeuger \mathcal{E}^d ist in jeder Dimension $d \in \mathbb{N}$ ein \mathbb{R}^d abzählbar überdeckender Semi-Ring.

Beweis: Wir können den entsprechenden Beweis von Satz 8 a) imitieren. Wir zeigen die Aussage zunächst für $d = 1$. Die Standard-Intervalle sind hier von der Form $I_{a,b} = (a, b]$ mit reellen $a < b$. Es ist folgendes zu zeigen:

- $\emptyset \in \mathcal{E}^1$ wegen $(a, b] = \emptyset$ für $b \leq a, a, b \in \mathbb{R}$.
- Seien $a_1, a_2, b_1, b_2 \in \mathbb{R}$ mit $a_1 < b_1, a_2 < b_2$. Dann ist

$$I_{a_1, b_1} \cap I_{a_2, b_2} = (a_1, b_1] \cap (a_2, b_2] = \begin{cases} \emptyset, & \text{falls } b_1 \leq a_2 \text{ oder } b_2 \leq a_1 \\ (\max\{a_1, a_2\}, \min\{b_1, b_2\}], & \text{falls } a_2 < b_1 \text{ und } a_1 < b_2 \end{cases} \in \mathcal{E}^1,$$

d.h. \mathcal{E}^1 ist durchschnittsstabil (die übrigen Fälle sind trivial).

Seien $a_1, a_2, b_1, b_2 \in \mathbb{R}$ mit $a_1 < b_1, a_2 < b_2$. Dann ist

$$I_{a_1, b_1} \setminus I_{a_2, b_2} = (a_1, b_1] \setminus (a_2, b_2] = \begin{cases} \emptyset, & \text{falls } a_2 \leq a_1 \text{ und } b_1 \leq b_2 \\ (a_1, a_2] \oplus (b_1, b_2], & \text{falls } a_2 > a_1 \text{ oder } b_1 > b_2 \end{cases} \in \mathcal{E}^1,$$

d.h. die Mengen-Differenz von Standard-Intervallen ist als disjunkte Vereinigung anderer Standard-Intervalle darstellbar (die übrigen Fälle sind trivial).

Insgesamt ist also \mathcal{E}^1 ein Semi-Ring über \mathbb{R}^1 .

I. Maßtheoretische Grundlagen

Sei nun $d \in \mathbb{N}$ beliebig. Es ist trivialerweise $\emptyset \in \mathcal{E}^d$ und \mathcal{E}^d durchschnittsstabil nach Lemma 1 wegen

$$I_{\mathbf{a},\mathbf{b}} \cap I_{\mathbf{c},\mathbf{d}} = \left(\prod_{i=1}^d (a_i, b_i] \right) \cap \left(\prod_{i=1}^d (c_i, d_i] \right) = \prod_{i=1}^d \underbrace{\{(a_i, b_i] \cap (c_i, d_i]\}}_{\in \mathcal{E}^1} \in \mathcal{E}^d \text{ für } \mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d} \in \mathbb{R}^d.$$

Für die Mengendifferenz genügt es wegen $I_{\mathbf{a},\mathbf{b}} \setminus I_{\mathbf{c},\mathbf{d}} = I_{\mathbf{a},\mathbf{b}} \setminus (I_{\mathbf{a},\mathbf{b}} \cap I_{\mathbf{c},\mathbf{d}})$ nur den nicht-trivialen Fall $I_{\mathbf{c},\mathbf{d}} \subseteq I_{\mathbf{a},\mathbf{b}}$, d.h. $a_i \leq c_i < d_i \leq b_i$ für $i=1, \dots, d$ zu betrachten. Zur Abkürzung benennen wir die Intervalle $A_i := (a_i, c_i]$, $C_i := (c_i, d_i]$, $D_i := (d_i, b_i]$, $i=1, \dots, d$. Dann gilt:

$$I_{\mathbf{a},\mathbf{b}} = \prod_{i=1}^d (A_i \oplus C_i \oplus D_i) = \bigoplus \left\{ \prod_{i=1}^d E_i \mid \forall i=1, \dots, d: E_i \in \{A_i, C_i, D_i\} \right\},$$

d.h. $I_{\mathbf{a},\mathbf{b}}$ ist darstellbar als disjunkte Vereinigung von maximal 3^d Standard-Intervallen, unter denen $I_{\mathbf{c},\mathbf{d}} = \prod_{i=1}^d C_i$ als Spezialfall auftritt. Es folgt

$$I_{\mathbf{a},\mathbf{b}} \setminus I_{\mathbf{c},\mathbf{d}} = \bigoplus \left\{ \prod_{i=1}^d E_i \mid \forall i=1, \dots, d: E_i \in \{A_i, C_i, D_i\} \right\} \setminus \prod_{i=1}^d C_i,$$

d.h. $I_{\mathbf{a},\mathbf{b}} \setminus I_{\mathbf{c},\mathbf{d}}$ ist darstellbar als disjunkte Vereinigung von maximal $3^d - 1$ Standard-Intervallen. Damit ist insgesamt gezeigt, dass \mathcal{E}^d ein Semi-Ring über \mathbb{R}^d ist. \mathcal{E}^d ist auch \mathbb{R}^d abzählbar überdeckend wegen

$$\mathbb{R}^d = \bigcup_{n=1}^{\infty} \prod_{i=1}^d (-n, n].$$

Damit ist alles bewiesen. ■

Es ist interessant zu untersuchen, welche Mengen die Borel'sche σ -Algebra insgesamt umfasst. Mit Hilfe des Auswahlaxioms werden wir später zeigen, dass $\mathcal{B}^d \neq \mathfrak{P}(\mathbb{R}^d)$ ist.

Satz 13. Die folgenden Mengensysteme sind – neben anderen – ebenfalls Erzeuger der Borel'schen σ -Algebra \mathcal{B}^d :

$$\begin{aligned} \mathcal{E}_O^d &:= \{O \subseteq \mathbb{R}^d \mid O \text{ offen}\} \\ \mathcal{E}_A^d &:= \{A \subseteq \mathbb{R}^d \mid A \text{ abgeschlossen}\} \\ \mathcal{E}_K^d &:= \{K \subseteq \mathbb{R}^d \mid K \text{ kompakt}\}. \end{aligned}$$

Insbesondere enthält \mathcal{B}^d auch alle "gemischten" Intervalle I der Form $I = \prod_{i=1}^d E_i$ mit $E_i \in \{(a_i, b_i), (a_i, b_i], [a_i, b_i), [a_i, b_i]\}$ und wird auch von den entsprechenden Mengensystemen erzeugt.

I. Maßtheoretische Grundlagen

Beweis: Wir zeigen zunächst $\mathcal{E}^d \subseteq \sigma(\mathcal{E}_O^d)$ und $\mathcal{E}_O^d \subseteq \sigma(\mathcal{E}^d)$, woraus $\sigma(\mathcal{E}_O^d) = \mathcal{B}^d$ folgt. Sei also

$I_{a,b} := \times_{i=1}^d (a_i, b_i]$ vorgegeben. Es ist $\times_{i=1}^d (a_i, b_i] = \bigcap_{n=1}^{\infty} \times_{i=1}^d \left(a_i, b_i + \frac{1}{n} \right) \in \sigma(\mathcal{E}_O^d)$ und damit $\mathcal{E}^d \subseteq \sigma(\mathcal{E}_O^d)$.

Sei umgekehrt O eine beliebige nicht-leere offene Menge im \mathbb{R}^d . Dann lässt sich (wegen der Dichtheit von \mathbb{Q}^d in \mathbb{R}^d) O darstellen als abzählbare Vereinigung nicht-leerer offener Intervalle

der Form $\times_{i=1}^d (a_i, b_i)$. Für ein solches Intervall gilt aber auch $\times_{i=1}^d (a_i, b_i) = \bigcup_{n=1}^{\infty} \times_{i=1}^d \left(a_i, b_i - \frac{\delta_i}{n} \right]$ mit

$\delta_i < b_i - a_i, i = 1, \dots, d$. Also ist auch $\mathcal{E}_O^d \subseteq \sigma(\mathcal{E}^d)$ und somit $\sigma(\mathcal{E}_O^d) = \mathcal{B}^d$.

Ferner ist $\sigma(\mathcal{E}_A^d) = \sigma(\mathcal{E}_O^d) = \mathcal{B}^d$, weil die abgeschlossenen Mengen die Komplemente der offenen Mengen im \mathbb{R}^d sind und umgekehrt, und jede σ -Algebra durchschnittsstabil ist.

Schließlich ist $\sigma(\mathcal{E}_K^d) = \sigma(\mathcal{E}_A^d) = \mathcal{B}^d$, weil jede kompakte Menge im \mathbb{R}^d abgeschlossen und be-

schränkt ist und für jede abgeschlossene Menge $A \subseteq \mathbb{R}^d$ gilt: $A = \bigcup_{n=1}^{\infty} \underbrace{\left(A \cap \times_{i=1}^d [-n, n] \right)}_{\text{kompakt}}$.

Die restlichen Aussagen ergeben sich auf analoge Weise, indem man die entsprechenden Approximationen für den linken bzw. rechten Intervall-Endpunkt vornimmt. ■

Insbesondere enthält \mathcal{B}^d also alle einelementigen (da kompakten) und damit auch alle abzählbaren Teilmengen von \mathbb{R}^d . Demnach existiert für \mathcal{B}^d auch ein vollständiges Atomsystem \mathcal{T}^d , welches durch

$$\mathcal{T}^d := \{ \{ \omega \} \mid \omega \in \mathbb{R}^d \}$$

gegeben ist. Allerdings ist dieses Atomsystem nicht mehr abzählbar und auch nicht erschöpfend, weil hier nur

$$\sigma(\mathcal{T}^d) = \{ A \subseteq \mathbb{R}^d \mid A \sim_{\#} \mathbb{N} \text{ oder } A^c \sim_{\#} \mathbb{N} \}$$

gilt (sog. σ -Algebra der *abzählbaren Komplemente*), welche offensichtlich nicht mit \mathcal{B}^d identisch ist, weil z.B. weder $(-\infty, 0) \in \mathcal{E}_O^d$ noch $[0, \infty) \in \mathcal{E}_A^d$ abzählbar sind. Damit ist die Borel'sche σ -Algebra eine in jeder Hinsicht nicht-triviale¹⁰ σ -Algebra über \mathbb{R}^d .

Wir wollen jetzt zeigen, wie die anschaulichen Begriffe "Länge", "Fläche" und "Volumen" mit einem geeigneten Maß auf der Borel'schen σ -Algebra \mathcal{B}^d in Zusammenhang gebracht werden können.

¹⁰ Eine "konstruktive" Beschreibung der Borel'schen σ -Algebra findet man z.B. in BILLINGSLEY (1986), S. 26ff.

I. Maßtheoretische Grundlagen

Definition 15. Die für Standard-Intervalle vermöge

$$m^d(I_{\mathbf{a},\mathbf{b}}) = m^d\left(\prod_{i=1}^d (a_i, b_i)\right) := \prod_{i=1}^d (b_i - a_i), \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^d, \quad a_j \leq b_j \text{ für } j = 1, \dots, d$$

erklärte Mengenfunktion m^d heißt *Elementarinhalt* auf dem Semi-Ring \mathcal{E}^d .

Der Elementarinhalt fällt also für Standard-Intervalle und für die Dimensionen $d = 1, 2, 3$ gerade mit den Begriffen "Länge", "Fläche" und "Volumen" zusammen. Mit Hilfe des "großen" Fortsetzungssatzes 7 läßt sich der Elementarinhalt nun eindeutig auf die Borel'sche σ -Algebra \mathcal{B}^d fortsetzen. Wie schon früher angekündigt, werden wir diesen Prozess hier einmal exemplarisch ausführlicher darstellen. Zuvor benötigen wir allerdings noch ein Hilfsresultat aus der Analysis.

Lemma 14. Für beliebige reelle Zahlen $x_1, \dots, x_d \geq 0$ und $0 \leq \delta \leq 1$ gilt:

$$\prod_{i=1}^d (x_i + \delta) - \prod_{i=1}^d x_i \leq \delta \cdot \prod_{i=1}^d (x_i + 1).$$

Beweis: Wir zeigen die Behauptung mit vollständiger Induktion nach d . Für $d = 1$ reduziert sich die Ungleichung auf

$$(x_1 + \delta) - x_1 = \delta \leq \delta \cdot (x_1 + 1),$$

was wegen $x_1 \geq 0$ richtig ist. Es gelte die Behauptung nun für ein $d \in \mathbb{N}$. Es folgt dann

$$\begin{aligned} \prod_{i=1}^{d+1} (x_i + \delta) - \prod_{i=1}^{d+1} x_i &= (x_{d+1} + \delta) \cdot \prod_{i=1}^d (x_i + \delta) - x_{d+1} \cdot \prod_{i=1}^d x_i \\ &= x_{d+1} \cdot \left(\prod_{i=1}^d (x_i + \delta) - \prod_{i=1}^d x_i \right) + \delta \cdot \prod_{i=1}^d (x_i + \delta) \\ &\leq x_{d+1} \cdot \delta \cdot \prod_{i=1}^d (x_i + 1) + \delta \cdot \prod_{i=1}^d (x_i + 1) = \delta \cdot \prod_{i=1}^{d+1} (x_i + 1), \end{aligned}$$

d.h. die Aussage gilt auch für $d + 1$. ■

Satz 14 (Fortsetzungssatz für den Elementarinhalt). Es existiert genau eine Fortsetzung des Elementarinhalts m^d von \mathcal{E}^d auf die Borel'sche σ -Algebra \mathcal{B}^d zu einem σ -endlichen Maß. Dieses wird unter Beibehaltung der Notation m^d als *Lebesgue-Maß* bezeichnet.

Beweis: Nach Satz 7 bleibt nur noch zu zeigen, dass $m^d\left(\bigoplus_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} m^d(S_n)$ gilt für alle paarweise disjunkten Mengenfolgen $\{S_n\}_{n \in \mathbb{N}} \subseteq \mathcal{E}^d$ mit der Eigenschaft $\bigoplus_{n=1}^{\infty} S_n \in \mathcal{E}^d$, denn \mathcal{E}^d ist ein Semi-Ring nach Lemma 13 und m^d erfüllt per Definition die Eigenschaften $m^d(\emptyset) = 0$ und $m^d \geq 0$; ferner ist $\mathbb{R}^d = \bigcup_{n=1}^{\infty} \prod_{i=1}^d (-n, n]$ mit $m^d\left(\prod_{i=1}^d (-n, n]\right) = (2n)^d < \infty$, d.h. der Elementarinhalt ist auf

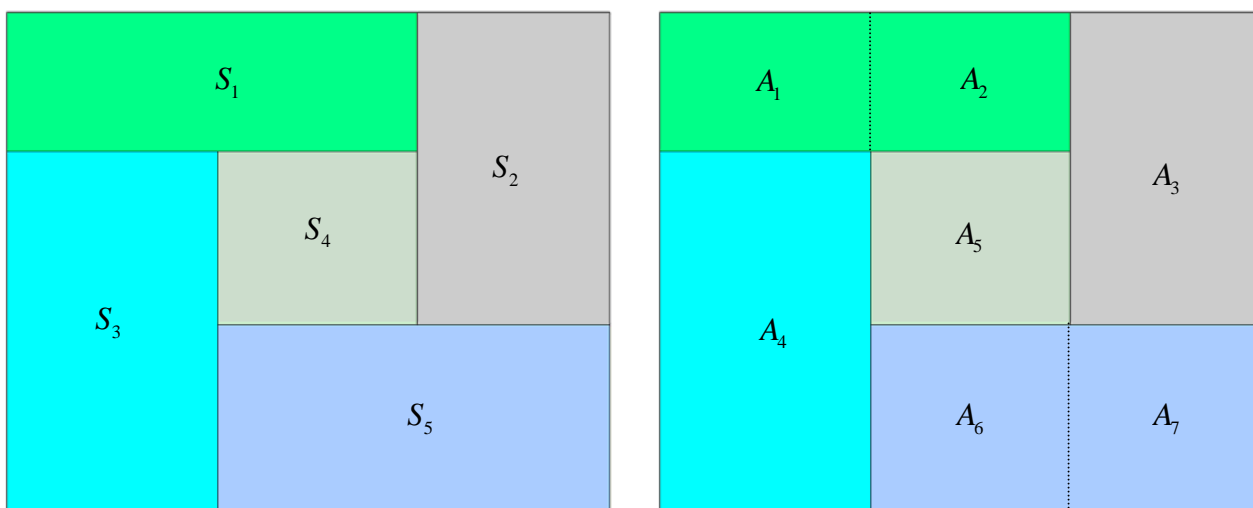
I. Maßtheoretische Grundlagen

\mathcal{E}^d σ -endlich. Wir schlagen dazu einen kleinen Umweg über ein Kompaktheitsargument ein, das ursprünglich auf Émile Borel zurückgeht.

Zunächst zeigen wir, dass m^d (endlich) additiv auf \mathcal{E}^d ist. Für $d=1$ sieht man das etwa so: ist $S_i = (a_{i-1}, a_i]$ für $i \in I = \{1, \dots, m\}$ mit $m \in \mathbb{N}$ und $a_0 < a_1 < \dots < a_m$, so ist

$$\bigoplus_{i \in I} S_i = (a_0, a_m] \text{ mit } \sum_{i \in I} m^d(S_i) = \sum_{i=1}^m (a_i - a_{i-1}) = a_m - a_0 = m^d\left(\bigoplus_{i \in I} S_i\right).$$

Man mache sich klar, dass für $d=1$ alle disjunkten Zerlegungen von Standard-Intervallen in Standard-Teilintervalle auf diesen Darstellungstyp zurückgeführt werden können. In höheren Dimensionen kann man ähnlich argumentieren, allerdings ist die Situation hier etwas komplizierter. Die nachfolgende Graphik zeigt dies anschaulich für den Fall $d=2$:



Im linken Bild lassen sich nicht direkt zwei benachbarte Standard-Intervalle durch Vereinigung zu *einem* (größeren) Standard-Intervall vereinigen. Man benötigt dafür einen oder mehrere geeignete *Hilfsschnitte* wie im rechten Bild. Damit wird es jedoch möglich, konsekutiv je zwei benachbarte Standard-Intervalle zu einem (größeren) Standard-Intervall zu vereinigen; hier etwa so:

$$S_1 \oplus S_2 \oplus S_3 \oplus S_4 \oplus S_5 = \{(A_1 \oplus A_4) \oplus (A_2 \oplus A_5 \oplus A_6)\} \oplus (A_3 \oplus A_7)$$

Es bleibt daher nur zu zeigen, dass die Additivität von m^d für zwei "benachbarte" disjunkte Standard-Intervalle, die durch Teilung eines größeren Standard-Intervalls entstehen, gewährleistet ist. Für das obige Beispiel folgt dann etwa, wie gewünscht:

I. Maßtheoretische Grundlagen

$$\begin{aligned}
 \mathfrak{m}^2(S_1 \oplus S_2 \oplus S_3 \oplus S_4 \oplus S_5) &= \mathfrak{m}^2\left(\{(A_1 \oplus A_4) \oplus (A_2 \oplus A_5 \oplus A_6)\} \oplus (A_3 \oplus A_7)\right) \\
 &= \mathfrak{m}^2(A_1 \oplus A_4) + \mathfrak{m}^2(A_2 \oplus A_5 \oplus A_6) + \mathfrak{m}^2(A_3 \oplus A_7) \\
 &= \mathfrak{m}^2(A_1) + \mathfrak{m}^2(A_4) + \mathfrak{m}^2(A_2) + \mathfrak{m}^2(A_5) + \mathfrak{m}^2(A_6) + \mathfrak{m}^2(A_3) + \mathfrak{m}^2(A_7) \\
 &= \mathfrak{m}^2(A_1) + \mathfrak{m}^2(A_2) + \mathfrak{m}^2(A_3) + \mathfrak{m}^2(A_4) + \mathfrak{m}^2(A_5) + \mathfrak{m}^2(A_6) + \mathfrak{m}^2(A_7) \\
 &= \mathfrak{m}^2(S_1) + \mathfrak{m}^2(S_2) + \mathfrak{m}^2(S_3) + \mathfrak{m}^2(S_4) + \mathfrak{m}^2(S_5).
 \end{aligned}$$

Sei also $I_{\mathbf{a},\mathbf{b}} = \times_{i=1}^d (a_i, b_i]$ gegeben. Eine Aufteilung in zwei Teilintervalle ist dann bestimmt durch die Angabe eines Index $j \in \{1, \dots, d\}$ und eines $c_j \in (a_j, b_j]$, so dass mit $\tilde{\mathbf{a}} := (a_1, \dots, a_{j-1}, c_j, a_{j+1}, \dots, a_d)$ und $\tilde{\mathbf{b}} := (b_1, \dots, b_{j-1}, c_j, b_{j+1}, \dots, b_d)$ folgt:

$$I_{\mathbf{a},\mathbf{b}} = \times_{i=1}^d (a_i, b_i] = \left(\times_{i=1}^{j-1} (a_i, b_i] \right) \times \left((a_j, c_j] \oplus (c_j, b_j] \right) \times \left(\times_{i=j+1}^d (a_i, b_i] \right) = I_{\mathbf{a},\tilde{\mathbf{b}}} \oplus I_{\tilde{\mathbf{a}},\mathbf{b}}$$

sowie

$$\begin{aligned}
 \mathfrak{m}^d(I_{\mathbf{a},\tilde{\mathbf{b}}}) + \mathfrak{m}^d(I_{\tilde{\mathbf{a}},\mathbf{b}}) &= \left(\prod_{i=1}^{j-1} (b_i - a_i) \right) \cdot (c_j - a_j) \cdot \left(\prod_{i=j+1}^d (b_i - a_i) \right) + \left(\prod_{i=1}^{j-1} (b_i - a_i) \right) \cdot (b_j - c_j) \cdot \left(\prod_{i=j+1}^d (b_i - a_i) \right) \\
 &= \left(\prod_{i=1}^{j-1} (b_i - a_i) \right) \cdot \{(c_j - a_j) + (b_j - c_j)\} \cdot \left(\prod_{i=j+1}^d (b_i - a_i) \right) \\
 &= \left(\prod_{i=1}^{j-1} (b_i - a_i) \right) \cdot (b_j - a_j) \cdot \left(\prod_{i=j+1}^d (b_i - a_i) \right) = \prod_{i=1}^d (b_i - a_i) = \mathfrak{m}^d(I_{\mathbf{a},\mathbf{b}}).
 \end{aligned}$$

Damit ist aber die Additivität für einen Teilungsschritt gezeigt und damit letztlich auch die (endliche) Additivität von \mathfrak{m}^d auf dem Semi-Ring \mathcal{E}^d .

Wir zeigen nun zum Abschluss, dass \mathfrak{m}^d auf dem Semi-Ring \mathcal{E}^d sogar σ -additiv ist. Mit Hilfe des großen Maßfortsetzungssatzes (Satz 7) folgt dann, dass der Elementarinhalt (der damit ein σ -endliches Prämaß auf \mathcal{E}^d ist) in eindeutiger Weise zu einem Maß auf die Borel'sche σ -Algebra fortgesetzt werden kann, was die Aussage des Satzes 14 beweist.

Sei also $\{S_n\}_{n \in \mathbb{N}} \subseteq \mathcal{E}^d$ eine Folge paarweise disjunkter Mengen mit $S := \bigoplus_{n=1}^{\infty} S_n \in \mathcal{E}^d$. Jede dieser Mengen ist ein Standard-Intervall, o.B.d.A. etwa der Form

$$S = I_{\mathbf{a},\mathbf{b}} \text{ mit } \mathbf{a} = (a_1, \dots, a_d), \mathbf{b} = (b_1, \dots, b_d) \text{ und}$$

$$S_n = I_{\mathbf{a}_n, \mathbf{b}_n} \text{ mit } \mathbf{a}_n = (a_{1n}, \dots, a_{dn}), \mathbf{b}_n = (b_{1n}, \dots, b_{dn}) \text{ und } a_{in} \leq b_{in}, 1 \leq i \leq d, n \in \mathbb{N}.$$

Wir betrachten die daraus abgeleiteten offenen Mengen

I. Maßtheoretische Grundlagen

$$S_n^o(\varepsilon) := \times_{i=1}^d (a_{i_n} - \varepsilon_n, b_{i_n} + \varepsilon_n)$$

sowie die größeren Standard-Intervalle

$$I_n(\varepsilon) := \times_{i=1}^d (a_{i_n} - \varepsilon_n, b_{i_n} + \varepsilon_n] \text{ mit } \varepsilon_n := \frac{\varepsilon}{2^n} \text{ für festes } \varepsilon \in \left(0, \frac{1}{2}\right).$$

Mit der Setzung $M(\varepsilon) := \prod_{i=1}^d (b_i - a_i + 2\varepsilon + 1)$ folgt nun nach Lemma 14

$$m^d(I_n(\varepsilon)) - m^d(S_n) \leq 2\varepsilon_n M(\varepsilon) \text{ für alle } n \in \mathbb{N} \text{ und } \varepsilon \in \left(0, \frac{1}{2}\right).$$

Wir betrachten jetzt das kompaktifizierte Intervall \bar{S} (d.h. S vereinigt mit der Menge seiner Häufungspunkte, also hier: mit seinem Rand). Wegen $S_n^o(\varepsilon) \supset S_n$ für alle $n \in \mathbb{N}$ und $\varepsilon \in \left(0, \frac{1}{2}\right)$ bildet also die Folge $\{S_n^o(\varepsilon)\}_{n \in \mathbb{N}}$ für jedes $\varepsilon \in \left(0, \frac{1}{2}\right)$ eine offene Überdeckung von \bar{S} ; da \bar{S} aber kompakt ist, existiert bereits eine *endliche* (aber i. Allg. von ε abhängende) Indexmenge $I \subset \mathbb{N}$ mit

$$S \subseteq \bar{S} \subseteq \bigcup_{i \in I} S_i^o(\varepsilon) \subseteq \bigcup_{i \in I} I_i(\varepsilon).$$

Damit erhalten wir aber aufgrund der Monotonie und Subadditivität von m^d auf \mathcal{E}^d folgende Ungleichungskette:

$$m^d(S) \leq \sum_{i \in I} m^d(I_i(\varepsilon)) \leq \sum_{i \in I} m^d(S_i) + 2M(\varepsilon) \sum_{i \in I} \varepsilon_i \leq \sum_{n=1}^{\infty} m^d(S_n) + 2\varepsilon M(\varepsilon) \text{ für jedes } \varepsilon \in \left(0, \frac{1}{2}\right)$$

und damit

$$m^d(S) \leq \sum_{n=1}^{\infty} m^d(S_n).$$

Umgekehrt gilt aber auch für beliebige endliche Indexmengen $J \subseteq \mathbb{N}$

$$\sum_{i \in J} m^d(S_i) = m^d\left(\bigoplus_{i \in J} S_i\right) \leq m^d(S)$$

und damit durch Grenzwertbildung

$$\sum_{n=1}^{\infty} m^d(S_n) \leq m^d(S).$$

Insgesamt erhält man also $m^d(S) = \sum_{n=1}^{\infty} m^d(S_n)$, was zu zeigen war. ■

I. Maßtheoretische Grundlagen

Eine charakteristische Eigenschaft des Lebesgue-Maßes ist seine *Translationsinvarianz*, die Gegenstand des nachfolgenden Lemmas ist.

Lemma 15. Das Lebesgue-Maß m^d ist translationsinvariant, d.h. es gilt

$$m^d = (m^d)^{T_a} \text{ mit den Abbildungen } T_a : \mathbb{R}^d \rightarrow \mathbb{R}^d : \mathbf{x} \mapsto \mathbf{x} + \mathbf{a} \text{ für jedes feste } \mathbf{a} \in \mathbb{R}^d.$$

Beweis: Es gilt

$$(T_a)^{-1}(I_{\mathbf{b},\mathbf{c}}) = I_{\mathbf{b}-\mathbf{a},\mathbf{c}-\mathbf{a}} \text{ für alle } \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^d.$$

Nach Lemma 10 sind also alle Abbildungen T_a messbar, mit

$$\prod_{i=1}^d (c_i - b_i) = \prod_{i=1}^d ((c_i - a_i) - (b_i - a_i)) = m^d(I_{\mathbf{b}-\mathbf{a},\mathbf{c}-\mathbf{a}}) = (m^d)^{T_a}(I_{\mathbf{b},\mathbf{c}}) = m^d(I_{\mathbf{b},\mathbf{c}}).$$

Da m^d und $(m^d)^{T_a}$ auf dem erzeugenden Semi-Ring \mathcal{E}^d übereinstimmen und dort σ -endlich sind, stimmen sie also nach dem großen Maßfortsetzungssatz 7 und Lemma 13 auch auf ganz \mathcal{B}^d überein. ■

Man kann zeigen, dass die Eigenschaft der Translationsinvarianz zusammen mit der Forderung, dass das Einheitsintervall $\times_{i=1}^d (0,1]$ das Maß 1 besitzt, das Lebesgue-Maß m^d eindeutig festlegt. Es ist darüber hinaus das einzige (derart normierte) Maß auf \mathcal{B}^d , welches *kongruenzinvariant* ist (also z.B. auch invariant gegenüber Drehungen und Spiegelungen).

Mit Hilfe des Lebesgue-Maßes und des Auswahlaxioms¹¹ können wir nun wie angekündigt zeigen, dass die Borel'sche σ -Algebra nicht mit der Potenzmenge $\mathfrak{P}(\mathbb{R}^d)$ zusammenfällt.

Satz 15 (Vitali 1905). Für jedes $d \in \mathbb{N}$ gilt: $\mathcal{B}^d \neq \mathfrak{P}(\mathbb{R}^d)$.

Beweis: Wir definieren die folgende Äquivalenzrelation: $\mathbf{x} \sim \mathbf{y} : \Leftrightarrow \mathbf{x} - \mathbf{y} \in \mathbb{Q}^d$ für alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$. K sei eine Repräsentantenmenge dieser Äquivalenzrelation. Dann zerfällt \mathbb{R}^d in (paarweise disjunkte) Äquivalenzklassen der Form $\mathbf{k} + \mathbb{Q}^d$ mit $\mathbf{k} \in K$. Zu $\mathbf{k} \in K$ existiert aber ein eindeutig bestimmtes $\mathbf{n} \in \mathbb{Z}^d$ mit $\mathbf{k} \in I_{\mathbf{n},\mathbf{n}+\mathbf{1}}$, wobei $\mathbf{1} = (1, \dots, 1) \in \mathbb{Z}^d$ bezeichne. Dann ist $\mathbf{k} \sim \mathbf{k} - \mathbf{n} \in \times_{i=1}^d (0,1]$, so dass wir also o.B.d.A. sogleich annehmen können, dass $K \subseteq \times_{i=1}^d (0,1]$ gilt.

Wenn wir nun annehmen, dass $K \in \mathcal{B}^d$ gilt, erhalten wir mit der gerade gezeigten Translationsinvarianz des Lebesgue-Maßes folgenden Widerspruch:

¹¹ Bezgl. der Notwendigkeit des Auswahlaxioms für den Nachweis der Existenz nicht-Borel'scher Mengen vgl. die Ausführungen in BENEDETTO (1976): Real Variable and Integration, Teubner Verlag, Remark 1, S. 50 und ELSTRODT (1996), S. 98.

I. Maßtheoretische Grundlagen

$$\mathbb{R}^d = \bigoplus_{\mathbf{k} \in K} (\mathbf{k} + \mathbb{Q}^d) = \bigoplus_{\mathbf{k} \in K} \bigoplus_{\mathbf{y} \in \mathbb{Q}^d} \{\mathbf{k} + \mathbf{y}\} = \bigoplus_{\mathbf{y} \in \mathbb{Q}^d} \bigoplus_{\mathbf{k} \in K} \{\mathbf{k} + \mathbf{y}\} = \bigoplus_{\mathbf{y} \in \mathbb{Q}^d} (\mathbf{y} + K) \text{ mit}$$

$$\infty = m^d(\mathbb{R}^d) = m^d\left(\bigoplus_{\mathbf{y} \in \mathbb{Q}^d} (\mathbf{y} + K)\right) = \sum_{\mathbf{y} \in \mathbb{Q}^d} m^d(\mathbf{y} + K) = \sum_{\mathbf{y} \in \mathbb{Q}^d} m^d(K),$$

also muss $m^d(K) > 0$ sein. Andererseits ist $\bigoplus_{\mathbf{y} \in \mathbb{Q}^d \cap \prod_{i=1}^d (0,1]} (\mathbf{y} + K) \subseteq \prod_{i=1}^d (0,2]$, so dass

$$\sum_{\mathbf{y} \in \mathbb{Q}^d \cap \prod_{i=1}^d (0,1]} m^d(K) = \sum_{\mathbf{y} \in \mathbb{Q}^d \cap \prod_{i=1}^d (0,1]} m^d(\mathbf{y} + K) = m^d\left(\bigoplus_{\mathbf{y} \in \mathbb{Q}^d \cap \prod_{i=1}^d (0,1]} (\mathbf{y} + K)\right) \leq m^d\left(\prod_{i=1}^d (0,2]\right) = 2^d < \infty$$

folgt: Widerspruch, da die linke Summe unendlich viele gleich große positive Summanden $m^d(K)$ enthält! Folglich kann K keine Borel'sche Menge sein, womit die Behauptung bewiesen ist. ■

Die gerade nachgewiesene Existenz nicht-Borel'scher Mengen in \mathbb{R}^d kann zu geradezu absonderlichen Paradoxien führen. So ist es nach einem berühmten Satz von Banach und Tarski aus dem Jahre 1924 möglich, eine Vollkugel vom Radius 1 im \mathbb{R}^3 derart in *endlich viele* (!), disjunkte (aber notwendigerweise nicht-Borel'sche) Mengen zu zerlegen und durch geeignete Bewegungen im \mathbb{R}^3 derart disjunkt wieder zusammensetzen, dass dabei *zwei* (!) disjunkte Vollkugeln vom Radius 1 (oder, noch abstruser, 1000 Vollkugeln vom Radius 10^6) herauskommen!

Einen anderen Beweis von Satz 15 findet man z.B. in BILLINGSLEY (1986), S. 26ff. Dort wird in Aufgabe 2.22 bemerkt, dass \mathcal{B}^d zu \mathbb{R} (bzw. äquivalent: zu \mathbb{R}^d) gleichmächtig ist (vgl. auch ELSTRODT (1996), S. 17 und S. 73). Damit kann \mathcal{B}^d aber nicht mit der Potenzmenge von \mathbb{R}^d übereinstimmen, da \mathbb{R}^d und $\mathfrak{P}(\mathbb{R}^d)$ nicht gleichmächtig sind.

Unter Verwendung der eingangs erwähnten *Kontinuumshypothese* kann man ferner noch zeigen, dass ein *endliches* (oder allgemeiner σ -endliches) Maß μ , welches auf ganz $\mathfrak{P}(\mathbb{R}^d)$ definiert ist, notwendig diskret in dem Sinne ist, dass dann eine abzählbare Menge $M \subset \mathbb{R}^d$ existiert mit $\mu(M^c) = 0$. Insbesondere existieren damit auf $\mathfrak{P}(\mathbb{R}^d)$ als Definitionsbereich ausschließlich diskrete Wahrscheinlichkeitsverteilungen (vgl. PLACHKY (1980), Abschnitt 1.1).

Zum Abschluss wollen wir nur noch bemerken, dass sowohl die Borel'sche σ -Algebra als auch das Lebesgue-Maß Produktstruktur besitzen, d.h. es gilt

$$\mathcal{B}^d = \bigotimes_{i=1}^n \mathcal{B}^{d_i} \text{ und } m^d = \bigotimes_{i=1}^n m^{d_i} \text{ für alle } d_1, \dots, d_n \in \mathbb{N} \text{ mit } d = \sum_{i=1}^n d_i, n \in \mathbb{N}.$$

I. 7. Das Lebesgue-Integral

In diesem Abschnitt werden wir in Grundzügen die Lebesgue'sche Integrationstheorie entwickeln, die in der Stochastik u.a. für eine einheitliche Behandlung des Begriffs des *Erwartungswerts* von Zufallsvariablen wichtig ist, der im schon genannten Gesetz der großen Zahlen eine wesentliche Rolle spielt.

Definition 16. Es sei (Ω, \mathcal{A}) ein Messraum und $A \in \mathcal{A}$ ein Ereignis. Die durch

$$\mathbb{1}_A : \omega \mapsto \begin{cases} 1, & \text{falls } \omega \in A \\ 0, & \text{falls } \omega \notin A \end{cases}$$

definierte reellwertige Funktion heißt *Indikatorfunktion* zum Ereignis A .

Die Indikatorfunktion $\mathbb{1}_A$ zum Ereignis $A \in \mathcal{A}$ ist natürlicherweise $(\mathcal{A}-\mathcal{B}^1)$ -messbar, denn es gilt:

$$(\mathbb{1}_A)^{-1}((a, b]) = \begin{cases} \emptyset, & (a, b] \cap \{0, 1\} = \emptyset \\ A, & (a, b] \cap \{0, 1\} = \{1\} \\ A^c, & (a, b] \cap \{0, 1\} = \{0\} \\ \Omega, & (a, b] \supset \{0, 1\} \end{cases} \in \mathcal{A} \text{ für alle reellen } a < b,$$

was nach Lemma 10 zum Nachweis der Messbarkeit ausreicht.

Definition 17. Es sei (Ω, \mathcal{A}) ein Messraum und f eine messbare, reellwertige Funktion auf Ω mit endlichem Wertebereich $f(\Omega)$. Dann heißt f *Elementarfunktion*.

Lemma 16. Es sei (Ω, \mathcal{A}) ein Messraum und f eine $(\mathcal{A}-\mathcal{B}^1)$ -messbare Elementarfunktion auf Ω . Dann besitzt f eine (i. Allg. nicht eindeutig bestimmte) Darstellung

$$f(\omega) = \sum_{i=1}^n a_i \mathbb{1}_{A_i}(\omega), \quad \omega \in \Omega$$

mit $n \in \mathbb{N}$, $a_1, \dots, a_n \in \mathbb{R}$, $A_1, \dots, A_n \in \mathcal{A}$. Umgekehrt ist jede Abbildung f mit einer solchen Darstellung eine Elementarfunktion.

Beweis: Es sei $f(\Omega) = \{a_1, \dots, a_n\} \subset \mathbb{R}$. Wegen der Messbarkeit von f liegen dann die Mengen $A_i := f^{-1}(\{a_i\})$, $i = 1, \dots, n$ in \mathcal{A} und bilden eine disjunkte Zerlegung von Ω . Zu jedem $\omega \in \Omega$ existiert daher ein eindeutig bestimmter Index k mit $\omega \in A_k$, mit $\mathbb{1}_{A_k}(\omega) = 1$ und $\mathbb{1}_{A_i}(\omega) = 0$ für $i \neq k$, so dass

$$f(\omega) = a_k = \sum_{i=1}^n a_i \mathbb{1}_{A_i}(\omega),$$

I. Maßtheoretische Grundlagen

wie behauptet. Umgekehrt ist sofort ersichtlich, dass f höchstens $n+1$ verschiedene Werte (ggf. etwa auch die Null) annimmt, also eine Elementarfunktion ist.

Die im Beweis benutzte eindeutige Darstellung für f , die auch als *kanonische Darstellung* bezeichnet wird, ist aber i. Allg. nicht die einzig mögliche Darstellung von f , jedenfalls dann nicht, wenn Ω mehr als n Elemente enthält und die σ -Algebra \mathcal{A} entsprechend reichhaltig ist. Es existiert dann nämlich wenigstens ein Index k so, dass das Urbild A_k mindestens zwei Elemente enthält, also ggf. eine disjunkte Darstellung $A_k = A_{k1} \oplus A_{k2}$ mit nicht-leeren messbaren Mengen $A_{i1}, A_{i2} \in \mathcal{A}$ erlaubt. Damit ist aber z.B. auch die formal andere Darstellung

$$f = \sum_{\substack{i=1 \\ i \neq k}}^n a_i \mathbb{1}_{A_i} + a_k \mathbb{1}_{A_{k1}} + a_k \mathbb{1}_{A_{k2}}$$

möglich. ■

Satz 16. Es sei (Ω, \mathcal{A}) ein Messraum und $\{f_n\}_{n \in \mathbb{N}}$ eine punktweise nach oben beschränkte Folge nicht-negativer $(\mathcal{A}, \mathcal{B}^1)$ -messbarer Funktionen auf Ω . Dann ist auch $f := \sup_{n \in \mathbb{N}} f_n$ $(\mathcal{A}, \mathcal{B}^1)$ -messbar.

Umgekehrt ist jede nicht-negative reelle $(\mathcal{A}, \mathcal{B}^1)$ -messbare Funktion auf Ω als Supremum einer monoton wachsenden Folge nicht-negativer, punktweise nach oben beschränkter Elementarfunktionen darstellbar.

Beweis: Es ist

$$\begin{aligned} f^{-1}((a, b]) &= \left\{ \omega \in \Omega \mid a < \sup_{n \in \mathbb{N}} f_n(\omega) \leq b \right\} = \left\{ \omega \in \Omega \mid \sup_{n \in \mathbb{N}} f_n(\omega) \leq b \right\} \setminus \left\{ \omega \in \Omega \mid \sup_{n \in \mathbb{N}} f_n(\omega) \leq a \right\} \\ &= \bigcap_{n \in \mathbb{N}} \underbrace{\left\{ \omega \in \Omega \mid f_n(\omega) \leq b \right\}}_{\in \mathcal{A}} \setminus \bigcap_{n \in \mathbb{N}} \underbrace{\left\{ \omega \in \Omega \mid f_n(\omega) \leq a \right\}}_{\in \mathcal{A}} \in \mathcal{A} \end{aligned}$$

für alle reellen $a < b$ (beachte: $(-\infty, b] = \bigcup_{k \in \mathbb{N}} (-k + b, b] \in \mathcal{B}^1$). Dies zeigt den ersten Teil.

Für die Umkehrung wählen wir die Elementarfunktionen

$$f_n := \sum_{i=1}^{n2^n} \frac{i-1}{2^n} \mathbb{1}_{A_{in}} + n \mathbb{1}_{B_n} \quad \text{mit}$$

$$A_{in} := \left\{ \omega \in \Omega \mid \frac{i-1}{2^n} < f(\omega) \leq \frac{i}{2^n} \right\} \quad \text{für } 1 \leq i \leq n2^n \quad \text{und} \quad B_n := \{ \omega \in \Omega \mid f(\omega) > n \}, \quad n \in \mathbb{N}.$$

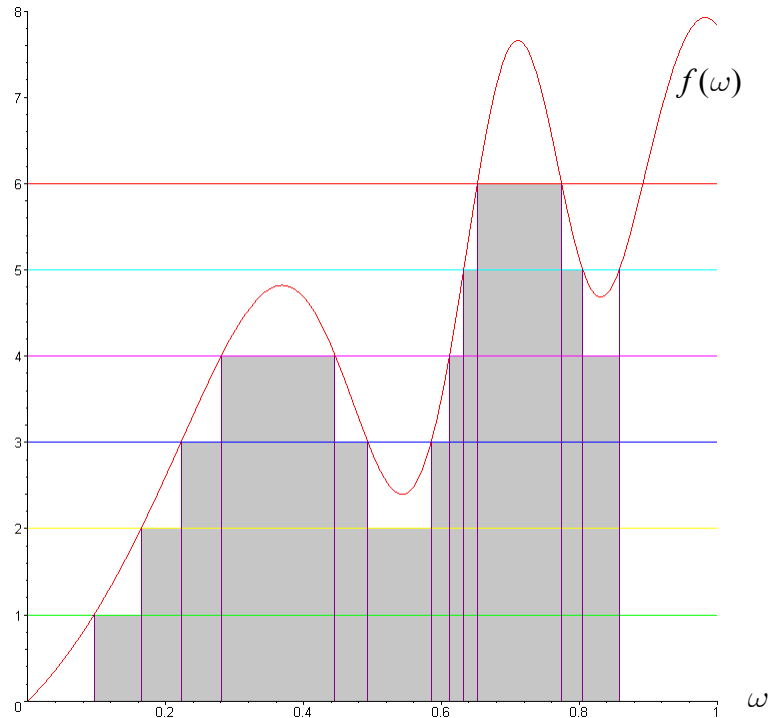
Aufgrund der Messbarkeit von f sind alle Mengen A_{in} und B_n Elemente von \mathcal{A} , mit

$$0 \leq f_n \leq f \quad \text{und} \quad \sup_{n \in \mathbb{N}} f_n = f,$$

wie behauptet. ■

I. Maßtheoretische Grundlagen

Satz 16 besagt also, dass die nicht-negativen $(\mathcal{A}-\mathcal{B}^1)$ -messbaren Abbildungen genau die sind, die sich als Suprema monoton wachsender Folgen von Elementarfunktionen darstellen lassen. Die nachfolgende Skizze zeigt, wie die monotone Approximation im Prinzip funktioniert.



Lemma 17. Es sei (Ω, \mathcal{A}) ein Messraum und f eine reelle, $(\mathcal{A}-\mathcal{B}^1)$ -messbare Abbildung auf Ω . Dann sind der Positivteil $f^+ := \max\{0, f\}$ und der Negativteil $f^- := -\min\{0, f\} = (-f)^+$ $\mathcal{A}-\mathcal{B}^1$ -messbare Funktionen auf Ω , mit $f = f^+ - f^-$.

Beweis: Es ist

$$\max\{a, b\} = \frac{a + b + |a - b|}{2} \quad \text{und} \quad \min\{a, b\} = \frac{a + b - |a - b|}{2} \quad \text{für } a, b \in \mathbb{R}.$$

Damit sind $\max\{\cdot, \cdot\}$ und $\min\{\cdot, \cdot\}$ stetige und damit $\mathcal{B}^2-\mathcal{B}^1$ -messbare Funktionen¹² zweier Variabler, so dass f^+ und f^- als Komposition messbarer Abbildungen messbar sind. Die Beziehung $f = f^+ - f^-$ ergibt sich durch Fallunterscheidung. ■

¹² Eine Abbildung $\mathbb{R}^2 \rightarrow \mathbb{R}^1$ ist genau dann stetig, wenn die Urbilder offener Mengen offen sind. Die Messbarkeit folgt somit aus Satz 13.

I. Maßtheoretische Grundlagen

Satz 17. Es sei (Ω, \mathcal{A}) ein Messraum, f, g und $\{f_n\}_{n \in \mathbb{N}}$ seien reelle $(\mathcal{A}, \mathcal{B}^1)$ -messbare Abbildungen. Dann sind auch folgende Abbildungen messbar:

- a) $f + g, f - g, f \cdot g, \frac{f}{g}$ (falls $g \neq 0$), $\max(f, g), \min(f, g)$,
- b) $\sup_{n \in \mathbb{N}} f_n, \inf_{n \in \mathbb{N}} f_n$ (falls $\{f_n\}_{n \in \mathbb{N}}$ punktweise nach oben bzw. nach unten beschränkt),
- c) $\limsup_{n \rightarrow \infty} f_n, \liminf_{n \rightarrow \infty} f_n$, und $\lim_{n \rightarrow \infty} f_n$ (falls $\{f_n\}_{n \in \mathbb{N}}$ punktweise beschränkt bzw. konvergent),
- d) $\sum_{n \in \mathbb{N}} f_n$ (falls absolut konvergent).

Beweis: Teil a) folgt aus der Tatsache, dass für jede stetige (und damit messbare) Abbildung $H: \mathbb{R}^2 \rightarrow \mathbb{R}$ auch $H(f, g)$ als Komposition messbarer Abbildungen messbar ist. Die angegebenen Verknüpfungen $+, -, \times, \div, \max, \min$ sind aber stetig auf \mathbb{R}^2 . Teil b) folgt aus Satz 16 mit $\inf_{n \in \mathbb{N}} f_n = -\sup_{n \in \mathbb{N}} (-f_n)$. Für Teil c) argumentiert man so: es ist $\limsup_{n \rightarrow \infty} f_n = \inf_{m \in \mathbb{N}} \sup_{n \geq m} f_n$ und $\liminf_{n \rightarrow \infty} f_n = \sup_{m \in \mathbb{N}} \inf_{n \geq m} f_n$, woraus der erste Teil der Behauptung nach Teil b) folgt. Ist ferner $f := \lim_{n \rightarrow \infty} f_n$ der punktweise Grenzwert der Folge, so ist auch $f = \limsup_{n \rightarrow \infty} f_n = \liminf_{n \rightarrow \infty} f_n$, also f messbar. Teil d) ist ein Spezialfall von c) für Partialsummenfolgen. ■

Gelegentlich ist eine Erweiterung des Wertebereichs messbarer Abbildungen von \mathbb{R} auf $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ notwendig. In diesem Fall sind die üblichen Rechenregeln geeignet zu modifizieren (vgl. etwa ELSTRODT (1996), Kapitel III, §4), und statt \mathcal{B}^1 ist die größere σ -Algebra $\bar{\mathcal{B}}^1 = \sigma(\mathcal{B}^1 \cup \{\{-\infty\}, \{\infty\}\})$ zu Grunde zu legen; entsprechend für höhere Dimensionen. Satz 17 b) ist dann ohne Einschränkung gültig, Satz 17 d) ohne Einschränkung für $f_n \geq 0$. Satz 16 und Lemma 17 gelten sinngemäß.

Wir kommen jetzt zum formalen Aufbau des Lebesgue-Integrals, der sich in drei konsekutiven Schritten vollzieht.

Definition 18. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f eine Elementarfunktion mit einer Darstellung

$$f(\omega) = \sum_{i=1}^n a_i \mathbb{1}_{A_i}(\omega), \quad \omega \in \Omega$$

wie in Lemma 16. Dann ist das μ -Integral (Lebesgue-Integral) von f definiert durch

$$\int f d\mu := \sum_{i=1}^n a_i \mu(A_i).$$

Insbesondere gilt

$$\int \mathbb{1}_A d\mu = \mu(A) \quad \text{für alle } A \in \mathcal{A}.$$

I. Maßtheoretische Grundlagen

Lemma 18. Unter den Voraussetzungen von Definition 18 gilt:

- a) Das μ -Integral ist wohldefiniert, d.h. es ist unabhängig von der speziellen Darstellung von f .
 b) Das μ -Integral ist linear und monoton auf der Menge der Elementarfunktionen, d.h. es gilt

$$\int \alpha f + \beta g \, d\mu = \alpha \int f \, d\mu + \beta \int g \, d\mu$$

für alle $\alpha, \beta \in \mathbb{R}$ und Elementarfunktionen f, g sowie

$$\int f \, d\mu \leq \int g \, d\mu \text{ für } f \leq g.$$

- c) Das μ -Integral ist auch linear und monoton auf der Menge der Maße auf \mathcal{A} , d.h. ist ν ein weiteres Maß auf \mathcal{A} , so gilt

$$\int f \, d(\alpha\mu + \beta\nu) = \alpha \int f \, d\mu + \beta \int f \, d\nu$$

für alle $\alpha, \beta \geq 0$ und Elementarfunktionen f sowie

$$\int f \, d\nu \leq \int f \, d\mu \text{ für } \nu \leq \mu \text{ und } f \geq 0.$$

Beweis:

- a) Sei $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i} = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$ mit $a_1, \dots, a_n, b_1, \dots, b_m \in \mathbb{R}$, $A_1, \dots, A_n, B_1, \dots, B_m \in \mathcal{A}$, $m \in \mathbb{N}$. Wir setzen $A_{n+j} := B_j$ für $j = 1, \dots, m$ und betrachten die Menge $\mathcal{D} := \left\{ \bigcap_{i=1}^{n+m} C_i \mid C_i \in \{A_i, A_i^c\} \right\}$ aller Durchschnitte aus A_i -Mengen und deren Komplementen. Die Elemente von \mathcal{D} sind dann offensichtlich paarweise disjunkt, und f besitzt genau eine Darstellung $f = \sum_{k=1}^K c_k \mathbb{1}_{D_k}$ mit reellen Koeffizienten c_k , $k = 1, \dots, K = 2^{n+m}$ und $D_k \in \mathcal{D}$. Es folgt

$$c_k = \sum_{i: D_k \subseteq A_i} a_i \text{ für } k = 1, \dots, K \text{ und somit}$$

$$\sum_{i=1}^n a_i \mu(A_i) = \sum_{i=1}^n a_i \sum_{k: D_k \subseteq A_i} \mu(D_k) = \sum_{k=1}^K \left(\sum_{i: D_k \subseteq A_i} a_i \right) \mu(D_k) = \sum_{k=1}^K c_k \mu(D_k).$$

Aus Symmetriegründen folgt analog

$$\sum_{j=1}^m b_j \mu(B_j) = \sum_{k=1}^K c_k \mu(D_k)$$

und damit die Behauptung.

- b) Sei $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$, $g = \sum_{j=1}^m b_j \mathbb{1}_{B_j}$. Dann ist $\alpha f + \beta g = \sum_{i=1}^n \alpha a_i \mathbb{1}_{A_i} + \sum_{j=1}^m \beta b_j \mathbb{1}_{B_j}$, also ist

I. Maßtheoretische Grundlagen

$$\int \alpha f + \beta g d\mu = \sum_{i=1}^n \alpha a_i \mu(A_i) + \sum_{j=1}^m \beta b_j \mu(B_j) = \alpha \int f d\mu + \beta \int g d\mu.$$

Die zweite Behauptung ergibt sich aus folgender Überlegung: die Differenz $g - f$ ist ebenfalls eine Elementarfunktion, etwa der Form $g - f = \sum_{j=1}^k c_j \mathbb{1}_{C_j}$ mit $c_j \geq 0$, $j = 1, \dots, k \in \mathbb{N}$. Also ist

$$\int g d\mu - \int f d\mu = \int (g - f) d\mu = \sum_{j=1}^k c_j \mu(C_j) \geq 0, \text{ woraus die zweite Behauptung folgt.}$$

- c) Es ist unmittelbar einsichtig, dass mit μ und ν auch $\alpha\mu + \beta\nu$ ein Maß ist, mit der Eigenschaft $(\alpha\mu + \beta\nu)(A) = \alpha\mu(A) + \beta\nu(A)$ für alle $A \in \mathcal{A}$. Es folgt

$$\int f d(\alpha\mu + \beta\nu) = \sum_{i=1}^n a_i (\alpha\mu + \beta\nu)(A_i) = \sum_{i=1}^n (\alpha a_i \mu(A_i) + \beta a_i \nu(A_i)) = \alpha \int f d\mu + \beta \int f d\nu.$$

Die zweite Behauptung folgt aus der Überlegung, dass bei $\nu \leq \mu$ und $f \geq 0$ gilt: $a_1, \dots, a_n \geq 0$ und somit

$$\int f d\nu = \sum_{i=1}^n a_i \nu(A_i) \leq \sum_{i=1}^n a_i \mu(A_i) = \int f d\mu. \quad \blacksquare$$

Lemma 19. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f eine nicht-negative reelle $(\mathcal{A}-\mathcal{B}^1)$ -messbare Abbildung. Dann gilt: $\sup_{n \in \mathbb{N}} \int g_n d\mu = \sup_{n \in \mathbb{N}} \int h_n d\mu$ für alle monoton wachsenden Folgen $\{g_n\}_{n \in \mathbb{N}}$, $\{h_n\}_{n \in \mathbb{N}}$ nicht-negativer Elementarfunktionen mit $f = \sup_{n \in \mathbb{N}} g_n = \sup_{n \in \mathbb{N}} h_n$.

Beweis: Wir zeigen zunächst, dass $\int g_n d\mu \leq \sup_{k \in \mathbb{N}} \int h_k d\mu$ für jedes $n \in \mathbb{N}$ gilt; entsprechend ist

dann auch $\int h_n d\mu \leq \sup_{k \in \mathbb{N}} \int g_k d\mu$ für jedes $n \in \mathbb{N}$. Sei etwa g_n von der Form $g_n = \sum_{j=1}^{m_n} b_{jn} \mathbb{1}_{B_{jn}}$.

Für $\alpha \in (0, 1)$ und $k \in \mathbb{N}$ sei ferner

$$C_{kn} := \{\omega \in \Omega \mid h_k(\omega) \geq \alpha g_n(\omega)\} = (\{\omega \in \Omega \mid h_k(\omega) \geq 0\} \cap \{\omega \in \Omega \mid g_n(\omega) = 0\}) \cup \dots \\ \dots \cup \left\{ \left\{ \omega \in \Omega \mid \frac{h_k}{g_n}(\omega) \geq \alpha \right\} \cap \{\omega \in \Omega \mid g_n(\omega) > 0\} \right\}.$$

Dann ist $C_{kn} \in \mathcal{A}$ für alle $k \in \mathbb{N}$. Die Monotonie der Folge $\{h_k\}_{k \in \mathbb{N}}$ impliziert aber die Monotonie der Mengenfolge $\{C_{kn}\}_{k \in \mathbb{N}}$ mit $\bigcup_{k=1}^{\infty} C_{kn} = \Omega$ bzw. $\bigcup_{k=1}^{\infty} (B_{jn} \cap C_{kn}) = B_{jn}$. Ferner ist $h_k \geq \alpha g_n \mathbb{1}_{C_{kn}}$ mit der Schlussfolgerung

I. Maßtheoretische Grundlagen

$$\int g_n d\mu = \sum_{j=1}^{m_n} b_{jn} \mu(B_{jn}) = \lim_{k \rightarrow \infty} \sum_{j=1}^{m_n} b_{jn} \mu(B_{jn} \cap C_{kn}) = \sup_{k \in \mathbb{N}} \sum_{j=1}^{m_n} b_{jn} \mu(B_{jn} \cap C_{kn}) = \sup_{k \in \mathbb{N}} \int g_n \mathbb{1}_{C_{kn}} d\mu,$$

woraus

$$\sup_{k \in \mathbb{N}} \int h_k d\mu \geq \sup_{k \in \mathbb{N}} \int \alpha g_n \mathbb{1}_{C_k} d\mu = \alpha \int g_n d\mu$$

folgt. Da $\alpha \in (0,1)$ beliebig war, folgt somit zunächst $\int g_n d\mu \leq \sup_{k \in \mathbb{N}} \int h_k d\mu$ für jedes $n \in \mathbb{N}$, wie behauptet. Damit erhalten wir aber

$$\sup_{n \in \mathbb{N}} \int g_n d\mu \leq \sup_{k \in \mathbb{N}} \int h_k d\mu \leq \sup_{\ell \in \mathbb{N}} \int g_\ell d\mu$$

und damit die Aussage des Lemmas. ■

Definition 19. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f eine nicht-negative reelle, $(\mathcal{A}-\mathcal{B}^1)$ -messbare Abbildung. Dann ist das nach Lemma 19 von der approximierenden Folge unabhängige μ -Integral von f definiert durch

$$\int f d\mu := \sup_{n \in \mathbb{N}} \int g_n d\mu,$$

wobei $\{g_n\}_{n \in \mathbb{N}}$ eine beliebige monoton wachsende Folge nicht-negativer Elementarfunktionen mit $f = \sup_{n \in \mathbb{N}} g_n$ ist.

Definition 20. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und f eine beliebige reelle, $(\mathcal{A}-\mathcal{B}^1)$ -messbare Abbildung. Dann ist das μ -Integral von f definiert durch

$$\int f d\mu := \int f^+ d\mu - \int f^- d\mu,$$

sofern mindestens eines der beiden rechten Integrale *endlich* ist. In diesem Fall heißt f μ -*integrierbar*. Sind beide rechten Integrale endlich, so heißt f *eigentlich μ -integrierbar*. Ist ferner $A \in \mathcal{A}$, so bezeichnet

$$\int_A f d\mu := \int f \cdot \mathbb{1}_A d\mu.$$

Alternativ schreibt man auch: $\int f d\mu = \int f(\omega) \mu(d\omega)$; diese Schreibweise ist insbesondere bei Produktmaßen von Vorteil, weil damit eine eventuelle Integrationsreihenfolge eindeutig bestimmt ist.

Wir listen im Folgenden kurz die wichtigsten Eigenschaften des Lebesgue-Integrals, die sich relativ problemlos zeigen lassen, nur mit Beweisskizze auf.

I. Maßtheoretische Grundlagen

Lemma 20. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum, f und g seien reelle, $(\mathcal{A}-\mathcal{B}^1)$ -messbare, μ -integrierbare Abbildungen. Dann gilt:

- a) $\int \alpha f + \beta g d\mu = \alpha \int f d\mu + \beta \int g d\mu$ für alle $\alpha, \beta \in \mathbb{R}$ (im Falle der Existenz)
- b) $f \leq g \Rightarrow \int f d\mu \leq \int g d\mu$
- c) $\left| \int f d\mu \right| \leq \int |f| d\mu$
- d) $\forall A \in \mathcal{A}: \mu(A) = 0 \Rightarrow \int_A f d\mu = 0$
- e) $f \geq 0$ und $\int f d\mu = 0 \Rightarrow f = 0$ μ -fast überall
- f) $f = g$ μ -fast überall $\Rightarrow \int f d\mu = \int g d\mu$
- g) $\forall A \in \mathcal{A}: \int_A f d\mu \leq \int_A g d\mu \Rightarrow f \leq g$ μ -fast überall
- h) $\forall A \in \mathcal{A}: \int_A f d\mu = \int_A g d\mu \Rightarrow f = g$ μ -fast überall
- i) $\forall \varepsilon > 0: \mu(\{\omega \in \Omega \mid |f(\omega)| > \varepsilon\}) \leq \frac{1}{\varepsilon} \int |f| d\mu$ (Markoff-Ungleichung).

Beweisskizze: Teil a) folgt sofort aus der Definition des Integrals. Für Teil b) betrachte man $g - f \geq 0$, so dass $\int (g - f) d\mu \geq 0$. Teil c) folgt aus der Darstellung $|f| = f^+ + f^-$. Für Teil d) sei $\{f_n\}_{n \in \mathbb{N}}$ eine Folge nicht-negativer Elementarfunktionen mit $|f| = \sup_{n \in \mathbb{N}} f_n$. Dann ist jedes f_n

beschränkt, etwa durch $M_n \geq 0$. Es folgt $\left| \int_A f_n d\mu \right| \leq \int_A |f_n| d\mu \leq M_n \mu(A) = 0$, also $\int_A f_n d\mu = 0$ für alle $n \in \mathbb{N}$ und damit auch $\int_A |f| d\mu = \sup_{n \in \mathbb{N}} \int_A f_n d\mu = 0$, also $\int_A f d\mu = 0$ nach Teil c). Für

Teil e) wähle $A_n := \left\{ \omega \in \Omega \mid f(\omega) > \frac{1}{n} \right\} \in \mathcal{A}$, dann gilt $0 = \int f d\mu \geq \int_{A_n} f d\mu \geq \frac{1}{n} \mu(A_n)$, also

$\mu(A_n) = 0$ für alle $n \in \mathbb{N}$ und damit auch $\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$, woraus $f = 0$ μ -fast überall folgt. Für

Teil f) wähle $A := \{\omega \in \Omega \mid f(\omega) \neq g(\omega)\} = \{\omega \in \Omega \mid (f - g)(\omega) \neq 0\} \in \mathcal{A}$ mit $\mu(A) = 0$; die Aussage folgt jetzt aus Teil d) wegen $\int (f - g) d\mu = \int_A (f - g) d\mu = 0$. Für Teil g) wähle

$A_n := \left\{ \omega \in \Omega \mid f(\omega) - g(\omega) > \frac{1}{n} \right\} \in \mathcal{A}$, dann ist $\mu(A_n) = 0$ wegen $\int_{A_n} f d\mu \geq \int_{A_n} \left(g + \frac{1}{n}\right) d\mu =$

$\int_{A_n} g d\mu + \frac{1}{n} \mu(A_n) \geq \int_{A_n} f d\mu + \frac{1}{n} \mu(A_n)$, woraus die Behauptung wie unter e) folgt. Teil h) folgt

aus Teil g), da $\int_A f d\mu = \int_A g d\mu$ mit $\int_A f d\mu \leq \int_A g d\mu$ und $\int_A g d\mu \leq \int_A f d\mu$ gleichbedeutend ist.

I. Maßtheoretische Grundlagen

Die Markoff-Ungleichung (Teil c)) folgt mit der Wahl $A := \{\omega \in \Omega \mid |f(\omega)| > \varepsilon\} \in \mathcal{A}$ aus $\int |f| d\mu \geq \int_A |f| d\mu \geq \varepsilon \int_A d\mu = \varepsilon \mu(A)$. ■

Wir zeigen jetzt einen der wichtigsten Konvergenzsätze der Integrationstheorie, der auf Beppo Levi aus dem Jahr 1906 zurückgeht.

Satz 18 (Satz von der monotonen Konvergenz). Es sei $\{f_n\}_{n \in \mathbb{N}}$ eine monoton wachsende Folge nicht-negativer reeller $(\mathcal{A}, \mathcal{B}^1)$ -messbarer Abbildungen auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$. Dann gilt:

$$\int \lim_{n \rightarrow \infty} f_n d\mu = \int \sup_{n \in \mathbb{N}} f_n d\mu = \sup_{n \in \mathbb{N}} \int f_n d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu.$$

Beweis: Zu jedem f_n existiert eine monoton approximierende Folge $\{f_{nk}\}_{k \in \mathbb{N}}$ nicht-negativer Elementarfunktionen mit $f_n = \sup_{k \in \mathbb{N}} f_{nk}$. Dann ist aber auch $g_n := \max\{f_{n1}, \dots, f_{nm}\}$ für jedes $n \in \mathbb{N}$ eine nicht-negative Elementarfunktion (die Messbarkeit ergibt sich etwa aus Satz 17 a)), die ebenfalls in n monoton wachsend ist mit $g_n \leq f_n$ und somit auch $g_n \leq f := \sup_{k \in \mathbb{N}} f_k$ für alle $n \in \mathbb{N}$. Andererseits ist aber auch $f_{nk} \leq g_n \leq \sup_{m \in \mathbb{N}} g_m$ für alle $n \geq k$ und somit $f_n = \sup_{k \in \mathbb{N}} f_{nk} \leq \sup_{m \in \mathbb{N}} g_m$, also $f = \sup_{n \in \mathbb{N}} f_n \leq \sup_{m \in \mathbb{N}} g_m \leq f$ und damit $f = \sup_{m \in \mathbb{N}} g_m$. Insgesamt folgt nach Definition des μ -Integrals

$$\begin{aligned} \int \sup_{n \in \mathbb{N}} g_n d\mu &\leq \sup_{n \in \mathbb{N}} \int \sup_{k \in \mathbb{N}} f_k d\mu = \int \sup_{k \in \mathbb{N}} f_k d\mu = \int f d\mu = \sup_{n \in \mathbb{N}} \int g_n d\mu \\ &\leq \sup_{n \in \mathbb{N}} \int \sup_{k \in \mathbb{N}} g_k d\mu = \int \sup_{k \in \mathbb{N}} g_k d\mu \end{aligned}$$

und damit überall Gleichheit, so dass sich insbesondere

$$\int f d\mu = \sup_{k \in \mathbb{N}} \int g_k d\mu \leq \sup_{k \in \mathbb{N}} \int f_k d\mu \leq \sup_{k \in \mathbb{N}} \int \sup_{n \in \mathbb{N}} f_n d\mu = \int \sup_{n \in \mathbb{N}} f_n d\mu = \int f d\mu,$$

also

$$\int \sup_{n \in \mathbb{N}} f_n d\mu = \sup_{n \in \mathbb{N}} \int f_n d\mu$$

ergibt, wie behauptet. ■

Die Voraussetzung der Monotonie in Satz 18 kann nicht fallen gelassen werden, wie das folgende Beispiel für den Maßraum $(\mathbb{R}^1, \mathcal{B}^1, m^1)$ zeigt: wählen wir $f_n := \frac{1}{n} \mathbb{1}_{[0, n]}$, so konvergiert $\{f_n\}_{n \in \mathbb{N}}$ auf ganz \mathbb{R}^1 gleichmäßig gegen 0, d.h. es gilt $\lim_{n \rightarrow \infty} f_n = 0$, es ist aber $\int f_n d m^1 = \frac{1}{n} m^1([0, n]) \equiv 1$.

I. Maßtheoretische Grundlagen

Eine einfache Folgerung aus Satz 18 für Reihen von Funktionen ist

Lemma 21. Es sei $\{f_n\}_{n \in \mathbb{N}}$ eine Folge nicht-negativer reeller $(\mathcal{A}-\mathcal{B}^1)$ -messbarer Abbildungen auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$. Dann gilt:

$$\int \sum_{n=1}^{\infty} f_n d\mu = \sum_{n=1}^{\infty} \int f_n d\mu.$$

Beweis: Folgt aus Satz 18 für die Partialsummenfolge $\left\{ \sum_{n=1}^m f_n \right\}_{m \in \mathbb{N}}$. ■

Eine natürliche Verallgemeinerung des Satzes von der monotonen Konvergenz stammt von Pierre Fatou aus dem Jahre 1906.

Lemma 22 (Lemma von Fatou). Es sei $\{f_n\}_{n \in \mathbb{N}}$ eine Folge nicht-negativer reeller $(\mathcal{A}-\mathcal{B}^1)$ -messbarer Abbildungen auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$. Dann gilt:

$$\int \liminf_{n \rightarrow \infty} f_n d\mu \leq \liminf_{n \rightarrow \infty} \int f_n d\mu.$$

Beweis: Wegen $\liminf_{n \rightarrow \infty} f_n = \sup_{n \in \mathbb{N}} \inf_{m \geq n} f_m$ folgt, da die Folge $\left\{ \inf_{m \geq n} f_m \right\}_{n \in \mathbb{N}}$ monoton wachsend ist, mit dem Satz von der monotonen Konvergenz:

$$\begin{aligned} \int \liminf_{n \rightarrow \infty} f_n d\mu &= \int \sup_{n \in \mathbb{N}} \inf_{m \geq n} f_m d\mu = \sup_{n \in \mathbb{N}} \int \inf_{m \geq n} f_m d\mu = \sup_{n \in \mathbb{N}} \inf_{k \geq n} \int \inf_{m \geq n} f_m d\mu \\ &\leq \sup_{n \in \mathbb{N}} \inf_{k \geq n} \int f_k d\mu = \liminf_{n \rightarrow \infty} \int f_n d\mu, \end{aligned}$$

womit alles gezeigt ist. ■

Ein weiterer wichtiger Satz über die Vertauschbarkeit von Limes und Integral stammt von Henri Lebesgue aus dem Jahr 1910.

Satz 19 (Satz von der majorisierten Konvergenz). Es sei $\{f_n\}_{n \in \mathbb{N}}$ eine Folge reeller $(\mathcal{A}-\mathcal{B}^1)$ -messbarer Abbildungen auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$, die punktweise gegen die (messbare) Abbildung f konvergiere. Ferner existiere eine nicht-negative $(\mathcal{A}-\mathcal{B}^1)$ -messbare und eigentlich μ -integrierbare Abbildung g auf $(\Omega, \mathcal{A}, \mu)$ mit $|f_n| \leq g$ für alle $n \in \mathbb{N}$. Dann gilt:

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int \lim_{n \rightarrow \infty} f_n d\mu \quad \text{und} \quad \lim_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

I. Maßtheoretische Grundlagen

Beweis: Setze $g_n := |f| + g - |f_n - f| \geq 0$. Mit dem gerade gezeigten Lemma von Fatou ergibt sich

$$0 \leq \int (|f| + g) d\mu = \int \liminf_{n \rightarrow \infty} g_n d\mu \leq \liminf_{n \rightarrow \infty} \int g_n d\mu = \int (|f| + g) d\mu - \limsup_{n \rightarrow \infty} \int |f_n - f| d\mu$$

(beachte: es ist $0 \leq \int (|f| + g) d\mu \leq 2 \int g d\mu < \infty$), also gilt

$$\lim_{n \rightarrow \infty} \int |f_n - f| d\mu = \limsup_{n \rightarrow \infty} \int |f_n - f| d\mu = 0.$$

Mit Lemma 20 c) ergibt sich nun auch die restliche Aussage. ■

Es soll hier noch darauf hingewiesen werden, dass aufgrund von Lemma 20, Teile d) bis h) die Voraussetzungen in Satz 19 abgeschwächt werden können zu den Bedingungen " $f = \lim_{n \rightarrow \infty} f_n$ μ -fast überall" und " $|f| \leq g$ μ -fast überall".

Auf die einschränkenden Voraussetzungen in Satz 19 kann nicht verzichtet werden, wie folgendes Gegenbeispiel zeigt: Sei $f_n = n \cdot \mathbb{1}_{(0, 1/n]}$, $n \in \mathbb{N}$. Dann konvergiert $\{f_n\}_{n \in \mathbb{N}}$ zwar punktweise gegen

$f = 0$, aber es ist $\int f_n d\mathbb{m}^1 = n \cdot \mathbb{m}^1\left(\left(0, \frac{1}{n}\right]\right) = 1 \neq 0 = \int f d\mathbb{m}^1$ für alle $n \in \mathbb{N}$. Man beachte, dass

hier die Abbildung $g = \sup_{n \in \mathbb{N}} |f_n| = \sum_{n=1}^{\infty} n \cdot \mathbb{1}_{\left(\frac{1}{n+1}, \frac{1}{n}\right]}$ zwar $(\mathcal{A}-\mathcal{B}^1)$ -messbar ist, nicht jedoch eigentlich

\mathbb{m}^1 -integrierbar wegen $\int g d\mathbb{m}^1 = \sum_{n=1}^{\infty} n \cdot \int \mathbb{1}_{\left(\frac{1}{n+1}, \frac{1}{n}\right]} d\mathbb{m}^1 = \sum_{n=1}^{\infty} n \cdot \mathbb{m}^1\left(\left(\frac{1}{n+1}, \frac{1}{n}\right]\right) = \sum_{n=1}^{\infty} \frac{1}{n+1} = \infty$.

Es gibt in diesem Fall also keine eigentlich \mathbb{m}^1 -integrierbare Majorante.

Mit Hilfe von (Lebesgue-)Integralen lassen sich selbst wieder Maße definieren, die vor allem in der Stochastik eine große Rolle spielen. Dies liegt an folgendem Sachverhalt, der sich sofort aus Lemma 21 ergibt: ist nämlich f eine nicht-negative, $\mathcal{A}-\mathcal{B}^1$ -messbare Abbildung auf einem Maßraum $(\Omega, \mathcal{A}, \mu)$, so gilt:

$$\int_{\bigoplus_{n=1}^{\infty} A_n} f d\mu = \int \sum_{n=1}^{\infty} f \cdot \mathbb{1}_{A_n} d\mu = \sum_{n=1}^{\infty} \int f \cdot \mathbb{1}_{A_n} d\mu = \sum_{n=1}^{\infty} \int_{A_n} f d\mu$$

für alle Folgen paarweise disjunkter Mengen $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$.

I. Maßtheoretische Grundlagen

Definition 21 (Maße mit Dichten). Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$ eine nicht-negative, messbare Abbildung. Dann wird durch

$$\nu(A) := \int_A f d\mu \quad \text{für alle } A \in \mathcal{A}$$

ein Maß auf \mathcal{A} definiert. Die Abbildung f heißt hier (eine) μ -Dichte von ν .

Die σ -Additivität von ν wurde gerade gezeigt; ferner ist nach Lemma 20 f) und c) die Mengenfunktion ν nicht-negativ mit $\nu(\emptyset) = 0$. Man beachte, dass man hier nicht von "der" Dichte sprechen kann, weil jede andere nicht-negative Abbildung $g : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$ mit $g = f$ μ -fast überall das Selbe leistet.

Eine wichtige Rechenregel für Integrale nach Maßen mit Dichten gibt

Satz 20. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $g : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$ eine nicht-negative, messbare Abbildung und ν das Maß mit Dichte g gemäß Definition 21. Dann gilt:

$$\int f d\nu = \int fg d\mu$$

für alle nicht-negativen, messbaren Abbildungen $f : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$. Die Gleichung bleibt auch dann richtig, wenn f beliebig reellwertig und ν -integrierbar bzw. fg μ -integrierbar ist.

Beweis: Sei zunächst $f = \mathbb{1}_A$ mit $A \in \mathcal{A}$. Dann ist

$$\int f d\nu = \nu(A) = \int_A g d\mu = \int fg d\mu.$$

Mit der Linearität des Integrals ergibt sich, dass diese Gleichung dann auch für beliebige Elementarfunktionen richtig bleibt. Ist f nun nicht-negativ und messbar, so gibt es nach Satz 16 eine monoton wachsende Folge nicht-negativer Elementarfunktionen $\{f_n\}_{n \in \mathbb{N}}$ mit $f = \sup_{n \in \mathbb{N}} f_n$. Nach dem Satz

18 von der monotonen Konvergenz folgt nun wegen $g \geq 0$

$$\int f d\nu = \sup_{n \in \mathbb{N}} \int f_n d\nu = \sup_{n \in \mathbb{N}} \int f_n g d\mu = \int \sup_{n \in \mathbb{N}} \{f_n g\} d\mu = \int fg d\mu,$$

wie behauptet. Die restliche Aussage folgt durch Zerlegung von f in Positiv- und Negativteil. ■

Eine sehr wichtige Charakterisierung von Maßen, die Dichten in dem allgemeinen Sinn von Definition 21 besitzen, gibt der folgende Satz, den wir hier aber nicht in voller Allgemeinheit beweisen werden.

I. Maßtheoretische Grundlagen

Satz 21 (Satz von Radon und Nikodym, 1930). Es sei (Ω, \mathcal{A}) ein Messraum; ferner seien μ und ν Maße auf \mathcal{A} mit der Eigenschaft: μ ist σ -endlich, und es gilt

$$\forall A \in \mathcal{A}: \mu(A) = 0 \Rightarrow \nu(A) = 0$$

(man sagt auch, μ dominiert ν .) Dann besitzt ν eine Dichte f bezüglich μ . f heißt auch Radon-Nikodym-Ableitung von ν bezgl. μ , in Zeichen: $f = \frac{d\nu}{d\mu}$.

Beweis für σ -Algebren mit abzählbarem, erschöpfendem Atomsystem $\mathcal{T} = \{T_i \mid i \in I\}$ (also mit abzählbarer Indexmenge I): die σ -Endlichkeit von μ impliziert, dass $\mu(T_i) < \infty$ ist für alle $i \in I$. Für das Nullmaß $\mu \equiv 0$, also $\mu(T_i) = 0$ für alle $i \in I$ ist nichts zu zeigen, weil in diesem Fall jede nicht-negative messbare Abbildung f eine geeignete Dichte ist. Sei also μ vom Nullmaß verschieden. Dann ist $J := \{i \in I \mid \mu(T_i) > 0\}$ nicht-leer. Definiere

$$f := \sum_{j \in J} \frac{\nu(T_j)}{\mu(T_j)} \mathbb{1}_{T_j}.$$

Dann gilt: für $A = \bigoplus_{k \in K} T_k \in \mathcal{A}$ mit $K \subseteq I$ ist

$$\begin{aligned} \nu(A) &= \sum_{k \in K} \nu(T_k) = \sum_{k \in K \cap J} \nu(T_k) = \sum_{k \in K \cap J} \frac{\nu(T_k)}{\mu(T_k)} \mu(T_k) = \sum_{k \in K \cap J} \frac{\nu(T_k)}{\mu(T_k)} \int \mathbb{1}_{T_k} d\mu \\ &= \int f \cdot \mathbb{1}_A d\mu = \int_A f d\mu, \end{aligned}$$

wie behauptet. ■

Für einen allgemeinen Beweis vgl. BAUER (1992), Satz 17.10 oder ELSTRODT (1996), Kapitel VII, Satz 2.3.

Eine weitere wichtige Rechenregel, die als verallgemeinerte Substitutionsregel für Lebesgue-Integrale angesehen werden kann, gibt

Satz 22. Es sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und (Ξ, \mathcal{C}) ein Messraum sowie $T: (\Omega, \mathcal{A}) \rightarrow (\Xi, \mathcal{C})$ eine messbare Abbildung. Dann gilt:

$$\int g \circ T d\mu = \int g d\mu^T$$

für alle nicht-negativen, messbaren Abbildungen $g: (\Xi, \mathcal{C}) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$. Die Gleichung bleibt auch dann richtig, wenn g beliebig reellwertig und $g \circ T$ μ -integrierbar bzw. g μ^T -integrierbar ist.

Beweis: Sei zunächst wieder $g = \mathbb{1}_C$ mit $C \in \mathcal{C}$. Dann ist

$$\int g \circ T d\mu = \int \mathbb{1}_{T^{-1}(C)} d\mu = \mu(T^{-1}(C)) = \mu^T(C) = \int \mathbb{1}_C d\mu^T = \int g d\mu^T.$$

I. Maßtheoretische Grundlagen

Für den weiteren Beweisgang argumentiert man analog wie im Beweis zu Satz 20. ■

Als einfache Folgerung zu Satz 22 erhält man noch die folgende Beziehung, die auch die Transformation der "Integrationsgrenzen" beinhaltet:

$$\int_{T^{-1}(C)} g \circ T d\mu = \int_C g d\mu^T \text{ für alle } C \in \mathcal{C}.$$

Dazu braucht in Satz 22 lediglich g durch $g \cdot \mathbb{1}_C$ ersetzt zu werden.

Als letzte wichtige Integralregel behandeln wir den Umkreis des Satzes von Fubini für Integrationen nach Produktmaßen. Wir beginnen zunächst mit dem einfacheren Fall von zwei "Faktoren".

Satz 23 (Satz von Fubini, 1907). Es seien $(\Omega_i, \mathcal{A}_i, \mu_i)$ Maßräume für $i=1,2$ mit σ -endlichen Maßen μ_i . Ferner sei $f : (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$ eine nicht-negative, (Produkt-)messbare Abbildung. Dann gilt: die Abbildung $f(\cdot, \omega_2)$ ist für jedes feste $\omega_2 \in \Omega_2$ \mathcal{A}_1 -messbar und die Abbildung $f(\omega_1, \cdot)$ ist für jedes feste $\omega_1 \in \Omega_1$ \mathcal{A}_2 -messbar, und es gilt

$$\int f d\mu_1 \otimes \mu_2 = \int \left\{ \int f(\omega_1, \omega_2) \mu_1(d\omega_1) \right\} \mu_2(d\omega_2) = \int \left\{ \int f(\omega_1, \omega_2) \mu_2(d\omega_2) \right\} \mu_1(d\omega_1).$$

Diese Gleichheit bleibt auch gültig, wenn f beliebig reell und eigentlich $\mu_1 \otimes \mu_2$ -integrierbar ist. In diesem Fall existieren die inneren Integrale $\int f(\omega_1, \omega_2) \mu_1(d\omega_1)$ und $\int f(\omega_1, \omega_2) \mu_2(d\omega_2)$ fast überall (als Funktionen von ω_2 bzw. von ω_1).

Beweis: Die durch $h_{\omega_2} : \Omega_1 \times \Omega_2 \rightarrow \Omega_1 \times \Omega_2 : (\eta_1, \eta_2) \mapsto (\eta_1, \omega_2)$ für jedes feste $\omega_2 \in \Omega_2$ definierte Abbildung ist $\mathcal{A}_1 \otimes \mathcal{A}_2$ - $\mathcal{A}_1 \otimes \mathcal{A}_2$ -messbar wegen

$$h_{\omega_2}^{-1}(A \times B) = \begin{cases} \emptyset, & \omega_2 \notin B \\ A \times \Omega_2, & \omega_2 \in B \end{cases} \in \mathcal{A}_1 \otimes \mathcal{A}_2 \text{ für alle } A \in \mathcal{A}_1, B \in \mathcal{A}_2.$$

Damit ist aber $f(\cdot, \omega_2) = f \circ h_{\omega_2}$ für jedes feste $\omega_2 \in \Omega_2$ \mathcal{A}_1 -messbar. Analog argumentiert man für die Abbildung $f(\omega_1, \cdot)$ für jedes feste $\omega_1 \in \Omega_1$. Also existiert jedes der Integrale $I_2(\omega_2) := \int f(\omega_1, \omega_2) \mu_1(d\omega_1)$ für jedes feste $\omega_2 \in \Omega_2$ und $I_1(\omega_1) := \int f(\omega_1, \omega_2) \mu_2(d\omega_2)$ für jedes feste $\omega_1 \in \Omega_1$. Es lässt sich nun zeigen, dass sowohl I_1 als auch I_2 nicht-negative messbare Abbildungen sind, so dass in jedem Fall die iterierten Integrale $\int \left\{ \int f(\omega_1, \omega_2) \mu_1(d\omega_1) \right\} \mu_2(d\omega_2)$ und $\int \left\{ \int f(\omega_1, \omega_2) \mu_2(d\omega_2) \right\} \mu_1(d\omega_1)$ existieren. Zur Überprüfung der Gleichheit aller drei Integrale wählt man zunächst $f = \mathbb{1}_{A \times B}$ mit $A \in \mathcal{A}_1, B \in \mathcal{A}_2$. Dann ist

I. Maßtheoretische Grundlagen

$$\begin{aligned}
 \int f d\mu_1 \otimes \mu_2 &= \mu_1 \otimes \mu_2(A \times B) = \mu_1(A) \cdot \mu_2(B) = \int \mu_1(A) \cdot \mathbb{1}_B d\mu_2 \\
 &= \int \left\{ \int \mathbb{1}_A(\omega_1) \mu_1(d\omega_1) \right\} \mathbb{1}_B(\omega_2) \mu_2(d\omega_2) = \int \left\{ \int \mathbb{1}_{A \times B}(\omega_1, \omega_2) \mu_1(d\omega_1) \right\} \mu_2(d\omega_2) \\
 &= \int \left\{ \int f(\omega_1, \omega_2) \mu_2(d\omega_2) \right\} \mu_1(d\omega_1).
 \end{aligned}$$

Analog ergibt sich die andere Gleichheit. Wir definieren nun ein Maß ν auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ vermöge

$$\nu(C) := \int \left\{ \int \mathbb{1}_C(\omega_1, \omega_2) \mu_2(d\omega_2) \right\} \mu_1(d\omega_1) \text{ für alle } C \in \mathcal{A}_1 \otimes \mathcal{A}_2.$$

Die definierenden Eigenschaften eines Maßes – insbesondere die σ -Additivität – lassen sich hier sofort verifizieren. Nach der vorigen Rechnung stimmt aber ν mit $\mu_1 \otimes \mu_2$ auf dem $\Omega_1 \times \Omega_2$ abzählbar überdeckenden Erzeuger bzw. Semi-Ring $\mathcal{E} = \{A \times B \mid A \in \mathcal{A}_1, B \in \mathcal{A}_2\}$ überein, so dass nach dem großen Maßfortsetzungssatz 7 beide Maße auf ganz $\mathcal{A}_1 \otimes \mathcal{A}_2$ übereinstimmen. Somit gilt

$$\int f d\mu_1 \otimes \mu_2 = \int \left\{ \int f(\omega_1, \omega_2) \mu_1(d\omega_1) \right\} \mu_2(d\omega_2)$$

auch für beliebige Indikatorfunktionen $f = \mathbb{1}_C$ mit $C \in \mathcal{A}_1 \otimes \mathcal{A}_2$ und damit wegen der Linearität des Integrals auch für beliebige $\mathcal{A}_1 \otimes \mathcal{A}_2$ -messbare Elementarfunktionen. Die allgemeine Aussage ergibt sich wie im Beweis zu Satz 21 bzw. 22 über monotone Approximation und Zerlegung von f in Positiv- und Negativteil. Die andere Gleichheit

$$\int f d\mu_1 \otimes \mu_2 = \int \left\{ \int f(\omega_1, \omega_2) \mu_2(d\omega_2) \right\} \mu_1(d\omega_1)$$

ergibt sich vollkommen analog, womit der Satz bewiesen ist. ■

Auf die Voraussetzung der σ -Endlichkeit der Maße kann übrigens nicht einfach verzichtet werden; für Gegenbeispiele vgl. die entsprechenden Ausführungen in BAUER (1992) oder ELSTRODT (1996).

Satz 23 läßt sich natürlich entsprechend auch auf den Fall von Produktmaßen mit mehr als zwei "Faktoren" übertragen. Die zugehörige Aussage wird dann ebenfalls als **Satz von Fubini** bezeichnet.

Zum Abschluss dieses Kapitels wollen wir noch kurz den Zusammenhang zwischen Riemann- und Lebesgue-Integral (bezüglich des Lebesgue-Maßes m^d) beleuchten, allerdings ohne Beweis.

Satz 24. Es sei $I \in \mathcal{B}^d$ ein nicht-leeres kompaktes Standard-Intervall. Eine beschränkte Funktion $f: I \rightarrow \mathbb{R}$ ist genau dann Riemann-integrierbar, wenn die Menge ihrer Unstetigkeitsstellen eine m^d -Nullmenge ist. In diesem Fall stimmt das Riemann-Integral von f mit dem Lebesgue-Integral überein. Der Satz bleibt richtig, wenn f allgemeiner auf einer beschränkten Jordan-messbaren¹³ Menge $M \subseteq \mathbb{R}^d$ definiert ist.

Ist ferner $I \in \mathcal{B}^1$ ein beliebiges Intervall und $f: I \rightarrow \mathbb{R}$ Riemann-integrierbar über jedem kompakten Teilintervall von I , so ist f genau dann (eigentlich) Lebesgue-integrierbar über I , wenn $|f|$ uneigentlich Riemann-integrierbar ist über I . In diesem Fall stimmen wieder beide Integrale überein.

Für einen Beweis siehe ELSTRODT (1996), Kapitel IV, §6.

Riemann- und Lebesgue-Integral stimmen allerdings nicht immer überein. So ist beispielsweise $\int \mathbb{1}_{\mathbb{Q} \cap [0,1]} dm^1 = 0$, jedoch ist $\mathbb{1}_{\mathbb{Q} \cap [0,1]}$ nicht Riemann-integrierbar. Umgekehrt ist die Funktion

$f := \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \mathbb{1}_{[k, k+1)}$ uneigentlich Riemann-integrierbar mit Wert $\int_0^{\infty} f(x) dx = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} = -\ln 2$,

jedoch nicht Lebesgue-integrierbar, da weder $f^+ = \sum_{k=1}^{\infty} \frac{1}{2k} \mathbb{1}_{[2k, 2k+1)}$ noch $f^- = \sum_{k=1}^{\infty} \frac{1}{2k-1} \mathbb{1}_{[2k-1, 2k)}$

eigentlich Lebesgue-integrierbar sind wegen $\int f^+ dm^1 = \int f^- dm^1 = \infty$.

Ein weiterer wichtiger Satz betrifft die mehrdimensionale Substitutionsregel der Integration, die üblicherweise im Rahmen der Analysis formuliert und bewiesen wird, und die in ihrer Grundform schon auf C.G.J. Jacobi (1841) zurückgeht. Sie gilt allerdings auch im Kontext der Maßtheorie, vgl. ELSTRODT (1996), Kapitel V, §4, insbesondere Satz 4.2 und Korollar 4.9.

Satz 25 (Transformationsformel). Es sei $d \in \mathbb{N}$ und $U \subseteq \mathbb{R}^d$ eine offene Menge. Ferner sei die Vektorfunktion $g = (g_1, \dots, g_d): U \rightarrow \mathbb{R}^d$ stetig differenzierbar, $V := g(U)$ das Bild von U unter g , $C := \{\mathbf{x} = (x_1, \dots, x_d) \in U \mid \det(\Delta g(\mathbf{x})) = 0\}$ die Menge der kritischen Punkte von g , wobei

$$\Delta g(\mathbf{x}) := \begin{bmatrix} \frac{\partial}{\partial x_1} g_1(\mathbf{x}) & \cdots & \frac{\partial}{\partial x_d} g_1(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial}{\partial x_1} g_d(\mathbf{x}) & \cdots & \frac{\partial}{\partial x_d} g_d(\mathbf{x}) \end{bmatrix}$$

die Funktionalmatrix von g im Punkt $\mathbf{x} \in U$ bezeichne, und g auf der Menge $U \setminus C$ injektiv. Eine Abbildung $f: V \rightarrow \mathbb{R}$ ist genau dann Lebesgue-integrierbar über V , wenn $f \circ g \cdot |\det(\Delta g)|$ Lebesgue-integrierbar über U ist. In diesem Fall gilt:

$$\int_V f dm^d = \int_U f \circ g \cdot |\det(\Delta g)| dm^d.$$

Hiermit eng verwandt ist das folgende Resultat:

¹³ vgl. ELSTRODT (1996), Kapitel I, §3.

I. Maßtheoretische Grundlagen

Satz 26 (Transformationssatz für Dichten). Ist unter den Voraussetzungen von Satz 25 die Abbildung f nicht-negativ, die Abbildung g injektiv und die Menge der kritischen Punkte leer, so ist die auf V definierte Abbildung h mit

$$h(\mathbf{y}) := \frac{f(g^{-1}(\mathbf{y}))}{\left| \det(\Delta g(g^{-1}(\mathbf{y}))) \right|}, \quad \mathbf{y} \in V$$

nicht-negativ und Borel-messbar, und es gilt:

$$\int_A h(\mathbf{y}) m^d(d\mathbf{y}) = \int_{g^{-1}(A)} f(\mathbf{x}) m^d(d\mathbf{x}) \quad \text{für alle } A \in V \cap \mathcal{B}^d.$$

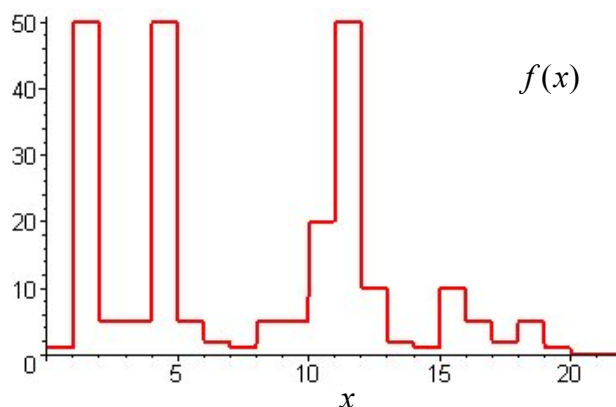
Bemerkung: Zur Vereinfachung der Schreibweise werden wir ab jetzt häufig das Lebesgue-Integral bezüglich des Lebesgue-Maßes m^d genau so notieren wie das Riemann-Integral, also z.B.

$\int_A f(x) m^1(dx)$ als $\int_A f(x) dx$ schreiben (für Borel-Mengen A).

Zum Abschluss dieses Kapitels soll noch einmal an einem einfachen plakativen Beispiel der inhaltliche Unterschied zwischen Riemann- und Lebesgue-Integral veranschaulicht werden. Dazu betrachten wir die folgende Kleingeld-Sammlung, deren Wert wir bestimmen wollen:



Wir können die Münzsammlung graphisch darstellen, indem wir die Centwerte der Reihe nach als Balken darstellen (von links nach rechts / oben nach unten):



Der Wert der Sammlung ist dann das Integral $\int_0^{20} f(x) dx$. Im Riemannschem Sinn addiert man dazu

die Werte der Münzen der Reihe nach, also $\int_0^{20} f(x) dx = 1 + 50 + 5 + 5 + 50 + \dots + 1 = 235$. Beim

Lebesgue-Integral fasst man aber zunächst die Münzen gleichen Werts zusammen und addiert

dann: $\int_0^{20} f(x) dx = 1 \cdot 4 + 2 \cdot 3 + 5 \cdot 7 + \dots + 50 \cdot 3 = 235$.

II. Grundprinzipien stochastischer Modellierung

In diesem zweiten Teil des Textes wollen wir die grundlegenden Ideen der eigentlichen "Stochastik" auf der Basis der zuvor eingeführten allgemeinen Maß- und Integrationstheorie entwickeln. Wir beginnen dabei mit den schon zu Beginn des Textes erwähnten historischen Anfangsgründen, in deren Mittelpunkt wesentlich der Laplace'sche Wahrscheinlichkeitsbegriff steht, der seinerseits stark auf kombinatorischen Überlegungen beruht.

II.1. Elemente der Kombinatorik

Ein wesentlicher Aspekt des Laplace'sche Wahrscheinlichkeitsbegriffs ist das zahlenmäßige Verhältnis von Mächtigkeiten endlicher Mengen. Die systematische Bestimmung solcher Mächtigkeiten ist Gegenstand der Kombinatorik, vor allem dann, wenn die betrachteten Mengen (kartesische) Produktstruktur besitzen.

Definition 22 (Permutationen und Kombinationen). Es sei $\Omega = \{\omega_1, \dots, \omega_n\}$ eine nicht-leere endliche Menge mit $n \in \mathbb{N}$. Für $1 \leq k \leq n$ definieren wir:

- Jedes Element $(\eta_1, \dots, \eta_k) \in \Omega^k = \prod_{i=1}^k \Omega$ heißt (k, n) -Permutation aus Ω (mit Wiederholung). Die Menge aller solchen Permutationen wird mit $Perm_k^n(\Omega; m.W.)$ bezeichnet.
- Jede (k, n) -Permutation (η_1, \dots, η_k) mit paarweise verschiedenen Komponenten heißt (k, n) -Permutation aus Ω (ohne Wiederholung). Die Menge dieser Permutationen wird mit $Perm_k^n(\Omega; o.W.)$ bezeichnet.
- Jede (k, n) -Permutation $(\omega_{i_1}, \dots, \omega_{i_k})$ mit $1 \leq i_1 \leq i_2 \leq \dots \leq i_k \leq n$ heißt (k, n) -Kombination aus Ω (mit Wiederholung). Die Menge dieser Kombinationen wird mit $Komb_k^n(\Omega; m.W.)$ bezeichnet.
- Jede (k, n) -Kombination $(\omega_{i_1}, \dots, \omega_{i_k})$ mit paarweise verschiedenen Komponenten heißt (k, n) -Kombination aus Ω (ohne Wiederholung). Die Menge dieser Kombinationen wird mit $Komb_k^n(\Omega; o.W.)$ bezeichnet.

Zur Veranschaulichung der genannten Begriffe betrachten wir beispielhaft die Menge $\Omega = \{1, 2, 3, 4, 5, 6\}$ (als Modell für einen Fairen Würfel). Dann gilt:

$$Perm_2^6(\Omega; m.W.) = \{(1,1); (1,2); (2,1); (1,3); (3,1); (1,4); (4,1); (1,5); (5,1); (1,6); (6,1); (2,2); (2,3); (3,2); (2,4); (4,2); (2,5); (5,2); (2,6); (6,2); (3,3); (3,4); (4,3); (3,5); (5,3); (3,6); (6,3); (4,4); (4,5); (5,4); (4,6); (6,4); (5,5); (5,6); (6,5); (6,6)\}$$

$$Perm_2^6(\Omega; o.W.) = \{(1,2); (2,1); (1,3); (3,1); (1,4); (4,1); (1,5); (5,1); (1,6); (6,1); (2,3); (3,2); (2,4); (4,2); (2,5); (5,2); (2,6); (6,2); (3,4); (4,3); (3,5); (5,3); (3,6); (6,3); (4,5); (5,4); (4,6); (6,4); (5,6); (6,5)\}$$

$$Komb_2^6(\Omega; m.W.) = \{(1,1); (1,2); (1,3); (1,4); (1,5); (1,6); (2,2); (2,3); (2,4); (2,5); (2,6); (3,3); (3,4); (3,5); (3,6); (4,4); (4,5); (4,6); (5,5); (5,6); (6,6)\}$$

II. Grundprinzipien stochastischer Modellierung

$$Komb_2^6(\Omega; o.W.) = \{(1, 2); (1, 3); (1, 4); (1, 5); (1, 6); (2, 3); (2, 4); (2, 5); (2, 6); (3, 4); (3, 5); (3, 6); (4, 5); (4, 6); (5, 6)\}$$

Über die Mächtigkeiten von Permutations- und Kombinationsmengen gibt das folgende Resultat Auskunft.

Lemma 23. Mit den Bezeichnungen aus Definition 22 gilt:

$$\#(Perm_k^n(\Omega; m.W.)) = n^k$$

$$\#(Perm_k^n(\Omega; o.W.)) = (n)_k = \binom{n}{k} k! = n(n-1)\cdots(n-k+1)$$

$$\#(Komb_k^n(\Omega; m.W.)) = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$$

$$\#(Komb_k^n(\Omega; o.W.)) = \binom{n}{k}.$$

Beweis: Für jede Komponente η_i einer (k, n) -Permutation mit Wiederholung (η_1, \dots, η_k) gibt es jeweils $n = \#(\Omega)$ Wahlmöglichkeiten, woraus die erste Beziehung unmittelbar folgt. Für eine entsprechende (k, n) -Permutation ohne Wiederholung (η_1, \dots, η_k) gibt es n Wahlmöglichkeiten für die erste Komponente η_1 , für die zweite Komponente η_2 gibt es $(n-1)$ Wahlmöglichkeiten, usw. bis schließlich $(n-k+1)$ Wahlmöglichkeiten für die letzte Komponente η_k . Dies ergibt die zweite Beziehung. Zum Beweis der vierten Beziehung beachte man, dass jede (k, n) -Permutation ohne Wiederholung $(\omega_{i_1}, \dots, \omega_{i_k})$ genau $k!$ verschiedene Permutationen ohne Wiederholung mit denselben Komponenten erzeugt. Die dritte Beziehung folgt schließlich, wenn man die Menge Ω um weitere $(k-1)$ Elemente $\omega_{n+1}, \dots, \omega_{n+k-1}$ zu der Menge $\Omega^* = \{\omega_1, \dots, \omega_{n+k-1}\}$ ergänzt und die Bijektion $(\omega_{i_1}, \omega_{i_2}, \dots, \omega_{i_k}) \mapsto (\omega_{i_1}, \omega_{i_2+1}, \dots, \omega_{i_k+k-1})$ von $Komb_k^n(\Omega; m.W.)$ auf $Komb_k^{n+k-1}(\Omega; o.W.)$ betrachtet. ■

Nach Lemma 23 erhält man für das obige Würfelbeispiel

$$\#(Perm_2^6(\Omega; m.W.)) = 6^2 = 36, \#(Perm_2^6(\Omega; o.W.)) = (6)_2 = 6 \cdot 5 = 30,$$

$$\#(Komb_2^6(\Omega; m.W.)) = \binom{7}{2} = \frac{7 \cdot 6}{1 \cdot 2} = 21, \#(Komb_2^6(\Omega; o.W.)) = \binom{6}{2} = \frac{6 \cdot 5}{1 \cdot 2} = 15$$

in Übereinstimmung mit den angegebenen Ergebnissen.

II. Grundprinzipien stochastischer Modellierung

Wir können damit nun das eingangs erwähnte Problem des französischen Adligen George Brossin Antoine Gombaud, Chevalier de Méré lösen, das wir hier noch einmal im Wortlaut wiedergeben:

Warum ist es unvorteilhafter, zum Erreichen einer Doppelsechs mit zwei Würfeln 24 Würfe zu tun, als zum Erreichen einer Sechs mit einem Würfel 4 Würfe zu tun, obwohl das Verhältnis 24 zu 36 (was die Anzahl der Ergebnisse bei 2 Würfeln ist) dasselbe ist wie 4 zu 6 (was die Anzahl der Ergebnisse bei einem Würfel ist)?

Als Grundmodell wählen wir für den einfachen Würfelwurf wie oben die Menge $\Omega = \{1, 2, 3, 4, 5, 6\}$, für den m -fachen Würfelwurf (entweder gleichzeitig oder nacheinander) die Menge Ω^m . Als kanonische σ -Algebra wählen wir jeweils die Potenzmenge, in Anlehnung an die Bemerkung im Anschluss an Satz 6 und als Wahrscheinlichkeitsmaß das Laplace-Maß nach Definition 7. In der ersten Situation wird dann das komplementäre Ereignis beschrieben durch die Menge

$$A^c = \bigtimes_{i=1}^{24} \{(6, 6)\}^c \subseteq \Omega^{48} \text{ mit } P(A^c) = \frac{\#(A^c)}{\#(\Omega^{48})} = \frac{35^{24}}{6^{48}} = 0,50859\dots,$$

(d.h. es tritt innerhalb von 24 Doppelwürfen keinmal der Sechserpasch auf), in der zweiten Situation durch

$$A^c = \bigtimes_{i=1}^4 \{6\}^c \subseteq \Omega^4 \text{ mit } P(A^c) = \frac{\#(A^c)}{\#(\Omega^4)} = \frac{5^4}{6^4} = 0,48225\dots$$

(d.h. es tritt innerhalb von 4 Einfachwürfen keinmal die Sechs auf). In der ersten Situation ist also

$$P(A) = 1 - \frac{35^{24}}{6^{48}} = 0,49140\dots < 0,5$$

und in der zweiten

$$P(A) = 1 - \frac{5^4}{6^4} = 0,51774\dots > 0,5.$$

Der Gedankenfehler des Chevalier de Méré besteht hier darin, dass die "Anzahl der Ergebnisse" jeweils in einem falschen Wahrscheinlichkeitsmodell bestimmt wird und nicht in dem korrekten Produktraum (mit unterschiedlichem "Exponenten"). Immerhin ist es erstaunlich, dass der doch recht kleine Unterschied in den tatsächlichen Wahrscheinlichkeiten empirisch festgestellt wurde, was zeigt, dass der damalige Adel wohl ziemlich viel Zeit mit Glücksspielen verbracht hat.

Zahlreiche weitere historische Beispiele kann man in dem Schulbuch von BARTH UND HALLER (1998) finden.

Kombinatorische Aufgaben sind häufig gar nicht so einfach zu lösen, weil oft nicht sofort klar ist, ob Situationen mit oder ohne Wiederholung vorliegen. Das folgende Beispiel möge dies veranschaulichen:

II. Grundprinzipien stochastischer Modellierung

Peter und Paul werfen beide gleichzeitig einen (fairen) Würfel. Peter sagt: „Die Wahrscheinlichkeit dafür, dass wir beide verschiedene Augenzahlen werfen, beträgt $5/6$, weil es insgesamt 36 Augenpaare gibt, davon 30 mit unterschiedlichen Augenzahlen.“ Paul antwortet: „Unsinn! Deine Rechnung geht davon aus, dass die Anordnung der Würfel eine Rolle spielt. Ich kann aber die beiden Würfel so vor mich hinlegen, dass der Würfel mit der kleineren Augenzahl links vor mir und der mit der höheren Augenzahl rechts vor mir liegt. Nach den vier kombinatorischen Grundformeln gibt es dann $\binom{6+2-1}{2} = \binom{7}{2} = 21$ mögliche Fälle mit Wiederholung der Augenzahl, davon sind $\binom{6}{2} = 15$ Fälle ohne Wiederholung der Augenzahl günstig. Die fragliche Wahrscheinlichkeit beträgt also nur $15/21 = 5/7$.“ Wer hat Recht?

Die richtige Antwort lautet: Peter hat Recht. Das Problem besteht darin, dass bei Augengleichheit keine eindeutige Anordnung der Würfel möglich ist. Ausweg: die Würfel werden unterschiedlich farbig markiert (z.B. rot, grün). Wenn Paul dann immer den roten Würfel links und den grünen rechts vor sich hinlegt, wird klar, warum Peter Recht hat.

Ein anderes verblüffendes Ergebnis verbirgt sich hinter der gemeinhin als „Geburtstagsparadox“ bekannten Fragestellung. Sie lautet: In einer Runde von k Personen soll darauf gewettet werden, dass wenigstens zwei Anwesende am gleichen Tag Geburtstag haben. Ab welcher Personenzahl würden Sie diese Wette stellen? Die überraschende Antwort lautet: bereits ab 23 Personen ist die Wahrscheinlichkeit für dieses Ereignis $A_{k,n}$ größer als $1/2$! Wir betrachten dazu sogleich das folgende allgemeinere Modell: als Ausgangspunkt wählen wir die Menge $\Omega = \{1, 2, \dots, n\}$, hier mit $n = 365$ für die möglichen Tage im Jahr. Dann beschreibt die Menge $Perm_k^n(\Omega; m.W.)$ die Menge der Möglichkeiten, wie sich die Geburtstage von k Personen auf das Jahr verteilen können. Diese Menge wählen wir als Grundraum für das Zufallsexperiment, darüber die Potenzmenge als σ -Algebra und die Laplace-Verteilung P als Wahrscheinlichkeitsverteilung. Damit nehmen wir implizit an, dass sich die Geburtstage gleichmäßig im Jahr verteilen.¹⁴ Zur Berechnung der gesuchten Wahrscheinlichkeit betrachten wir zunächst das komplementäre Ereignis $A_{k,n}^c$, dass alle k Personen an *verschiedenen* Tagen Geburtstag haben. Dann ist offensichtlich $A_{k,n}^c = Perm_k^n(\Omega; o.W.)$, und es gilt

$$\begin{aligned} P(A_{k,n}) &= 1 - P(A_{k,n}^c) = 1 - \frac{\#(Perm_k^n(\Omega; o.W.))}{\#(Perm_k^n(\Omega; m.W.))} = 1 - \frac{(n)_k}{n^k} = 1 - \prod_{i=0}^{k-1} \left(1 - \frac{i}{n}\right) \\ &= 1 - \exp\left(\sum_{i=0}^{k-1} \ln\left(1 - \frac{i}{n}\right)\right) \geq 1 - \exp\left(-\sum_{i=0}^{k-1} \frac{i}{n}\right) = 1 - \exp\left(-\frac{k(k-1)}{2n}\right) \geq 1 - \exp\left(-\frac{(k-1)^2}{2n}\right). \end{aligned}$$

Dabei haben wir von der aus der Analysis bekannten Ungleichung $\ln(1-x) \leq -x$, $0 < x < 1$ Gebrauch gemacht. Löst man nun die Ungleichung

$$1 - \exp\left(-\frac{(k-1)^2}{2n}\right) \geq \frac{1}{2}$$

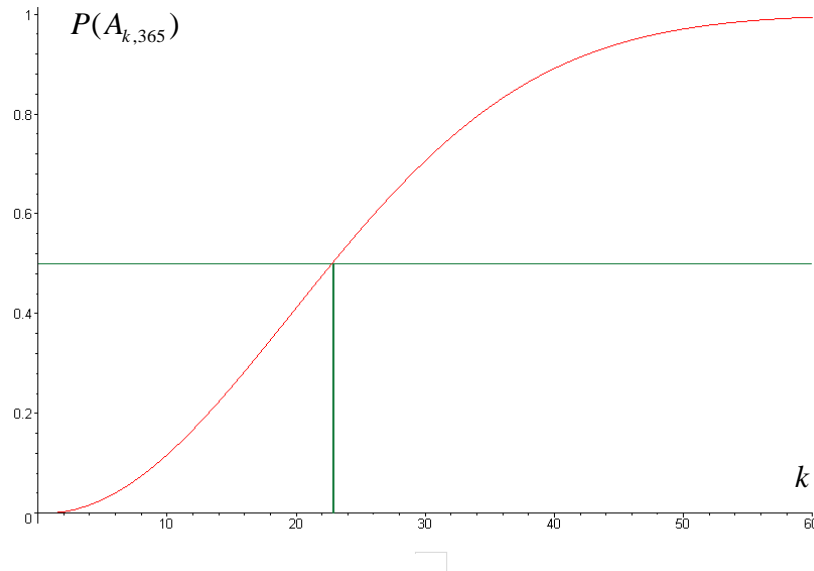
nach k auf, so erhält man

¹⁴ Diese Annahme ist, wie man zeigen kann, sogar die ungünstigste, d.h. bei einer hiervon abweichenden Wahrscheinlichkeitsverteilung wird die Wahrscheinlichkeit für wenigstens einen gemeinsamen Geburtstag sogar noch größer.

II. Grundprinzipien stochastischer Modellierung

$$k \geq 1 + \sqrt{n \ln 4} \geq 1 + \sqrt{1,386n}.$$

Für $n = 365$ liefert diese approximative Lösung $k \geq 23,49\dots$, die exakte Lösung ist $k \geq 23$ mit $P(A_{23,365}) = 0,50729\dots$. Die folgende Graphik zeigt die Wahrscheinlichkeiten $P(A_{k,365})$ in Abhängigkeit von k im Bereich $1 \leq k \leq 60$.



II.2. Zufallsvariablen und ihre Verteilung, stochastische Unabhängigkeit

Wir beginnen mit dem einführenden Beispiel aus BARTH UND HALLER (1998), Kapitel 11.

Auf einem Jahrmarkt wird in einer Glücksbude folgendes Spiel angeboten: Der Spieler leistet 1 € Einsatz, darf eine der Zahlen 1,2,3,4,5,6 nennen und dann drei Würfel werfen. Zeigt mindestens einer der Würfel seine Zahl, so erhält er vom Budenbesitzer seinen Einsatz zurück und ausserdem für jeden Würfel der diese Zahl zeigt, noch zusätzlich 1 €. Erscheint seine Zahl nicht, so verfällt der Einsatz.

Ein solches Spiel ist in den USA als *chuck-a-luck* bekannt.

Wir interessieren uns zunächst für den bei dem Spiel erzielbaren *Gewinn*, d.h. für die Größe *Auszahlung* minus *Einsatz*. Dazu ordnen wir den möglichen Ergebnissen des Spiels die entsprechende Gewinngröße zu. Aus Ausgangsmodell wählen wir die Menge $\Omega = \{1,2,3,4,5,6\}^3$ aller Tripel, die aus den Zahlen 1,2,3,4,5,6 gebildet werden können. Sie symbolisiert die möglichen Wurfresultate mit drei Würfeln. Wählt der Spieler nun die Zahl $k \in \{1,2,3,4,5,6\}$, so kann sein Gewinn formal durch die auf Ω definierte Abbildung X_k beschrieben werden, die gegeben ist durch

$$X_k(\omega) = \#\{i | \omega_i = k\} - 1 + \min\{\#\{i | \omega_i = k\}, 1\} \text{ für } \omega = (\omega_1, \omega_2, \omega_3) \in \Omega.$$

Die resultierende Wertemenge ist offensichtlich gegeben durch $\mathfrak{X} = \{-1, 1, 2, 3\}$. Der negative Gewinn -1 wird dabei genau dann realisiert, wenn $\#\{i | \omega_i = k\} = 0$ ist, also keiner der Würfel die gewählte Zahl des Spielers zeigt. Wenn die vom Spieler genannte Zahl auf einem der Würfel er-

II. Grundprinzipien stochastischer Modellierung

scheint, gilt dagegen gerade $X_k(\omega) = \#\{i | \omega_i = k\}$, wie gefordert. Wenn wir nun die Mengen Ω und \mathfrak{X} mit der jeweiligen Potenzmenge \mathcal{A} bzw. \mathcal{B} als zugehöriger σ -Algebra ausstatten, können wir die Abbildung $X_k(\cdot)$ als messbare Abbildung zwischen den Messräumen (Ω, \mathcal{A}) und $(\mathfrak{X}, \mathcal{B})$ auffassen. Wenn wir uns jetzt für die Wahrscheinlichkeitsverteilung Q_k über \mathcal{B} der möglichen Gewinne interessieren, so entspricht das mathematisch der Frage nach dem Bildmaß gemäß Definition 12. Dazu nehmen wir sinnvollerweise an, dass auf dem originären Messraum die Laplace-Verteilung P gegeben ist. Demnach ist also $Q_k = P^{X_k}$. Zur expliziten Berechnung benötigen wir hierfür die Urbildmengen $X_k^{-1}(B)$ mit $B \in \mathcal{B}$. Da die Menge \mathfrak{X} endlich ist und die als Potenzmenge darauf gegebene σ -Algebra \mathcal{B} die einelementigen Teilmengen als Atome besitzt, reicht es nach Satz 6, die Urbildmengen $X_k^{-1}(\{x\})$ mit $x \in \mathfrak{X}$ zu betrachten. Zur Vereinfachung der Schreibweise benutzen wir als übliche Konvention die Notation

$$X_k^{-1}(\{x\}) = \{X_k = x\} \quad \text{bzw. allgemeiner} \quad X_k^{-1}(B) = \{X_k \in B\} \quad \text{für } x \in \mathfrak{X} \text{ bzw. } B \in \mathcal{B}.$$

Demnach ist nun

$$Q_k(\{x\}) = P(X_k = x) = \frac{\#\{X_k = x\}}{\#\Omega} \quad \text{für } x \in \mathfrak{X}.$$

Es gilt hier:

$$\#\{X_k = 3\} = 1$$

[alle drei Würfel zeigen die gewählte Zahl k]

$$\#\{X_k = 2\} = 15 = \binom{3}{2} \cdot 5$$

[genau zwei von drei Würfeln zeigen die gewählte Zahl k , d.h. es gibt zunächst $\binom{3}{2} = 3$

Möglichkeiten der Auswahlen von zwei von drei Komponenten für k , dann noch 5 Möglichkeiten für die restliche Komponente ohne k]

$$\#\{X_k = 1\} = 75 = \binom{3}{1} \cdot 5^2$$

[genau einer von drei Würfeln zeigen die gewählte Zahl k , d.h. es gibt zunächst $\binom{3}{1} = 3$

Möglichkeiten der Auswahl von einer von drei Komponenten für k , dann noch $5^2 = 25$ Möglichkeiten für die restlichen beiden Komponenten ohne k]

$$\#\{X_k = -1\} = 125 = 5^3$$

[keiner von drei Würfeln zeigen die gewählte Zahl k , d.h. alle Würfel zeigen Ergebnisse aus der kleineren Menge $\{1, 2, 3, 4, 5\}$]

II. Grundprinzipien stochastischer Modellierung

Insgesamt erhält man damit die folgenden Wahrscheinlichkeiten:

x	-1	1	2	3
$Q_k(\{x\}) = P(X_k = x)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

Diese Wahrscheinlichkeiten sind erwartungsgemäß für alle Wahlen k gleich. Die Wahrscheinlichkeit für einen Verlust des Einsatzes beträgt dabei $\frac{125}{216} = 0,5787... > 0,5$.

Definition 23 (Zufallsvariable / Zufallsvektor). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(\mathfrak{X}, \mathcal{B})$ ein Messraum. $X : (\Omega, \mathcal{A}) \rightarrow (\mathfrak{X}, \mathcal{B})$ sei eine messbare Abbildung. Dann heißt X

Zufallsvariable, wenn $\mathfrak{X} = \mathbb{R}$ und \mathcal{B} die zugehörige Borel'sche σ -Algebra ist;

Zufallsvektor, wenn $\mathfrak{X} = \mathbb{R}^d$ für ein $d \in \mathbb{N}$, $d \geq 2$ und \mathcal{B} die zugehörige Borel'sche σ -Algebra ist;

Zufallselement, wenn \mathfrak{X} ein allgemeinerer topologischer Raum und \mathcal{B} die zugehörige Borel'sche σ -Algebra¹⁵ ist.

Das durch die Abbildung X induzierte Maß $Q = P^X$ auf \mathcal{B} heißt (*Wahrscheinlichkeits-*)*Verteilung von X* .

Man spricht dabei auch dann von Zufallsvariablen bzw. Zufallsvektoren, wenn die Bildmenge \mathfrak{X} lediglich ein Teilraum von \mathbb{R} bzw. \mathbb{R}^d ist, und die σ -Algebra \mathcal{B} die entsprechende Spur- σ -Algebra in der Borel'schen σ -Algebra ist. Beispielsweise kann die für das *chuck-a-luck*-Spiel verwendete Abbildung X_k als Zufallsvariable aufgefasst werden, weil die Bildmenge $\mathfrak{X} = \{-1, 1, 2, 3\}$ Teilmenge von \mathbb{R} und die Potenzmenge $\mathcal{B} = \mathfrak{P}(\mathfrak{X})$ zugleich ihre eigene Spur- σ -Algebra ist, da alle endlichen Teilmengen von \mathbb{R} Borelsch sind.

Zufallsvektoren entstehen natürlicherweise immer dann, wenn man mehrere Zufallsvariablen gleichzeitig betrachtet, z.B. das Werfen von drei Würfeln. Sind etwa X_1, X_2, X_3 die Zufallsvariablen, die auf $\Omega = \{1, 2, 3, 4, 5, 6\}^3$ vermöge $X_i(\omega_1, \omega_2, \omega_3) := \omega_i$ für $i = 1, 2, 3$ definiert sind, so beschreiben die X_i das Wurfresultat des i -ten Würfels. Der Zufallsvektor $X = (X_1, X_2, X_3)$ ist dann eine \mathbb{R}^3 -wertige Abbildung auf Ω und beschreibt das Ergebnis eines gleichzeitigen Wurfes dreier Würfel. Lemma 12 regelt hier die entsprechenden Messbarkeitsfragen.

¹⁵d.h. die σ -Algebra, die von den offenen Mengen erzeugt wird.

II. Grundprinzipien stochastischer Modellierung

Definition 24 (Stochastischer Prozess). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und \mathfrak{X} ein Banach-Raum reeller Funktionen; d.h. \mathfrak{X} ist ein vollständiger, linearer, normierter topologischer Raum. \mathcal{B} bezeichne die von der Topologie des Raumes erzeugte Borel'sche σ -Algebra, $X : (\Omega, \mathcal{A}) \rightarrow (\mathfrak{X}, \mathcal{B})$ sei eine messbare Abbildung. Dann heißt X stochastischer Prozess.

Stochastische Prozesse sind demnach spezielle Zufallselemente. Jedes $X(\omega)$, $\omega \in \Omega$ ist also eine reelle Funktion. Wählen wir beispielsweise $\mathfrak{X} = C[0,1]$, den Raum der auf dem Intervall $[0,1]$ (gleichmäßig) stetigen Funktionen mit der Norm

$$\|f\| := \max \{ |f(x)| \mid x \in [0,1] \} \text{ für } f \in \mathfrak{X},$$

so wird durch

$$K_\varepsilon(f) := \{ g \in X \mid \|f - g\| < \varepsilon \} \text{ für } f \in \mathfrak{X} \text{ und } \varepsilon > 0$$

eine *offene Kugel* $K_\varepsilon(f)$ mit Mittelpunkt $f \in \mathfrak{X}$ und Radius $\varepsilon > 0$ definiert. *Offene Mengen* in $\mathfrak{X} = C[0,1]$ sind dann allgemeiner solche Teilmengen O von \mathfrak{X} , für die gilt, dass zu jedem Element $f \in O$ ein $\varepsilon > 0$ existiert mit $K_\varepsilon(f) \subseteq O$ (f ist dann ein sog. *innerer Punkt*). Ist nun (Ω, \mathcal{A}, P) ein beliebiger Wahrscheinlichkeitsraum und $X : (\Omega, \mathcal{A}) \rightarrow (\mathfrak{X}, \mathcal{B})$, so sind also die einzelnen Bilder $X(\omega)$ sämtlich stetige Funktionen. Wertet man diese Funktionen an der Stelle $t \in [0,1]$ aus, erhält man die Zufallsvariable $X_t(\omega)$; in diesem Sinne kann man X auch als unendlichdimensionalen Zufallsvektor auffassen, allerdings in einem Produktraum mit überabzählbar vielen Komponenten. Solche stochastischen Prozesse mit – wie man sagt: stetigen Pfaden – spielen in der modernen Finanzmathematik zur Modellierung von Aktienkursen oder Zinsstrukturkurven eine zentrale Rolle, Stichwort: Brown'sche Bewegung oder Wiener-Prozess.

Zu beachten ist hier noch, dass die obige σ -Algebra \mathcal{B} echt größer ist als die Produkt- σ -Algebra $\bigotimes_{i \in [0,1]} \mathcal{B}^1_{[0,1]}$ der eindimensionalen Spur- σ -Algebren \mathcal{B}^1 in $[0,1]$, da etwa die Menge \mathfrak{X} selbst kein Element von $\bigotimes_{i \in [0,1]} \mathcal{B}^1_{[0,1]}$ ist, weil die Stetigkeit einer Funktion eine Forderung ist, die sich nicht nur auf höchstens abzählbar viele Argumente beschränken lässt; vgl. hierzu Satz 11.

Eine sehr wesentliche Eigenschaft von Zufallsvariablen und –vektoren ist die so genannte *stochastische Unabhängigkeit*; vgl. die Bemerkung im Anschluss an Satz 12. Man kann sie recht allgemein für beliebige Familien von messbaren Abbildungen definieren.

Definition 25 (stochastische Unabhängigkeit). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(\mathfrak{X}, \mathcal{B})$ ein Messraum. $X_i : (\Omega, \mathcal{A}) \rightarrow (\mathfrak{X}, \mathcal{B})$ seien messbare Abbildungen mit $i \in I$; die Indexmenge I darf dabei beliebig nicht-leer gewählt sein. Die Familie $\{X_i \mid i \in I\}$ heißt *stochastisch unabhängig*, wenn die auf dem Produktraum $\left(\prod_{i \in I} \mathfrak{X}, \bigotimes_{i \in I} \mathcal{B} \right)$ induzierte Verteilung der Familie das Produktmaß $Q = \bigotimes_{i \in I} P^{X_i}$ ist.

II. Grundprinzipien stochastischer Modellierung

Aufgrund der Sätze 11 und 12 kann diese Eigenschaft alternativ auch so ausgedrückt werden:

Lemma 24: In der Situation von Definition 25 ist die Familie $\{X_i \mid i \in I\}$ genau dann stochastisch unabhängig, wenn jede endliche Teilfamilie $\{X_j \mid j \in J_m\}$ mit $J_m = \{j_1, \dots, j_m\} \subseteq I$, $m \in \mathbb{N}$ stochastisch unabhängig ist, d.h. wenn gilt

$$P^{(X_{j_1}, \dots, X_{j_m})} = \bigotimes_{k=1}^m P^{X_{j_k}}.$$

Beweis: Es gibt nach Satz 11 für jede Menge $C \in \bigotimes_{i \in I} \mathcal{B}$ eine Darstellung $C = B \times \prod_{i \in I \setminus J} \mathcal{X}$ mit $B \in \bigotimes_{j \in J} \mathcal{B}$ und einer abzählbaren Teilindexmenge $J \subseteq I$. Demnach ist die Aussage des Lemmas jedenfalls richtig für abzählbare Teilindexmengen $J \subseteq I$, und es reicht, die Aussage für abzählbare Indexmengen I zu zeigen.

" \Rightarrow ": Folgt sofort aus der Definition des Produktmaßes.

" \Leftarrow ": O.B.d.A. können wir $I = \mathbb{N}$ und $J_m = \{1, \dots, m\}$ mit $m \in \mathbb{N}$ wählen. Sei nun $C = \prod_{i=1}^{\infty} B_i \in \bigotimes_{i \in I} \mathcal{B}$ mit $B_i \in \mathcal{B}$ für $i \in I$. Die durch $C_m = \prod_{i=1}^m B_i \times \prod_{i=m+1}^{\infty} \mathcal{X} \in \bigotimes_{i \in I} \mathcal{B}$ mit $m \in \mathbb{N}$ gegebene Mengenfolge ist dann monoton wachsend mit $C = \bigcup_{m=1}^{\infty} C_m$, so dass wegen der Stetigkeit von Maßen von unten (Satz 5) folgt

$$P^{\{X_i\}_{i \in I}}(C) = \lim_{m \rightarrow \infty} P^{\{X_i\}_{i \in I}}(C_m) = \lim_{m \rightarrow \infty} P^{(X_1, \dots, X_m)}(C_m) = \lim_{m \rightarrow \infty} \bigotimes_{i=1}^m P^{X_i}(C_m) = \bigotimes_{i=1}^{\infty} P^{X_i}(C).$$

Damit ist das Lemma bewiesen (vgl. die Bemerkung vor Satz 11). ■

Eine weitere Vereinfachung ergibt sich mit Satz 7 bzw. Satz 12 und der Bemerkung vor Satz 11 folgendermaßen.

Lemma 25. In der Situation von Definition 25 ist die Familie $\{X_i \mid i \in I\}$ genau dann stochastisch unabhängig, wenn für jede endliche Teilindexmenge $J_m = \{j_1, \dots, j_m\} \subseteq I$, $m \in \mathbb{N}$ gilt:

$$P(X_{j_1} \in B_1, \dots, X_{j_m} \in B_m) := P\left(\bigcap_{k=1}^m \{X_{j_k} \in B_k\}\right) = \prod_{k=1}^m P(\{X_{j_k} \in B_k\}) \text{ für alle } B_k \in \mathcal{B}, 1 \leq k \leq m.$$

Hierbei kann die Bedingung $B_k \in \mathcal{B}$ noch ersetzt werden durch $B_k \in \mathcal{S}$, wobei \mathcal{S} einen Semi-Ring bezeichne, der \mathcal{B} erzeugt.

Es sollte noch bemerkt werden, dass in Definition 25 und den nachfolgenden Lemmata der Zielraum $(\mathcal{X}, \mathcal{B})$ nicht notwendig für alle X_i derselbe sein muss. Eine entsprechende allgemeinere Formulierung – bei gleicher Argumentation – überlassen wir dem Leser.

II. Grundprinzipien stochastischer Modellierung

Mit Hilfe der Indikatorfunktion (Definition 16) lässt sich der Begriff der stochastischen Unabhängigkeit auch auf beliebige Familien messbarer Mengen und Mengensysteme übertragen.

Definition 26 (stochastische Unabhängigkeit von Ereignissen). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\{A_i\}_{i \in I}$ eine beliebige Familie messbarer Mengen. Die Familie heißt *stochastisch unabhängig*, wenn die Familie $\{\mathbb{1}_{A_i}\}_{i \in I}$ der zugehörigen Indikatorfunktionen stochastisch unabhängig ist. Eine Familie $\{\mathcal{F}_i\}_{i \in I}$ mit $\mathcal{F}_i \subseteq \mathcal{A}$ für alle $i \in I$ heißt entsprechend stochastisch unabhängig, wenn jede Familie von Auswahlmengen $\{F_i\}_{i \in I}$ mit $F_i \in \mathcal{F}_i$ für alle $i \in I$ stochastisch unabhängig ist.

Im Sinne von Lemma 25 bedeutet dies äquivalent:

Lemma 26. In der Situation von Definition 26 ist die Familie $\{A_i\}_{i \in I}$ genau dann stochastisch unabhängig, wenn für jede endliche Teilindexmenge $J_m = \{j_1, \dots, j_m\} \subseteq I$, $m \in \mathbb{N}$ gilt:

$$P\left(\bigcap_{k=1}^m A_{j_k}\right) = \prod_{k=1}^m P(A_{j_k}).$$

Ferner gilt: eine endliche Familie $\{A_i\}_{i \in I}$ von Ereignissen mit $I = \{1, \dots, m\}$, $m \in \mathbb{N}$ ist genau dann stochastisch unabhängig, wenn gilt:

$$P\left(\bigcap_{k=1}^m C_k\right) = \prod_{k=1}^m P(C_k) \text{ für alle Wahlen } C_k \in \{A_k, A_k^c\}, 1 \leq k \leq m.$$

In diesem Fall gilt auch noch

$$P\left(\bigcup_{i=1}^m A_i\right) = 1 - \prod_{i=1}^m (1 - P(A_i)).$$

Beweis: Der erste Teil folgt sofort aus Lemma 25 wegen $P(\mathbb{1}_A = 1) = P(A)$, $P(\mathbb{1}_A = 0) = P(A^c)$ für jedes Ereignis $A \in \mathcal{A}$ (vgl. die Rechnung nach Definition 16). Zum zweiten Teil:

" \Rightarrow ": Zunächst gilt $P\left(\bigcap_{k=1}^m C_k\right) = \prod_{k=1}^m P(C_k)$ für $C_k = A_k$, $1 \leq k \leq m$. Wir zeigen, dass diese Beziehung richtig bleibt, wenn ein beliebiges C_r durch A_r^c mit $1 \leq k \leq m$ ersetzt wird. Dann ist nämlich

$$\begin{aligned} P\left(\bigcap_{\substack{k=1 \\ k \neq r}}^m A_k\right) &= P\left(A_r \cap \bigcap_{\substack{k=1 \\ k \neq r}}^m A_k\right) + P\left(A_r^c \cap \bigcap_{\substack{k=1 \\ k \neq r}}^m A_k\right), \text{ also} \\ P\left(A_r^c \cap \bigcap_{\substack{k=1 \\ k \neq r}}^m A_k\right) &= P\left(\bigcap_{\substack{k=1 \\ k \neq r}}^m A_k\right) - P\left(A_r \cap \bigcap_{\substack{k=1 \\ k \neq r}}^m A_k\right) = \prod_{\substack{k=1 \\ k \neq r}}^m P(A_k) - P(A_r) \prod_{\substack{k=1 \\ k \neq r}}^m P(A_k) \\ &= [1 - P(A_r)] \prod_{\substack{k=1 \\ k \neq r}}^m P(A_k) = P(A_r^c) \prod_{\substack{k=1 \\ k \neq r}}^m P(A_k) \end{aligned}$$

II. Grundprinzipien stochastischer Modellierung

wie behauptet. Iterativ kann man auf dieselbe Weise weitere Mengen C_r durch A_r^c mit $1 \leq k \leq m$ ersetzen, woraus die Behauptung folgt.

" \Leftarrow ": Zunächst gilt $P\left(\bigcap_{k=1}^m C_k\right) = \prod_{k=1}^m P(C_k)$ für $C_k = A_k$, $1 \leq k \leq m$ und für $C_k = A_k$, $1 \leq k \leq m-1$, sowie $C_m = A_m^c$. Es folgt ähnlich zuvor

$$\begin{aligned} P\left(\bigcap_{k=1}^{m-1} A_k\right) &= P\left(A_m \cap \bigcap_{k=1}^{m-1} A_k\right) + P\left(A_m^c \cap \bigcap_{k=1}^{m-1} A_k\right) = P(A_m) \prod_{k=1}^{m-1} P(A_k) + P(A_m^c) \prod_{k=1}^{m-1} P(A_k) \\ &= \underbrace{\left[P(A_m) + P(A_m^c)\right]}_{=1} \prod_{k=1}^{m-1} P(A_k) = \prod_{k=1}^{m-1} P(A_k), \end{aligned}$$

d.h. die Aussage gilt auch für $m-1$ statt m Mengen. Wiederum iterativ folgt, dass die Aussage damit für je k aus der Familie $\{A_i\}_{i \in I}$ ausgewählte Mengen gilt mit $1 \leq k \leq m$, was zu zeigen war. Der Rest ergibt sich aus

$$P\left(\bigcup_{i=1}^m A_i\right) = 1 - P\left(\bigcap_{i=1}^m A_i^c\right) = 1 - \prod_{i=1}^m P(A_i^c) = 1 - \prod_{i=1}^m (1 - P(A_i)). \quad \blacksquare$$

Auf den "Test" mit allen endlichen Teilindexmengen in Lemma 26 kann man nicht verzichten, wie das folgende Beispiel zeigt, in dem je zwei Ereignisse stochastisch unabhängig sind, nicht aber alle zugleich. Dazu sei $\Omega = \{1, 2, 3, 4, 5, 6, 7, 8\}$ und $A = \{1, 2, 3, 4\}$, $B = \{3, 4, 5, 6\}$ und $C = \{1, 2, 5, 6\}$. Als σ -Algebra über Ω wählen wir die Potenzmenge und darauf die Laplace-Verteilung. Nach Konstruktion gilt $\#(A \cap B) = \#(A \cap C) = \#(B \cap C) = 2$ und somit

$$P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} = \frac{1}{2} \cdot \frac{1}{2} = P(A) \cdot P(B) = P(A) \cdot P(C) = P(B) \cdot P(C),$$

d.h. A, B, C sind paarweise stochastisch unabhängig. Es gilt jedoch

$$0 = P(\emptyset) = P(A \cap B \cap C) \neq \frac{1}{8} = P(A) \cdot P(B) \cdot P(C),$$

womit eine der erforderlichen Bedingungen für stochastische Unabhängigkeit der Familie $\{A, B, C\}$ verletzt ist.

Der Beweis zu Lemma 26 zeigt allerdings, dass für lediglich *zwei* Ereignisse A und B stochastische Unabhängigkeit genau dann vorliegt, wenn

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

gilt.

Das folgende überraschende Resultat zeigt, dass schon in einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit *abzählbarer* Grundmenge Ω keine abzählbaren Familien stochastisch unabhängiger nicht-trivialer gleichwahrscheinlicher Ereignisse $\{A_n \mid n \in \mathbb{N}\}$ mit $0 < P(A_n) = p < 1$ existieren.

II. Grundprinzipien stochastischer Modellierung

Damit lassen sich bereits recht einfache Experimente wie das Warten auf die erste Sechs beim Würfeln nicht mehr im diskreten Rahmen modellieren.

Satz 27. Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum mit abzählbarer Grundmenge Ω und $p \in (0,1)$ beliebig, aber fest. Dann existiert keine abzählbare Familie stochastisch unabhängiger Ereignisse $\{A_n \mid n \in \mathbb{N}\} \subseteq \mathcal{A}$ mit $P(A_n) = p$ für alle $n \in \mathbb{N}$.

Beweis: Wir führen einen Widerspruchsbeweis und nehmen dazu an, dass eine solche Familie existiert. Für jedes $\omega \in \Omega$ definieren wir die Ereignisse

$$B_n(\omega) := \begin{cases} A_n & \text{falls } \omega \in A_n \\ A_n^c & \text{falls } \omega \in A_n^c \end{cases} \quad \text{für alle } n \in \mathbb{N}.$$

Nach Konstruktion gilt $\omega \in B_n(\omega)$ für alle $\omega \in \Omega$ und $n \in \mathbb{N}$, also auch $\omega \in \bigcap_{n=1}^{\infty} B_n(\omega)$. Wir bezeichnen mit A_ω das jeweils von ω induzierte Atom. Dann gilt auch $A_\omega \subseteq \bigcap_{n=1}^{\infty} B_n(\omega)$. Mit der Stetigkeit von P von oben, der Monotonie von P (Satz 5) und der vorausgesetzten Unabhängigkeit folgt nun für alle $\omega \in \Omega$:

$$P(A_\omega) \leq P\left(\bigcap_{n=1}^{\infty} B_n(\omega)\right) = \lim_{m \rightarrow \infty} P\left(\bigcap_{n=1}^m B_n(\omega)\right) = \lim_{m \rightarrow \infty} \prod_{n=1}^m P(B_n(\omega)) \leq \lim_{m \rightarrow \infty} q^m = 0$$

mit $q := \max\{p, 1-p\} < 1$, also $P(A_\omega) = 0$ für alle $\omega \in \Omega$. Wegen der Abzählbarkeit von Ω gelangt man damit zu folgenden Widerspruch:

$$1 = P(\Omega) = P\left(\bigcup_{\omega \in \Omega} A_\omega\right) \leq \sum_{\omega \in \Omega} P(A_\omega) = 0 \quad \leftarrow$$

Damit ist der Satz bewiesen. ■

Die Veranschaulichung der stochastischen Unabhängigkeit von Ereignissen lässt sich am besten anhand der Definition der elementaren bedingten Wahrscheinlichkeit zeigen. Ein etwas abstrakterer Begriff von bedingten Wahrscheinlichkeiten, der maßtheoretisch motiviert ist, wird in Abschnitt II.8 gesondert behandelt.

Definition 27 (elementare bedingte Wahrscheinlichkeit). Es sei (Ω, \mathcal{A}, P) ein beliebiger Wahrscheinlichkeitsraum und $B \in \mathcal{A}$ ein Ereignis mit $P(B) > 0$. Dann heißt das normierte Spurmaß P_B bezüglich B die *durch B induzierte elementare bedingte Verteilung*, in expliziter Form:

$$P_B(A) := \frac{P(A \cap B)}{P(B)} \quad \text{für alle } A \in \mathcal{A}.$$

Der Vollständigkeit halber definiert man P_B auch noch für Null-Ereignisse B , also solche mit $P(B) = 0$, indem man in diesem Fall $P_B = P$ wählt.

II. Grundprinzipien stochastischer Modellierung

Üblicherweise wählt man für einzelne bedingte Wahrscheinlichkeiten noch die folgende Schreibweise:

$$P(A|B) := P_B(A) \text{ für } A \in \mathcal{A}$$

und nennt dies die "bedingte Wahrscheinlichkeit für das Ereignis A unter [der Hypothese] B ".

Lemma 27. Die stochastische Unabhängigkeit zweier Ereignisse $A, B \in \mathcal{A}$ kann unter den Voraussetzungen von Definition 27 äquivalent charakterisiert werden durch die Gültigkeit wahlweise einer der Beziehungen

$$P(A|B) = P(A) \text{ bzw. } P(B|A) = P(B).$$

Beweis: " \Rightarrow ": Es sei zunächst $P(B) > 0$ angenommen und A und B als unabhängig vorausgesetzt. Dann gilt

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

wie behauptet. Die zweite Gleichung folgt durch Vertauschen von A und B unter der Voraussetzung $P(A) > 0$. Ist nun $P(B) = 0$, so gilt die Gleichung per Definition, analog für die zweite Gleichung, wenn $P(A) = 0$ ist.

" \Leftarrow ": Es gelte zunächst $P(A|B) = P(A)$ mit $P(B) > 0$. Dann folgt

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

und damit nach der Bemerkung hinter dem Beispiel zu Lemma 26 die stochastische Unabhängigkeit von A und B . Im Fall $P(B) = 0$ erhalten wir entsprechend wegen der Monotonie von P

$$0 \leq P(A \cap B) \leq P(B) = 0, \text{ also } P(A \cap B) = 0 = P(A) \cdot P(B),$$

also wieder die stochastische Unabhängigkeit von A und B . Damit ist das Lemma bewiesen. ■

Zur Veranschaulichung der stochastischen Unabhängigkeit betrachten wir jetzt einen beliebigen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) sowie den zugehörigen Produktraum $(\Omega \times \Omega, \mathcal{A} \otimes \mathcal{A}, P \otimes P)$.

Wir können den Produktraum als Beschreibung eines sequenziellen Experiments auffassen, bei dem nacheinander (oder zeitgleich) zwei einzelne Experimente ablaufen, die jeweils durch (Ω, \mathcal{A}, P) beschrieben werden (etwas das Werfen zweier Würfel). Für ein Produktereignis $A \times B \in \mathcal{A} \otimes \mathcal{A}$ mit $P \otimes P(A \times B) > 0$ gilt dann nach Definition des Produktmaßes:

$$P \otimes P((A \times \Omega) \cap (\Omega \times B)) = P \otimes P(A \times B) = P(A) \cdot P(B) = P \otimes P(A \times \Omega) \cdot P \otimes P(\Omega \times B),$$

bzw.

$$P \otimes P(\Omega \times B | A \times \Omega) = \frac{P \otimes P(A \times B)}{P \otimes P(A \times \Omega)} = \frac{P(A) \cdot P(B)}{P(A)} = P(B) = P \otimes P(\Omega \times B).$$

II. Grundprinzipien stochastischer Modellierung

Diese letzte Gleichung besagt gerade, dass die Ereignisse $A \times \Omega$ ("im ersten Experiment tritt A ein") und $\Omega \times B$ ("im zweiten Experiment tritt B ein") unter dem Produktmaß stochastisch unabhängig sind, und dass die Kenntnis des Ausgangs des ersten Experiments (gegeben durch die Hypothese $A \times \Omega$) die (bedingte) Wahrscheinlichkeit des Eintretens von B (unter dieser Hypothese) im zweiten Experiment nicht verändert. Plakativ gesprochen bedeutet stochastische Unabhängigkeit von Ereignissen in Produkträumen also, dass die Kenntnis über das Eintreten oder Nichteintreten von Ereignissen aus vorangehenden Experimenten keinen Einfluss auf die Wahrscheinlichkeiten des Eintretens oder Nichteintretens späterer Experimente hat.

Beim Würfelwurf bedeutet das etwa, dass jede noch so lange Beobachtung von Wurfsergebnissen keinerlei Einfluss auf die Wahrscheinlichkeiten zukünftiger Wurfsergebnisse hat. Aus diesem Grund ist es auch völlig sinnlos, jahrelange Statistiken über die Ergebnisse der Ziehung der Lottozahlen zu studieren in der Hoffnung, man könne hieraus Erkenntnisse über das Auftreten zukünftiger Ziehungsergebnisse gewinnen.

Stochastische Unabhängigkeit bedeutet in diesem Sinne also eine „Informationsunabhängigkeit“. Die oft als plakative Veranschaulichung gebrauchte Formulierung, dass sich die Experimente nicht „gegenseitig beeinflussen“, trifft also nicht den Kern der stochastischen Unabhängigkeit.

Das folgende einfache Beispiel zeigt dagegen eine Situation auf, in der die stochastische Unabhängigkeit sequentieller Ereignisse verletzt wird. Wir "simulieren" darin einen zweifachen (abhängigen) Münzwurf mit Hilfe eines Würfelwurfs. Als Ausgangsraum für das Würfelexperiment verwenden wir wieder die Menge $\Omega = \{1, 2, 3, 4, 5, 6\}$, darüber die Potenzmenge als σ -Algebra \mathcal{A} und darauf die Laplace-Verteilung P . Als spezielle Ereignisse wählen wir noch $G := \{2, 4, 6\}$, $U := G^c = \{1, 3, 5\}$ (gerade und ungerade Wurfsergebnisse) sowie $E := \{1, 2, 3\}$ und $L := E^c = \{4, 5, 6\}$ (erste und letzte Hälfte). Dann definieren wir den Zufallsvektor $X = (X_1, X_2)$ durch

$$X_1 := \mathbb{1}_G \text{ und } X_2 := \mathbb{1}_E.$$

Beide Zufallsvariablen besitzen jeweils eine Laplace-Verteilung auf der Menge $\{0, 1\}$ (als Repräsentanten von etwa Kopf / Zahl) und können daher auch wegen $P(X_1 = 1) = P(X_2 = 1) = \frac{3}{6} = \frac{1}{2}$ jeweils als Beschreibung des Werfens einer fairen Münze dienen. Es gilt aber

$$P(X_1 = 1, X_2 = 1) = P(G \cap E) = P(\{2\}) = \frac{1}{6} \neq \frac{1}{4} = P(X_1 = 1) \cdot P(X_2 = 1),$$

so dass X_1 und X_2 nicht stochastisch unabhängig sind.

Wir wollen zum Abschluss dieses Abschnitts noch einige wichtige Aussagen betrachten, die mit den Eigenschaften der stochastischen Unabhängigkeit im Zusammenhang stehen.

Lemma 28 (Unabhängigkeit bei Blockbildung). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(\mathcal{X}, \mathcal{B})$ ein Messraum. $X_i : (\Omega, \mathcal{A}) \rightarrow (\mathcal{X}, \mathcal{B})$ seien stochastisch unabhängig (Zufallsvariablen, Zufallsvektoren oder Zufallselemente) mit $i \in I$; die Indexmenge I darf dabei beliebig nicht-leer gewählt sein. Sind dann g_j geeignet messbare Abbildungen auf dem Produktraum $\left(\prod_{i \in I_j} \mathcal{X}, \otimes_{i \in I_j} \mathcal{B} \right)$ mit

II. Grundprinzipien stochastischer Modellierung

Werten in einem Messraum $(\mathfrak{X}_j, \mathcal{B}_j)$, wobei die I_j , $j \in J$ nichtleere, paarweise disjunkte Teilmengen von I seien, dann ist auch die Familie $\left\{ g_j \left(\{X_i\}_{i \in I_j} \right) \right\}_{j \in J}$ stochastisch unabhängig.

Beweis: Auf Grund der Struktur von Produkt- σ -Algebren und Lemma 24 genügt es, die Aussage für den Fall zu beweisen, dass die Indexmenge $I = \{1, \dots, m\}$ mit $m \in \mathbb{N}$ endlich ist. Zur Vereinfachung nehmen wir ferner an, dass $J = \{1, 2\}$ zweielementig ist (der allgemeine Fall geht analog), und $I_1 = \{1, \dots, k\}$, $I_2 = \{k+1, \dots, m\}$ gilt mit $k \in \{1, \dots, m\}$. Dann folgt für $B_j \in \mathcal{B}_j$, $j \in J$:

$$\begin{aligned}
 P^{(g_1(X_1, \dots, X_k), g_2(X_{k+1}, \dots, X_m))} (B_1 \times B_2) &= P \left(\{g_1(X_1, \dots, X_k) \in B_1\} \cap \{g_2(X_{k+1}, \dots, X_m) \in B_2\} \right) \\
 &= P \left(\{(X_1, \dots, X_k) \in g_1^{-1}(B_1)\} \cap \{(X_{k+1}, \dots, X_m) \in g_2^{-1}(B_2)\} \right) \\
 &= P \left((X_1, \dots, X_m) \in g_1^{-1}(B_1) \times g_2^{-1}(B_2) \right) = P^{(X_1, \dots, X_m)} (g_1^{-1}(B_1) \times g_2^{-1}(B_2)) \\
 &= \bigotimes_{i=1}^m P^{X_i} (g_1^{-1}(B_1) \times g_2^{-1}(B_2)) = \bigotimes_{i=1}^k P^{X_i} (g_1^{-1}(B_1)) \cdot \bigotimes_{i=k+1}^m P^{X_i} (g_2^{-1}(B_2)) \\
 &= P^{(g_1(X_1, \dots, X_k))} (B_1) \cdot P^{(g_2(X_{k+1}, \dots, X_m))} (B_2),
 \end{aligned}$$

d.h. $g_1(X_1, \dots, X_k)$ und $g_2(X_{k+1}, \dots, X_m)$ sind stochastisch unabhängig. ■

Lemma 28 besagt also, dass die Eigenschaft der stochastischen Unabhängigkeit erhalten bleibt, wenn man aus beliebigen Familien von unabhängigen Zufallsvariablen / Zufallsvektoren / Zufallselementen durch messbare Transformationen aus disjunkten Teilauswahlen neue Zufallsvariablen / Zufallsvektoren / Zufallselemente bildet. Werfen wir beispielsweise in unabhängigen Versuchen einen Würfel neun Mal und bilden aus je drei konsekutiven Wurfsergebnissen X_i die Augensumme, so erhalten wir drei neue unabhängige Zufallsvariablen. Dagegen sind die Augensummen der ersten drei Würfe und der Würfe 3,4,5 nicht stochastisch unabhängig, weil der dritte Wurf in beiden Augensummen vorkommt, wie man an der nachfolgenden Rechnung sieht:

$$\begin{aligned}
 P \left(\sum_{i=1}^3 X_i = 3 \right) &= P \left(\sum_{i=3}^5 X_i = 18 \right) = \frac{1}{216}, \text{ aber} \\
 P \left(\left\{ \sum_{i=1}^3 X_i = 3 \right\} \cap \left\{ \sum_{i=3}^5 X_i = 18 \right\} \right) &= P(\emptyset) = 0 \neq \frac{1}{46656} = P \left(\sum_{i=1}^3 X_i = 3 \right) \cdot P \left(\sum_{i=3}^5 X_i = 18 \right).
 \end{aligned}$$

Zum Themenkreis "bedingte Wahrscheinlichkeiten" gehört noch der folgende Satz:

II. Grundprinzipien stochastischer Modellierung

Satz 28 (von der totalen Wahrscheinlichkeit bzw. von Bayes): Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\{B_i\}_{i \in I} \subseteq \mathcal{A}$ eine disjunkte Zerlegung von Ω , also $\Omega = \bigoplus_{i \in I} B_i$ mit einer höchstens abzählbaren Indexmenge I . Ferner gelte $P(B_i) > 0$ für alle $i \in I$. Dann gilt:

a) $P(A) = \sum_{i \in I} P(A | B_i) \cdot P(B_i)$ für alle $A \in \mathcal{A}$

b) $P(B_j | A) = \frac{P(A | B_j) \cdot P(B_j)}{\sum_{i \in I} P(A | B_i) \cdot P(B_i)}$ für alle $A \in \mathcal{A}$ mit $P(A) > 0$ und alle $j \in I$.

Beweis: Es ist $P(A | B_i) \cdot P(B_i) = \frac{P(A \cap B_i)}{P(B_i)} \cdot P(B_i) = P(A \cap B_i)$ für alle $i \in I$, also

$$\sum_{i \in I} P(A | B_i) \cdot P(B_i) = \sum_{i \in I} P(A \cap B_i) = P\left(\bigcup_{i \in I} (A \cap B_i)\right) = P\left(A \cap \bigcup_{i \in I} B_i\right) = P(A \cap \Omega) = P(A),$$

wie in Teil a) behauptet. Teil b) ergibt sich entsprechend aus

$$\frac{P(A | B_j) \cdot P(B_j)}{\sum_{i \in I} P(A | B_i) \cdot P(B_i)} = \frac{P(A \cap B_j)}{P(A)} = P(B_j | A) \text{ für } P(A) > 0. \quad \blacksquare$$

Bemerkung: Die Voraussetzung $P(B_i) > 0$ für alle $i \in I$ in Satz 28 ist keine wesentliche Einschränkung, da man anderenfalls alle eventuellen Nullmengen B_j mit $j \in J \subset I$ zu einer gemeinsamen Nullmenge $N := \bigcup_{j \in J} B_j$ vereinigen kann und dann eine (feste) Menge B_k zu $B_k \cup N$ erweitert.

Satz 28 ist von besonderer Bedeutung in den so genannten *Expertensystemen* (z.B. in der Medizin). Dazu betrachten wir folgendes Beispiel.

Aus medizinischen Untersuchungen sei bekannt, dass die Symptome A_1 und A_2 bei (genau) drei Krankheiten B_1 , B_2 und B_3 auftreten können, und zwar mit den (bedingten) Wahrscheinlichkeiten $\pi_{ij} = P(A_j | B_i)$, $i = 1, 2, 3$, $j = 1, 2$, die wir zweckmäßiger Weise in Matrixform

$$\Pi = (\pi_{ij}) = \begin{pmatrix} 0,8 & 0,2 \\ 0,2 & 0,8 \\ 0,4 & 0,6 \end{pmatrix}$$

notieren. Die Eintrittswahrscheinlichkeiten für die Krankheiten B_1 , B_2 und B_3 notieren wir als Zeilenvektor in der Form

$$\mathbf{p} = (p_i) = (0,3 \quad 0,6 \quad 0,1).$$

II. Grundprinzipien stochastischer Modellierung

Nach Satz 28 lässt sich dann der Vektor $\mathbf{q} = (q_j) = (P(A_1) \ P(A_2))$ der Eintrittswahrscheinlichkeiten für die Symptome darstellen als

$$\mathbf{q} = \mathbf{p}\Pi = (0,3 \ 0,6 \ 0,1) \cdot \begin{pmatrix} 0,8 & 0,2 \\ 0,2 & 0,8 \\ 0,4 & 0,6 \end{pmatrix} = (0,4 \ 0,6).$$

Für die bedingten Wahrscheinlichkeiten

$$\pi_{ji}^* = P(B_i | A_j), \quad j = 1, 2, \quad i = 1, 2, 3$$

dafür, dass Krankheit B_i vorliegt, wenn man Symptom A_j beobachtet hat, erhält man nun aus Teil b) des Satzes

$$\pi_{ji}^* = \frac{\pi_{ij} P_i}{q_j}$$

oder gleich in Matrixform (auf 4 Stellen gerundet)

$$\Pi^* = (\pi_{ji}^*) = \begin{pmatrix} 0,6 & 0,3 & 0,1 \\ 0,1 & 0,8 & 0,1 \end{pmatrix}.$$

Beachtenswert ist hierbei, dass zwar – wie erwartet – bei Auftreten des Symptoms A_1 ein Patient am wahrscheinlichsten an der Krankheit B_1 (mit Wahrscheinlichkeit 0,6) und bei Auftreten des Symptoms A_2 am häufigsten an der Krankheit B_2 (mit Wahrscheinlichkeit 0,8) leidet, dass er aber z.B. bei Auftreten des Symptoms A_1 – wenn er *nicht* an B_1 erkrankt ist – häufiger an der Krankheit B_2 (mit Wahrscheinlichkeit 0,75¹⁶) als an B_3 (mit Wahrscheinlichkeit 0,25) leidet, obwohl A_1 häufiger unter der Krankheit B_3 (mit Wahrscheinlichkeit 0,4) als unter B_2 (mit Wahrscheinlichkeit 0,2) vorkommt. Dies liegt daran, dass die Krankheit B_3 selbst nur relativ selten auftritt: B_2 kommt etwa dreimal so häufig vor wie B_3 .

Im Zusammenhang mit bedingten Wahrscheinlichkeiten gibt es eine ganze Reihe von Seltsamkeiten, darunter das bekannte *Simpson-Paradox* aus dem Jahr 1951 (siehe etwa BARTH UND HALLER (1998), S. 140, Aufgabe 18) oder die Problematik der richtigen Diagnose beim AIDS-Test oder ähnlichen medizinischen Screening-Verfahren (vgl. etwa HENZE (2003), Beispiele 16.11 und 16.12 und BEHENEN UND NEUHAUS (2003), Beispiel 6.9.)

¹⁶ es geht hier um die bedingte Wahrscheinlichkeit

$$P_{A_1}(B_2 | B_1^c) = \frac{P_{A_1}(B_2 \cap B_1^c)}{P_{A_1}(B_1^c)} = \frac{P_{A_1}(B_2)}{P_{A_1}(B_1^c)} = \frac{P(B_2 | A_1)}{P(B_1^c | A_1)} = \frac{0,3}{0,3 + 0,1} = 0,75$$

II.3. Verteilungsfunktionen und Momente

Wir greifen hier die schon in Abschnitt I.3 behandelte Frage wieder auf, durch welche Angaben ein (Wahrscheinlichkeits-)Maß bereits eindeutig bestimmt ist. Für Zufallsvariablen und -vektoren gelangt man damit in natürlicher Weise zum Begriff der *Verteilungsfunktion*.

Definition 28 (Verteilungsfunktion). Es sei P ein Wahrscheinlichkeitsmaß auf der Borel'schen σ -Algebra \mathcal{B}^d , $d \in \mathbb{N}$. Die durch

$$F_P(x_1, \dots, x_d) := P\left(\prod_{i=1}^d (-\infty, x_i]\right), \quad (x_1, \dots, x_d) \in \mathbb{R}^d$$

definierte Funktion heißt die zu P gehörige (d -dimensionale) *Verteilungsfunktion*.

Verteilungsfunktionen besitzen wesentliche strukturelle Eigenschaften, von denen wir zeigen werden, dass sie Wahrscheinlichkeitsmaße eindeutig festlegen.

Definition 29 (Δ -Monotonie). Eine Funktion $g: \mathbb{R}^d \rightarrow \mathbb{R}$ mit $d \in \mathbb{N}$ heißt Δ -*monoton*, wenn sie die folgende Eigenschaft besitzt:

$$\Delta g_{\mathbf{x}}^{\mathbf{y}} := \sum_{(\varepsilon_1, \dots, \varepsilon_d) \in \{0,1\}^d} (-1)^{\sum_{i=1}^d \varepsilon_i} g(\varepsilon_1 x_1 + (1-\varepsilon_1)y_1, \dots, \varepsilon_d x_d + (1-\varepsilon_d)y_d) \geq 0$$

für alle $\mathbf{x} = (x_1, \dots, x_d)$, $\mathbf{y} = (y_1, \dots, y_d) \in \mathbb{R}^d$ mit $\mathbf{x} \leq \mathbf{y}$, d.h. $x_i \leq y_i$ für $i = 1, \dots, d$.

Für $d = 1$ fällt also die Δ -Monotonie mit der üblichen (schwachen) Monotonie zusammen.

Lemma 29 (Eigenschaften von Verteilungsfunktionen). Es sei $F = F_P$ die Verteilungsfunktion einer Wahrscheinlichkeitsverteilung P auf der Borel'schen σ -Algebra \mathcal{B}^d , $d \in \mathbb{N}$. Dann gilt:

$\Delta F_{\mathbf{x}}^{\mathbf{y}} \geq 0$ für alle $\mathbf{x}, \mathbf{y} \in \mathbb{R}^d$ mit $\mathbf{x} \leq \mathbf{y}$ (Δ -Monotonie von F)

$$P\left(\prod_{i=1}^d (x_i, y_i]\right) = \Delta F_{\mathbf{x}}^{\mathbf{y}} \text{ für alle } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d \text{ mit } \mathbf{x} \leq \mathbf{y}$$

$\lim_{y_1 \downarrow x_1, \dots, y_d \downarrow x_d} F(y_1, \dots, y_d) = F(x_1, \dots, x_d)$ für alle $\mathbf{x} \in \mathbb{R}^d$ (rechtsseitige Stetigkeit)

$\lim_{x_1 \uparrow \infty, \dots, x_d \uparrow \infty} F(x_1, \dots, x_d) = 1$, $\lim_{x_i \downarrow -\infty} F(x_1, \dots, x_d) = 0$ für $i = 1, \dots, d$

(Normiertheit / rechtsseitige Stetigkeit).

Beweis: Die erste Aussage folgt aus der zweiten, die wir hier der Einfachheit halber nur für den Fall $d = 2$ zeigen (der allgemeine Fall kann ähnlich gezeigt werden wie die Siebformel aus Satz 5). Es ist nämlich

$$(-\infty, y_1] \times (-\infty, y_2] \setminus (x_1, y_1] \times (x_2, y_2] = (-\infty, x_1] \times (-\infty, y_2] \cup (-\infty, y_1] \times (-\infty, x_2]$$

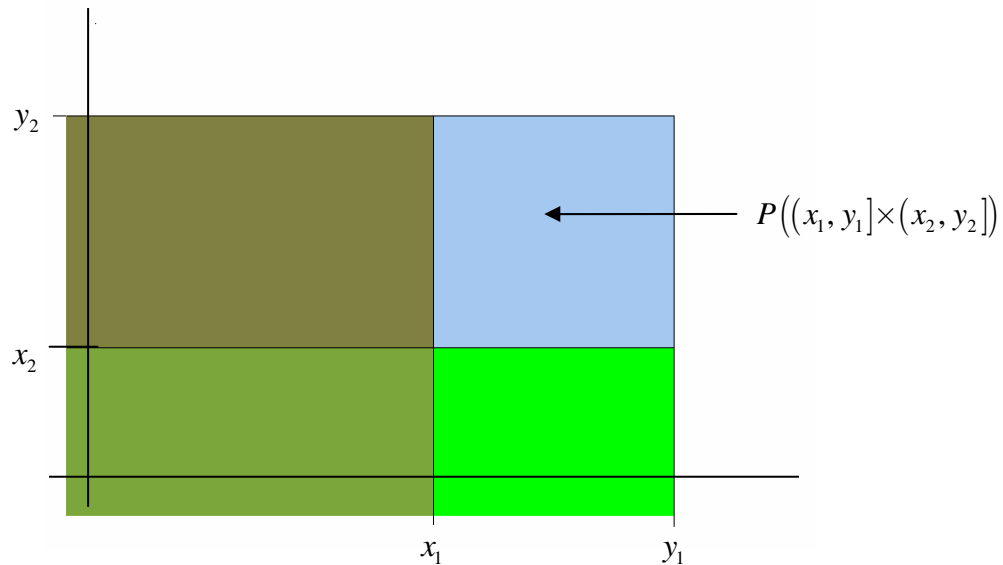
und somit

II. Grundprinzipien stochastischer Modellierung

$$F(y_1, y_2) - P((x_1, y_1] \times (x_2, y_2]) = F(x_1, y_2) + F(y_1, x_2) - F(x_1, x_2),$$

also

$$P((x_1, y_1] \times (x_2, y_2]) = F(x_1, x_2) - F(x_1, y_2) - F(y_1, x_2) + F(y_1, y_2) = \Delta F_{\mathbf{x}}^y.$$



Der Rest ergibt sich aus der Stetigkeit von P von oben und unten (da P ein beschränktes und damit endliches Maß ist). ■

Die zentrale Bedeutung von Verteilungsfunktionen wird aus dem folgenden Resultat ersichtlich.

Satz 29 (Charakterisierungssatz). Es sei F eine Δ -monotone Funktion auf \mathbb{R}^d für $d \in \mathbb{N}$, die die drei Grenzwerteigenschaften von Lemma 29 erfülle. Dann gibt es genau ein Wahrscheinlichkeitsmaß P auf der Borel'schen σ -Algebra \mathcal{B}^d , so dass $F = F_P$ ist.

Beweis: Wir zeigen hier nur die Beweiseidee auf und verweisen für Einzelheiten auf BAUER (1992). Durch

$$P\left(\prod_{i=1}^d (x_i, y_i]\right) = \Delta F_{\mathbf{x}}^y \text{ für alle } \mathbf{x}, \mathbf{y} \in \mathbb{R}^d \text{ mit } \mathbf{x} \leq \mathbf{y}$$

wird zunächst eine nicht-negative, normierte Mengenfunktion P auf dem Semi-Ring \mathcal{E}^d der Intervalle definiert, die additiv ist. Mit Hilfe der ersten und dritten (rechtsseitigen) Stetigkeitseigenschaft lässt sich wie beim Lebesgue-Maß (Satz 14) mit einem geeigneten Kompaktheitsargument dann zeigen, dass P auf \mathcal{E}^d sogar σ -additiv ist. Damit existiert nach Satz 7 (großer Maßfortsetzungssatz) eine Fortsetzung von P zu einem Maß auf der Borel'schen σ -Algebra \mathcal{B}^d , die wir wieder mit P bezeichnen wollen. P ist eindeutig, weil P beschränkt und damit insbesondere σ -endlich ist. Die zweite Stetigkeitseigenschaft impliziert dann noch

$$P(\mathbb{R}^d) = \lim_{x_i \uparrow \infty, \dots, x_d \uparrow \infty} P\left(\prod_{i=1}^d (-\infty, x_i]\right) = \lim_{x_i \uparrow \infty, \dots, x_d \uparrow \infty} F(x_1, \dots, x_d) = 1,$$

so dass P tatsächlich auch ein Wahrscheinlichkeitsmaß ist. ■

II. Grundprinzipien stochastischer Modellierung

Eine Verteilungsfunktion spielt damit dieselbe Rolle wie der Elementarinhalt beim Lebesgue-Maß; sie beinhaltet insbesondere schon sämtliche notwendigen Informationen über das zugehörige Wahrscheinlichkeitsmaß.

Beispielsweise ist die Funktion $F(x_1, x_2) := x_1 + x_2$, $x_1, x_2 \in \mathbb{R}$ Δ -monoton wegen

$$\begin{aligned}\Delta F_{\mathbf{x}}^y &= F(x_1, x_2) - F(x_1, y_2) - F(x_2, y_1) + F(y_1, y_2) \\ &= (x_1 + x_2) - (x_1 + y_2) - (x_2 + y_1) + (y_1 + y_2) = 0,\end{aligned}$$

jedoch ist F keine Verteilungsfunktion wegen z.B. $\lim_{x_1 \uparrow \infty, \dots, x_d \uparrow \infty} F(x_1, \dots, x_d) = \infty$.

Ein wichtiger Spezialfall von Verteilungsfunktionen liegt vor, wenn das zugehörige Wahrscheinlichkeitsmaß P eine Dichte f bezüglich des Lebesgue-Maßes m^d besitzt. In diesem Fall gilt nämlich

$$\Delta F_{\mathbf{x}}^y = P\left(\bigotimes_{i=1}^d (x_i, y_i]\right) = \int_{x_d}^{y_d} \cdots \int_{x_1}^{y_1} f(u_1, \dots, u_d) du_1 \cdots du_d.$$

Ist die Dichte f darüber hinaus in dem Punkt $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$ stetig, folgt nach dem (mehrdimensionalen) Hauptsatz der Differential- und Integralrechnung sogar noch

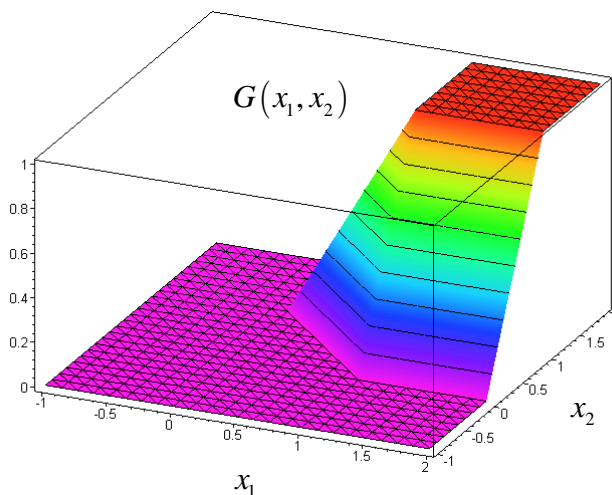
$$f(x_1, \dots, x_d) = \frac{\partial^d}{\partial x_1 \cdots \partial x_d} F(x_1, \dots, x_d).$$

Damit kann bei Riemann-Integrierbarkeit die Verteilungsfunktion F als Stammfunktion zur Dichte f angesehen werden.

Die oben als Beispiel betrachtete Δ -monotone Funktion $F(x_1, x_2) = x_1 + x_2$, $x_1, x_2 \in \mathbb{R}$ kann durch eine kleine Modifikation zu einer „echten“ Verteilungsfunktion gemacht werden, indem man

$$G(x_1, x_2) := \max\{\min\{x_1, 1\} + \min\{x_2, 1\} - 1, 0\}, \quad x_1, x_2 \in \mathbb{R}$$

setzt. Dann gilt $G(x_1, x_2) = F(x_1, x_2) - 1$ in dem Dreieck $D := \{\mathbf{x} \in \mathbb{R}^2 \mid 0 \leq x_1, x_2 \leq 1, x_1 + x_2 \geq 1\}$.



II. Grundprinzipien stochastischer Modellierung

Die zugehörige Wahrscheinlichkeitsverteilung P besitzt aber keine Dichte bezüglich des Lebesgue-Maßes m^2 , da z.B. im Inneren D° des Dreiecks D gilt:

$$\frac{\partial^2}{\partial x_1 \partial x_2} G(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} (x_1 + x_2 - 1) = 0 \text{ für alle } \mathbf{x} \in D^\circ$$

und ebenso

$$\frac{\partial^2}{\partial x_1 \partial x_2} G(x_1, x_2) = 0 \text{ für alle } \mathbf{x} \in D^c.$$

Solche Verteilungen, die auch offensichtlich nicht diskret sind, heißen *singulär* (bezüglich des Lebesgue-Maßes).

Satz 30 (Verteilungsfunktionen bei Unabhängigkeit). Es seien X_1, \dots, X_d reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit den Verteilungsfunktionen F_1, \dots, F_d (d.h. zu den Verteilungen P^{X_1}, \dots, P^{X_d}). Dann gilt: X_1, \dots, X_d sind stochastisch unabhängig genau dann, wenn für die Verteilungsfunktion F der gemeinsamen Verteilung $P^{(X_1, \dots, X_d)}$ gilt:

$$F(x_1, \dots, x_d) = \prod_{i=1}^d F_i(x_i), \quad \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d.$$

Beweis: " \Rightarrow ": Es ist mit $\mathbf{X} = (X_1, \dots, X_d)$ bei Unabhängigkeit

$$\begin{aligned} F(x_1, \dots, x_d) &= P\left(\mathbf{X} \in \bigtimes_{i=1}^d (-\infty, x_i]\right) = P^{\mathbf{X}}\left(\bigtimes_{i=1}^d (-\infty, x_i]\right) \\ &= \bigotimes_{i=1}^d P^{X_i}\left(\bigtimes_{i=1}^d (-\infty, x_i]\right) = \prod_{i=1}^d P^{X_i}\left((-\infty, x_i]\right) \\ &= \prod_{i=1}^d F_i(x_i), \quad \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d. \end{aligned}$$

" \Leftarrow ": Es ist

$$P^{\mathbf{X}}\left(\bigtimes_{i=1}^d (-\infty, x_i]\right) = F(x_1, \dots, x_d) = \prod_{i=1}^d F_i(x_i) = \prod_{i=1}^d P^{X_i}\left((-\infty, x_i]\right) = \bigotimes_{i=1}^d P^{X_i}\left(\bigtimes_{i=1}^d (-\infty, x_i]\right)$$

für alle $\mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d$. Also stimmt $P^{\mathbf{X}}$ auf dem Semi-Ring \mathcal{E}^d der Intervalle mit dem Produktmaß $\bigotimes_{i=1}^d P^{X_i}$ überein und ist daher nach Satz 8 mit diesem identisch, d.h. X_1, \dots, X_d sind stochastisch unabhängig. ■

II. Grundprinzipien stochastischer Modellierung

Im Fall der Existenz einer Dichte f von F bezüglich des Lebesgue-Maßes m^d wie oben ist die fundamentale Gleichung in Satz 30 auch noch äquivalent zu

$$f(x_1, \dots, x_d) = \prod_{i=1}^d f_i(x_i) \quad m^d\text{-fast überall, } \mathbf{x} = (x_1, \dots, x_d) \in \mathbb{R}^d,$$

wobei die f_i die (eindimensionalen) Lebesgue-Dichten der P^{x_i} bezeichnen.

Definition 30 (Randverteilungen). Es sei $\left(\prod_{i \in I} \Omega_i, \otimes_{i \in I} \mathcal{A}_i\right)$ ein Produkt-Messraum mit einer abzählbaren Indexmenge $I \subseteq \mathbb{N}$ und P ein Wahrscheinlichkeitsmaß auf $\otimes_{i \in I} \mathcal{A}_i$. Das für jede Teilindexmenge $J \subseteq I$ auf dem Produkt-Messraum $\left(\prod_{i \in J} \Omega_i, \otimes_{i \in J} \mathcal{A}_i\right)$ durch

$$P_J\left(\prod_{j \in J} A_j\right) := P\left(\prod_{i \in I} B_i\right) \quad \text{mit } B_i := \begin{cases} A_i, & i \in J \\ \Omega_i, & i \in I \setminus J \end{cases}$$

eindeutig bestimmte Wahrscheinlichkeitsmaß P_J heißt Randverteilung der Ordnung J zu P .

Man beachte dabei, dass $\mathcal{E}_J := \left\{ \prod_{j \in J} A_j \mid \forall k \in J : A_k \in \mathcal{A}_k \right\}$ ein $\prod_{j \in J} \Omega_j$ abzählbar überdeckender Semi-Ring und Erzeuger der Produkt- σ -Algebra $\otimes_{i \in J} \mathcal{A}_i$ ist, auf dem P_J σ -additiv ist (weil P auf ganz $\otimes_{i \in I} \mathcal{A}_i$ diese Eigenschaft besitzt). Die Existenz und Eindeutigkeit von P_J ergibt sich damit nach dem (großen Maßfortsetzungs-) Satz 7.

Im Kontext von Verteilungsfunktionen besagt dies Folgendes:

Lemma 30. Es sei F die zu einem Wahrscheinlichkeitsmaß P auf \mathcal{B}^d gehörige Verteilungsfunktion auf \mathbb{R}^d für $d \geq 2$ und $J = \{i_1, \dots, i_r\} \subseteq I := \{1, \dots, d\}$ mit einer natürlichen Zahl $r < d$. Ferner sei $K := I \setminus J = \{j_1, \dots, j_{d-r}\}$ die der Größe nach angeordnete Menge der komplementären Indices. Dann ist die zu dem Wahrscheinlichkeitsmaß P_J gehörige Verteilungsfunktion F_J gegeben durch

$$F_J(x_{i_1}, \dots, x_{i_r}) = \lim_{x_{j_1} \rightarrow \infty} \cdots \lim_{x_{j_{d-r}} \rightarrow \infty} F(x_1, \dots, x_d), \quad (x_{i_1}, \dots, x_{i_r}) \in \mathbb{R}^r.$$

Besitzt F darüber hinaus eine Dichte f bezüglich des Lebesgue-Maßes m^d , so besitzt F_J ebenfalls eine Dichte (bezüglich des Lebesgue-Maßes m^r), die gegeben ist durch

$$f_J(x_{i_1}, \dots, x_{i_r}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_d) dx_{j_1} \cdots dx_{j_{d-r}}, \quad (x_{i_1}, \dots, x_{i_r}) \in \mathbb{R}^r.$$

II. Grundprinzipien stochastischer Modellierung

Beweis: Der erste Teil folgt sofort aus Definition 30 mit der Wahl $A_i := (-\infty, x_i]$, $x_i \in \mathbb{R}$. Der zweite Teil ergibt sich aus der Darstellung der (mehrdimensionalen) Verteilungsfunktion als Lebesgue-Integral. ■

Wir betrachten beispielhaft die Dichte

$$f(x_1, x_2) = (x_1 + x_2) \cdot \mathbb{1}_Q(x_1, x_2), \quad \mathbf{x} \in \mathbb{R}^2$$

mit dem (abgeschlossenen) Einheitsquadrat Q . Man beachte, dass hier tatsächlich

$$\begin{aligned} \iint_{\mathbb{R}^2} f(x_1, x_2) dx_1 dx_2 &= \int_0^1 \int_0^1 x_1 + x_2 dx_1 dx_2 = \int_0^1 \left. \frac{x_1^2}{2} + x_1 x_2 \right|_0^1 dx_2 \\ &= \int_0^1 \left. \frac{1}{2} + x_2 \right|_0^1 dx_2 = \left. \frac{x_2}{2} + \frac{x_2^2}{2} \right|_0^1 = 1 \end{aligned}$$

gilt. Es folgt mit analoger Rechnung

$$F(x_1, x_2) = \int_0^{x_2} \int_0^{x_1} u_1 + u_2 du_1 du_2 = \int_0^{x_2} \left. \frac{x_1^2}{2} + x_1 u_2 \right|_0^{x_1} du_2 = \frac{x_1^2 x_2 + x_1 x_2^2}{2} = \frac{1}{2} x_1 x_2 (x_1 + x_2), \quad (x_1, x_2) \in Q$$

und daraus für die Randverteilungsfunktionen

$$F_1(x_1) = \lim_{x_2 \rightarrow \infty} F(x_1, x_2) = F(x_1, 1) = \frac{1}{2} x_1 (x_1 + 1), \quad 0 \leq x_1 \leq 1$$

sowie aus Symmetriegründen

$$F_2(x_2) = F(1, x_2) = \frac{1}{2} x_2 (x_2 + 1), \quad 0 \leq x_2 \leq 1.$$

Für die Randdichten ergibt sich durch Integration (oder Differenzieren der gerade berechneten Verteilungsfunktionen) noch

$$f_1(x_1) = \int_0^1 f(x_1, x_2) dx_2 = \int_0^1 x_1 + x_2 dx_2 = x_1 + \frac{1}{2}, \quad 0 < x_1 < 1$$

$$f_2(x_2) = \int_0^1 f(x_1, x_2) dx_1 = x_2 + \frac{1}{2}, \quad 0 < x_2 < 1.$$

Hieran sieht man übrigens, dass das zugehörige Wahrscheinlichkeitsmaß P kein Produktmaß ist, weil im Allgemeinen nicht $F(x_1, x_2) = F_1(x_1) \cdot F_2(x_2)$ für $(x_1, x_2) \in Q$ ist.

II. Grundprinzipien stochastischer Modellierung

Wir kehren noch einmal ausführlicher zu dem vorigen Beispiel

$$G(x_1, x_2) := \max \{ \min \{x_1, 1\} + \min \{x_2, 1\} - 1, 0 \}, \quad x_1, x_2 \in \mathbb{R}$$

zurück. Hier gilt entsprechend

$$\begin{aligned} G_1(x_1) &= \lim_{x_2 \rightarrow \infty} G(x_1, x_2) = G(x_1, 1) = x_1, \quad 0 \leq x_1 \leq 1 \\ G_2(x_2) &= \lim_{x_1 \rightarrow \infty} G(x_1, x_2) = G(1, x_2) = x_2, \quad 0 \leq x_2 \leq 1. \end{aligned}$$

Die Randverteilungsfunktionen G_1 und G_2 gehören damit zur *stetigen Gleichverteilung über dem Einheitsintervall* $E := [0, 1]$ mit den Randdichten

$$f_1 = f_2 = \mathbb{1}_E.$$

Hier liegt also der interessante Fall vor, dass zwar die beiden Randverteilungen Dichten bezüglich des Lebesgue-Maßes m^1 besitzen, die gemeinsame Verteilung allerdings keine Dichte bezüglich des Lebesgue-Maßes m^2 . Man kann sich diese Verteilung jedoch folgendermaßen vorstellen: sei X eine über dem Einheitsintervall E stetig gleichverteilte Zufallsvariable. Dann besitzt der Zufallsvektor $\mathbf{X} := (X, 1 - X)$ bzw. die Verteilung $P^{\mathbf{X}}$ die (singuläre) Verteilungsfunktion G . Dies sieht man durch die Rechnung

$$P(X \leq x_1, 1 - X \leq x_2) = P(1 - x_2 \leq X \leq x_1) = \begin{cases} x_1 + x_2 - 1, & \text{falls } 1 - x_2 \leq x_1 \\ 0, & \text{sonst} \end{cases} = G(x_1, x_2)$$

für $\mathbf{x} \in D$. Ferner ist

$$P(X \leq x_1, 1 - X \leq x_2) = P(1 - x_2 \leq X) = 1 - (1 - x_2) = x_2 \quad \text{für } x_1 > 1, x_2 \in E \text{ sowie}$$

$$P(X \leq x_1, 1 - X \leq x_2) = P(X \leq x_1) = x_1 \quad \text{für } x_2 > 1, x_1 \in E,$$

woraus insgesamt folgt, dass G tatsächlich die Verteilungsfunktion von \mathbf{X} ist. Darüber hinaus sieht man, dass mit X auch $1 - X$ über dem Einheitsintervall E stetig gleichverteilt ist. Allerdings sind X und $1 - X$ in extremer Weise voneinander funktional (nämlich linear) abhängig, damit ist die (messbare) Menge D , für die $P(\mathbf{X} \in D) = 1$ gilt, die Anti-Diagonale

$$D = \{(x_1, x_2) \in E^2 \mid x_1 + x_2 = 1\}$$

mit dem Lebesgue-Maß $m^2(D) = 0$. Dies erklärt, warum die Verteilung $P^{\mathbf{X}}$ keine Dichte bezüglich des Lebesgue-Maßes m^2 hat, wie schon oben festgestellt wurde.

Man nennt eine Verteilungsfunktion G über dem Einheitsquadrat bzw. allgemeiner über dem d -dimensionalen Einheitswürfel im \mathbb{R}^d , die Randverteilungsfunktionen besitzt, die zu einer stetigen Gleichverteilung über E gehören, auch eine *Copula*. Das folgende berühmte Ergebnis von Sklar (das wir hier nur ohne Beweis zitieren, siehe etwa NELSEN (1999)) zeigt, dass sich alle Verteilun-

II. Grundprinzipien stochastischer Modellierung

gen über \mathbb{R}^d in natürlicher Weise durch Randverteilungen und eine Copula beschreiben lassen, d.h. alle denkbaren Arten von stochastischer Abhängigkeit stecken bereits im Begriff der Copula.

Satz 31 (Sklar 1958). Es sei F eine beliebige Verteilungsfunktion auf \mathbb{R}^d mit $d \geq 2$. Dann existiert eine Copula C so, dass

$$F(x_1, \dots, x_d) = C(F_1(x_1), \dots, F_d(x_d)) \text{ für alle } \mathbf{x} \in \mathbb{R}^d$$

gilt, wobei F_1, \dots, F_d die zu F gehörigen Randverteilungsfunktionen bezeichnen. Die Copula ist dabei eindeutig bestimmt, wenn die Randverteilungsfunktionen F_1, \dots, F_d stetig sind.

Das letzte Beispiel zeigt, dass die Abbildung

$$\underline{C}_2(u_1, u_2) := \max\{u_1 + u_2 - 1, 0\} \text{ für } \mathbf{u} \in Q$$

eine solche Copula ist. Sie ist in gewisser Weise extrem, denn sie ist eine untere Schranke für *alle* zweidimensionalen Copulas (so genannte *untere Fréchet-Hoeffding-Schranke*). Die entsprechende obere Fréchet-Hoeffding-Schranke ist gegeben durch

$$\bar{C}_2(u_1, u_2) := \min\{u_1, u_2\} \text{ für } \mathbf{u} \in Q,$$

sie ist die Verteilungsfunktion des Zufallsvektors $\mathbf{X} = (X, X)$. Die Unabhängigkeitscopula ergibt sich als „intermediärer“ Fall zu

$$\dot{C}_2(u_1, u_2) = u_1 \cdot u_2 \text{ für } \mathbf{u} \in Q$$

(vgl. Satz 30). Die beiden letzten Copulas können leicht auf mehrere Dimensionen verallgemeinert werden, nämlich zu

$$\bar{C}_d(u_1, \dots, u_d) := \min\{u_1, \dots, u_d\}, \quad \dot{C}_d(u_1, \dots, u_d) = \prod_{i=1}^d u_i \text{ für } \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d, \quad d \in \mathbb{N}.$$

Die analoge untere Fréchet-Hoeffding-Schranke

$$\underline{C}_d(u_1, \dots, u_d) := \max\left\{\sum_{i=1}^d u_i - d + 1, 0\right\} \text{ für } \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d, \quad d \in \mathbb{N}$$

ist allerdings für $d \geq 3$ keine Copula mehr, jedoch bleibt für alle d -dimensionalen Copulas C die Ungleichungskette

$$\underline{C}_d(u_1, \dots, u_d) \leq C(u_1, \dots, u_d) \leq \bar{C}_d(u_1, \dots, u_d) \text{ für } \mathbf{u} = (u_1, \dots, u_d) \in [0, 1]^d, \quad d \in \mathbb{N}$$

bestehen. Copulas spielen aktuell eine große Rolle in der Versicherungs- und Finanzmathematik, etwa um im Rahmen eines mathematischen Risikomanagements die verschiedenen beobachtbaren Risiken und ihre gegenseitigen Abhängigkeiten modellhaft angemessen zu erfassen und zu bewerten. Solche Bewertungen werden seit einiger Zeit verstärkt durch aufsichtsrechtliche Vorgaben auf nationaler und europäischer Ebene vorgeschrieben.

II. Grundprinzipien stochastischer Modellierung

Verteilungsfunktionen können recht merkwürdige Eigenschaften besitzen. Wir zeigen das an dem folgenden eindimensionalen Beispiel.

Dazu betrachten wir zunächst einen diskreten Zufallsvektor (X, Y) mit Werten in $\mathbb{N} \times \mathbb{N}$. Für Y nehmen wir eine spezielle *geometrische Verteilung* an (diese wird noch genauer in Abschnitt II.5 behandelt), konkret:

$$P(Y = n) = \frac{1}{2^n} \text{ für alle } n \in \mathbb{N}.$$

Die Verteilung von X spezifizieren wir über bedingte Wahrscheinlichkeiten:

$$P(X = k | Y = n) = \frac{1}{n} \text{ für alle } k \in \{1, \dots, n\} \text{ bei gegebenem } n \in \mathbb{N}.$$

Die bedingte Verteilung von X unter dem Ereignis $\{Y = n\}$ ist also eine Laplace-Verteilung. Die unbedingte Verteilung von X ist demnach gegeben durch

$$P(X = k) = \sum_{n=1}^{\infty} P(X = k | Y = n) \cdot P(Y = n) = \sum_{n=k}^{\infty} \frac{1}{n2^n} \text{ für alle } k \in \mathbb{N}.$$

Dieser Ausdruck ist nicht weiter elementar zu vereinfachen ausser für $k = 1$ mit dem Wert

$$P(X = 1) = \sum_{n=1}^{\infty} \frac{1}{n2^n} = -\ln\left(1 - \frac{1}{2}\right) = \ln 2 = 0,6314\dots$$

unter Ausnutzung der Reihendarstellung $-\ln(1-x) = \sum_{n=1}^{\infty} \frac{x^n}{n}$ für $-1 \leq x < 1$. Wir betrachten jetzt

die reelle Zufallsvariable $Z := \frac{X}{Y}$, die Werte in $\mathbb{Q}^* = \mathbb{Q} \cap (0, 1]$ annimmt und sogar surjektiv ist.

Wir bestimmen die zugehörige Verteilungsfunktion F_Z wie folgt:

$$\begin{aligned} F_Z(x) &= P(Z \leq x) = P(X \leq xY) = \sum_{n=1}^{\infty} P(X \leq nx | Y = n) \cdot P(Y = n) \\ &= \sum_{n=1}^{\infty} P(X \leq \lfloor nx \rfloor | Y = n) \cdot P(Y = n) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{n2^n} \end{aligned}$$

für $0 \leq x \leq 1$ mit $F_Z(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 1. \end{cases}$ Dabei bezeichne $\lfloor z \rfloor := \max \{m \in \mathbb{Z} | m \leq z\}$ die größte ganze

Zahl unterhalb von $z \in \mathbb{R}$.

Diese Funktion ist in *jedem* Punkt $x \in \mathbb{Q}^*$ *unstetig* und in *jedem irrationalen* Punkt $x \in (0, 1)$ *stetig!!* Um dies einzusehen, betrachten wir die Elementarwahrscheinlichkeiten zu Z , also $P(Z = q)$ für alle $q \in \mathbb{Q}^*$. Jedes solche q kann in eindeutiger Weise als *teilerfremder* Bruch (nach Kürzen aller gemeinsamen Faktoren in der Primfaktorzerlegung) in der Form $q = \frac{a}{b}$ geschrieben werden mit $a, b \in \mathbb{N}$, $a \leq b$.

II. Grundprinzipien stochastischer Modellierung

Es folgt dann für $q \in \mathbb{Q}^*$ in dieser Darstellung, wieder mit der obigen Reihendarstellung für den Logarithmus:

$$\begin{aligned} P(Z = q) &= P\left(\frac{X}{Y} = \frac{a}{b}\right) = P\left(\bigcup_{n=1}^{\infty} \left\{\frac{X}{Y} = \frac{an}{bn}\right\}\right) = \sum_{n=1}^{\infty} P(\{X = an\} \cap \{Y = bn\}) \\ &= \sum_{n=1}^{\infty} P(X = an | Y = bn) \cdot P(Y = bn) = \sum_{n=1}^{\infty} \frac{1}{bn} \cdot \frac{1}{2^{bn}} = -\frac{1}{b} \ln\left(1 - \frac{1}{2^b}\right) > 0. \end{aligned}$$

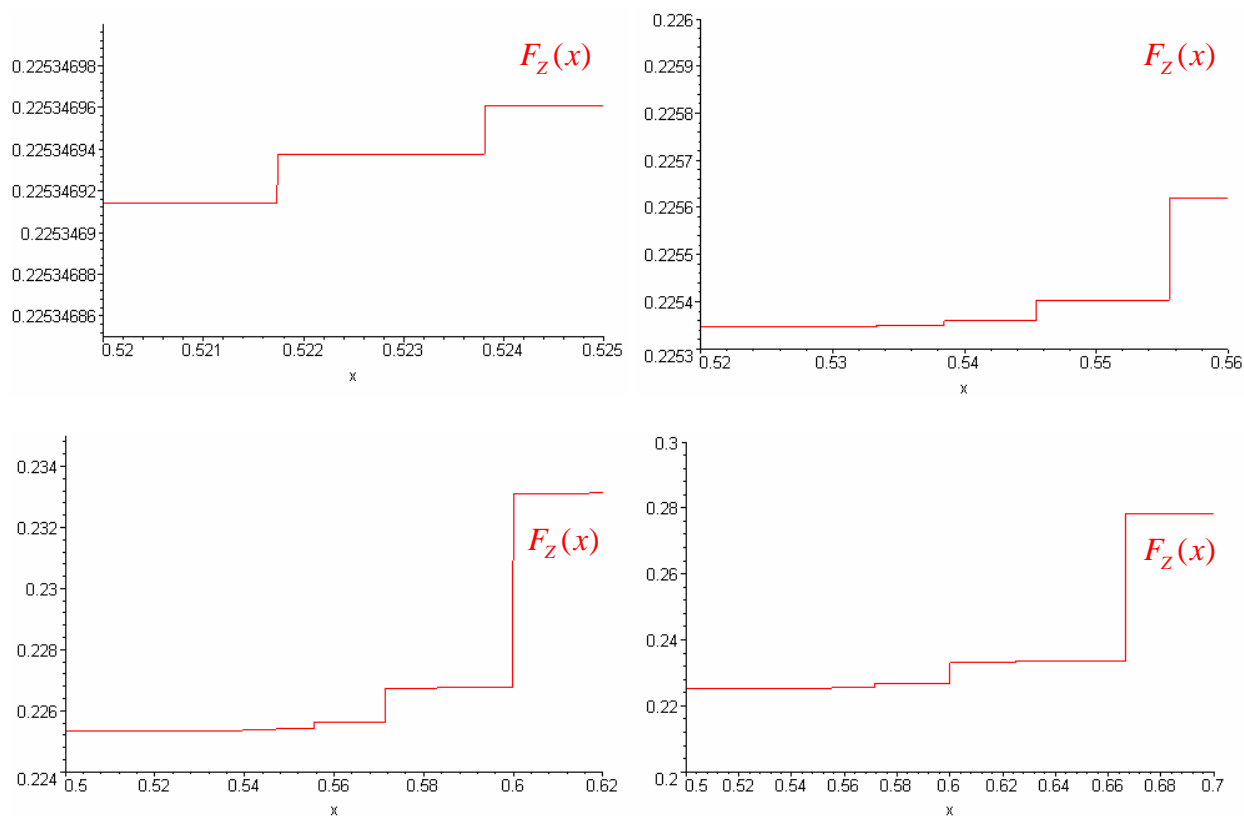
Dies bedeutet, dass $P(Z = q) > 0$ ist für *alle* $q \in \mathbb{Q}^*$, so dass folgt:

$$P(Z = q) = P(Z \leq q) - P(Z < q) = F_Z(q) - \lim_{h \downarrow 0} F_Z(q - h) > 0 \text{ für alle } q \in \mathbb{Q}^*,$$

d.h. F_Z ist in allen $q \in \mathbb{Q}^*$ nicht linksseitig stetig und damit dort unstetig, wie behauptet. Diese Eigenschaft gilt – wie man sofort sieht – übrigens allgemein immer dann, wenn die Zufallsvariable, deren Verteilung einer Verteilungsfunktion zu Grunde liegt, eine positive Wahrscheinlichkeit in einem einzelnen Punkt besitzt. Mit analoger Argumentation folgt auch, dass F_Z in allen *irrationalen* Punkten $x \in (0,1)$ stetig ist, weil offensichtlich $P^Z(\mathbb{Q}^*) = 1$ und damit $P^Z((0,1) \setminus \mathbb{Q}^*) = 0$ ist, also insbesondere $P(Z = x) = 0$ ist für alle $x \in (0,1) \setminus \mathbb{Q}^*$, was natürlich sofort zu

$$F_Z(x) = P(Z \leq x) = P(Z < x) \text{ für alle } x \in (0,1) \setminus \mathbb{Q}^*$$

führt. Die folgenden Bilder zeigen einige Ausschnittvergrößerungen aus dem Graphen von F_Z . Die unendlich dichte Lage der Unstetigkeitsstellen ist allerdings nicht abbildbar.



II. Grundprinzipien stochastischer Modellierung

Zur Veranschaulichung der „praktischen Nützlichkeit“ von Verteilungsfunktionen behandeln wir jetzt das so genannte *Achensee-Paradox*, das seinen Namen von einer mathematischen Fachtagung erhalten hat, die vor einigen Jahren am Achensee (Tirol) stattgefunden hat und auf der das Paradox intensiver diskutiert wurde. Es stammt aber wohl ursprünglich von T. Cover („Pick the largest number“, in COVER AND GOPINATH (1987), S. 152).

Achensee-Paradox: Jemand notiert zwei beliebige *verschiedene* reelle Zahlen auf zwei Zetteln (je eine pro Zettel), die verdeckt auf den Tisch gelegt werden. Eine zweite Person deckt nach Wahl zufällig einen der Zettel auf. Dann gibt es eine Strategie, mit der mit *mehr* als 50%iger Wahrscheinlichkeit entschieden werden kann, ob die noch verdeckte Zahl größer oder kleiner als die aufgedeckte Zahl ist.

Dies erscheint einem unbedarften Betrachter als unmöglich, da ja keinerlei Information über die noch verdeckte Zahl vorliegt. Daher würde ein solcher Betrachter davon ausgehen, dass die Wahrscheinlichkeit einer korrekten Vorhersage über die Größenordnung der zweiten Zahl nicht anders sein kann als bei einem reinen Raten, wo diese Wahrscheinlichkeit genau 50% und *nicht mehr* beträgt.

Die versprochene Strategie besteht nun darin, dass man unabhängig von der Größe der aufgedeckten Zahl ein Zufallsexperiment durchführt, in dem man eine Zufallszahl Y erzeugt, die aus einer Wahrscheinlichkeitsverteilung Q stammt, deren zugehörige Verteilungsfunktion F lediglich *stetig* und *streng monoton wachsend* auf \mathbb{R} sei. Solche Verteilungen und die Erzeugung passender Zufallszahlen werden wir später in diesem Abschnitt bzw. genauer im übernächsten Abschnitt kennen lernen.

Die Entscheidungsregel zu der Strategie lautet nun wie folgt: Ist die Zufallszahl Y *größer* als die aufgedeckte Zahl, so entscheiden wir, dass die noch verdeckte Zahl ebenfalls größer als die aufgedeckte Zahl ist. Ist die Zufallszahl Y *kleiner* als die aufgedeckte Zahl, so entscheiden wir, dass die noch verdeckte Zahl ebenfalls kleiner als die aufgedeckte Zahl ist. Mit anderen Worten: wir besorgen uns ersatzweise die zweite Zahl aus einem eigenen unabhängigen Zufallsexperiment und entscheiden dann so, als wenn die „gezogene“ Zahl tatsächlich die noch verdeckte wäre.

Diese Strategie führt tatsächlich mit mehr als 50%iger Wahrscheinlichkeit zum Erfolg! Dies kann man wie folgt einsehen: wir bezeichnen die kleinere der notierten Zahlen mit a , die größere mit b . Das Aufdecken der ersten Zahl kann beschrieben werden durch eine Zufallsvariable X mit der Wertemenge $\mathfrak{X} = \{a, b\}$ und den Wahrscheinlichkeiten

$$P(X = a) = P(X = b) = \frac{1}{2}.$$

Die genaue Struktur des zu Grunde liegenden Wahrscheinlichkeitsraumes (Ω, \mathcal{A}, P) spielt dabei keine Rolle. Die obige Ersatz-Zufallsvariable Y werde voraussetzungsgemäß als von X unabhängig angenommen. Dann kann zunächst das Ereignis einer korrekten Vorhersage beschrieben werden als die Menge

$$A := (\{X = a\} \cap \{Y > a\}) \oplus (\{X = b\} \cap \{Y < b\}).$$

Es folgt

II. Grundprinzipien stochastischer Modellierung

$$\begin{aligned}
 P(A) &= P(\{X = a\} \cap \{Y > a\}) + P(\{X = b\} \cap \{Y < b\}) \\
 &= P(X = a) \cdot P(Y > a) + P(X = b) \cdot P(Y < b) \\
 &= \frac{1}{2} \cdot (1 - F(a)) + \frac{1}{2} \cdot F(b) = \frac{1}{2} + \frac{F(b) - F(a)}{2} > \frac{1}{2}!
 \end{aligned}$$

Man beachte hierbei, dass $P(Y < b) = P(Y \leq b) = F(b)$ ist wegen

$$\begin{aligned}
 P(Y = b) &= Q(\{b\}) = \lim_{h \downarrow 0} Q((b-h, b]) = \lim_{h \downarrow 0} (F(b) - F(b-h)) \\
 &= F(b) - \lim_{h \downarrow 0} F(b-h) = F(b) - F(b) = 0
 \end{aligned}$$

auf Grund der vorausgesetzten Stetigkeit von F .

Leider kann i.Allg. der Trefferwahrscheinlichkeits-Zuwachs $\frac{F(b) - F(a)}{2}$ nicht konkret bestimmt werden, da die Zahlen a und b ja nicht bekannt sind. Man kann das allerdings dann, wenn bekannt ist, nach welchem Mechanismus die Zahlen a und b ermittelt werden, etwa dann, wenn diese selbst das Ergebnis einer zufälligen „Ziehung“ sind nach dem Muster der obigen Strategie. Wir werden später noch zeigen, dass die Trefferwahrscheinlichkeit sogar $\frac{2}{3}$ (!) beträgt, wenn alle Zahlen unabhängig voneinander nach derselben Verteilung Q mit der stetigen, streng monoton wachsenden Verteilungsfunktion F gezogen werden.

Für die praktische Durchführung des Spiels kann man sich z.B. vorab eine Liste mit Zufallszahlen nach der selbst gewählten Verteilung Q ausdrucken und muss dann nur noch in jeder Runde die aufgedeckte Zahl mit der entsprechenden Zahl der eigenen Liste vergleichen.

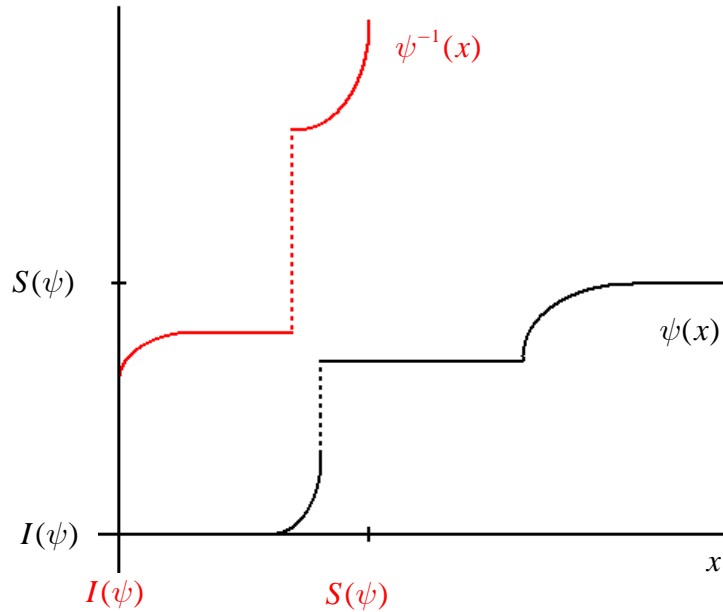
Bei der oben angesprochenen Erzeugung von Zufallszahlen spielt die so genannte *Pseudo-Inverse* einer Verteilungsfunktion eine wichtige Rolle.

Definition 31 (Pseudo-Inverse). Es sei $\psi: \mathbb{R} \rightarrow \mathbb{R}$ eine schwach monoton wachsende, rechtsseitig stetige Funktion. Ferner sei $I(\psi) := \inf \{\psi(x) \mid x \in \mathbb{R}\}$, $S(\psi) := \sup \{\psi(x) \mid x \in \mathbb{R}\}$. Dann ist auf dem offenen Intervall $(I(\psi), S(\psi))$ die *Pseudo-Inverse* ψ^{-1} von ψ definiert durch

$$\psi^{-1}(y) := \inf \{x \in \mathbb{R} \mid \psi(x) \geq y\}, \quad y \in (I(\psi), S(\psi)).$$

Bemerkung: Bei der Bildung der Pseudo-Inversen gehen Konstanzbereiche von ψ in Sprungstellen von ψ^{-1} über und umgekehrt, wie die folgende Skizze zeigt:

II. Grundprinzipien stochastischer Modellierung



Lemma 31 (Eigenschaften der Pseudo-Inversen). Es seien ψ und ψ^{-1} wie in Definition 31. Dann gilt:

- ψ^{-1} ist auf $(I(\psi), S(\psi))$ schwach monoton wachsend und linksseitig stetig.
- $\psi(\psi^{-1}(y)) \geq y$ für alle $y \in (I(\psi), S(\psi))$.
- Ist ψ in $\psi^{-1}(y)$ stetig für ein $y \in (I(\psi), S(\psi))$, so gilt $\psi(\psi^{-1}(y)) = y$.
- $\psi^{-1}(\psi(x)) \leq x$ für alle $x \in \mathbb{R}$ mit $\psi(x) \in (I(\psi), S(\psi))$.
- Ist ψ^{-1} in $\psi(x)$ stetig für ein $x \in \mathbb{R}$ mit $\psi(x) \in (I(\psi), S(\psi))$, so gilt $\psi^{-1}(\psi(x)) = x$.

Beweis: Es seien $y, z \in (I(\psi), S(\psi))$ mit $y \leq z$ und $A_y := \{x \in \mathbb{R} \mid \psi(x) \geq y\}$, $A_z := \{x \in \mathbb{R} \mid \psi(x) \geq z\}$. Dann ist $A_z \subseteq A_y$ und somit $\psi^{-1}(y) = \inf(A_y) \leq \inf(A_z) = \psi^{-1}(z)$. Dies ist der erste Teil von a). Sei nun $y \in (I(\psi), S(\psi))$. Definitionsgemäß existiert eine schwach monoton fallende Folge $\{x_n\}_{n \in \mathbb{N}}$ mit $\psi(x_n) \geq y$ und $\lim_{n \rightarrow \infty} x_n = \psi^{-1}(y)$. Da ψ rechtsseitig stetig ist, folgt $\psi(\psi^{-1}(y)) = \lim_{n \rightarrow \infty} \psi(x_n) \geq y$. Dies ist Teil b). Für ein $x \in \mathbb{R}$ mit $\psi(x) \in (I(\psi), S(\psi))$ gilt definitionsgemäß $\psi^{-1}(\psi(x)) = \inf\{z \in \mathbb{R} \mid \psi(z) \geq \psi(x)\} \leq x$. Dies ist Teil d). Sei nun $\{y_n\}_{n \in \mathbb{N}}$ eine schwach monoton wachsende Folge in $(I(\psi), S(\psi))$ mit $y = \lim_{n \rightarrow \infty} y_n$. Nach dem gerade Gezeigten ist dann auch die Folge $\{\psi^{-1}(y_n)\}_{n \in \mathbb{N}}$ schwach monoton wachsend mit $\psi^{-1}(y_n) \leq \psi^{-1}(y) < \infty$, d.h. die Folge $\{\psi^{-1}(y_n)\}_{n \in \mathbb{N}}$ ist (schwach monoton) konvergent, etwa gegen $z \leq \psi^{-1}(y)$. Angenommen, es sei $z < \psi^{-1}(y)$. Dann ist $\psi(z) < y$ nach Definition von ψ^{-1} . Nach Teil b) ist aber $\psi(\psi^{-1}(y_n)) \geq y_n$, d.h. es ist wegen $\psi^{-1}(y_n) \leq z$ und der schwachen Monotonie von ψ : $\psi(z) \geq \liminf_{n \rightarrow \infty} \psi(\psi^{-1}(y_n)) \geq y > \psi(z)$. \Leftarrow , also ist $z = \psi^{-1}(y)$. Das ist der zweite Teil von a). Wir zeigen noch die Teile c) und e). Sei also ψ in $\psi^{-1}(y)$ stetig für ein $y \in (I(\psi), S(\psi))$. Ist nun

II. Grundprinzipien stochastischer Modellierung

$\{x_n\}_{n \in \mathbb{N}}$ eine schwach monoton wachsende Folge mit $x_n < \psi^{-1}(y)$ und $\lim_{n \rightarrow \infty} x_n = \psi^{-1}(y)$, so gilt nach Definition der Pseudo-Inversen $\psi(x_n) < y$ und somit nach Teil b): $\lim_{n \rightarrow \infty} \psi(x_n) \leq y \leq \psi(\psi^{-1}(y)) = \psi(\lim_{n \rightarrow \infty} x_n) = \lim_{n \rightarrow \infty} \psi(x_n)$, also Gleichheit. Dies ist Teil c). Sei schließlich ψ^{-1} in $\psi(x)$ stetig für ein $x \in \mathbb{R}$ mit $\psi(x) \in (I(\psi), S(\psi))$. Ist $\{y_n\}_{n \in \mathbb{N}}$ eine schwach monoton fallende Folge mit $y_n > \psi(x)$ und $\lim_{n \rightarrow \infty} y_n = \psi(x)$, so gilt wieder nach Definition der Pseudo-Inversen $\psi^{-1}(y_n) > x$ und somit nach Teil d): $\lim_{n \rightarrow \infty} \psi^{-1}(y_n) \geq x \geq \psi^{-1}(\psi(x)) = \psi^{-1}(\lim_{n \rightarrow \infty} y_n) = \lim_{n \rightarrow \infty} \psi^{-1}(y_n)$, also auch hier die Gleichheit. Dies ist Teil e). Damit ist das Lemma vollständig bewiesen. ■

Lemma 32. Es sei X eine Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit stetiger Verteilungsfunktion F . Dann ist die Zufallsvariable $Y := F(X)$ über dem Einheitsintervall $[0, 1]$ stetig gleichverteilt. Ist umgekehrt F eine beliebige Verteilungsfunktion und Y eine über dem Einheitsintervall $[0, 1]$ stetig gleichverteilte Zufallsvariable, so besitzt die Zufallsvariable $X := F^{-1}(Y)$ die Verteilungsfunktion F .

Beweis: Erster Teil: Zunächst ist bei Verteilungsfunktionen $I(F) = 0$, $S(F) = 1$. Für $0 < x < 1$ sei nun $A_x := \{y \in \mathbb{R} \mid F(y) \leq x\}$, $B_x := \{y \in \mathbb{R} \mid y \leq F^{-1}(x)\}$. Wenn F stetig ist, folgt aus der schwachen Monotonie von F und Teil c) von Lemma 31: $B_x \subseteq A_x$ für alle $0 < x < 1$. Ist ferner $x \in (0, 1)$ ein Stetigkeitspunkt von F^{-1} , so folgt umgekehrt auch $A_x \subseteq B_x$ aus den Teilen a) und e) von Lemma 31. Wegen der schwachen Monotonie von F^{-1} ist die Menge der Unstetigkeitspunkte im Intervall $(0, 1)$, $U(F^{-1})$, höchstens abzählbar, d.h. für stetiges F ist die Menge $U(F^{-1}) = \{x \in (0, 1) \mid A_x \neq B_x\}$ eine Lebesgue-Nullmenge. Für $x \in (0, 1) \setminus U(F^{-1})$ folgt dann also

$$\{F(X) \leq x\} = X^{-1}(A_x) = X^{-1}(B_x) = \{X \leq F^{-1}(x)\}$$

und somit

$$P(F(X) \leq x) = P(X \leq F^{-1}(x)) = F(F^{-1}(x)) = x \text{ für alle } x \in (0, 1),$$

d.h. $F(X)$ ist über $[0, 1]$ stetig gleichverteilt (die Menge $\{0, 1\}$ ist ebenfalls eine Lebesgue-Nullmenge).

Zweiter Teil: Aus den Teilen b) und d) von Lemma 31 folgt wegen der schwachen Monotonie von F und F^{-1} sofort

$$\{F^{-1}(Y) \leq x\} = \{Y \leq F(x)\} \text{ für alle } x \in (0, 1),$$

also gilt

$$P(X \leq x) = P(F^{-1}(Y) \leq x) = P(Y \leq F(x)) = F(x) \text{ für alle } x \in (0, 1),$$

so dass X die Verteilungsfunktion F besitzt. Damit ist das Lemma bewiesen. ■

II. Grundprinzipien stochastischer Modellierung

Lemma 32 zeigt damit auf, wie man bei dem Achensee-Paradox vorgehen kann. Eine geeignete Verteilungsfunktion F erhält man beispielsweise, wenn man die Dichte

$$f(x) = \frac{1}{\pi} \frac{1}{1+x^2}, \quad x \in \mathbb{R}$$

betrachtet. Man erhält damit die so genannte *Cauchy-Verteilung*, deren Verteilungsfunktion F gegeben ist durch

$$F(x) = \int_{-\infty}^x f(u) du = \frac{1}{\pi} \int_{-\infty}^x \frac{1}{1+u^2} du = \frac{1}{\pi} \arctan u \Big|_{-\infty}^x = \frac{1}{\pi} \left(\arctan x + \frac{\pi}{2} \right), \quad x \in \mathbb{R}.$$

Wenn man also auf einfache Weise eine Zufallsvariable U erzeugen kann, die über dem Einheitsintervall $[0,1]$ stetig gleichverteilt sind, so ist die Zufallsvariable

$$X := F^{-1}(U) = \tan \left(\pi U - \frac{\pi}{2} \right) = -\cot(\pi U)$$

Cauchy-verteilt; ihre „Realisationen“ können damit für die Strategie-Liste beim Achensee-Paradox herangezogen werden.

Realisationen solcher Zufallsvariablen U werden häufig näherungsweise durch rekursive Algorithmen folgender Art erzeugt:

1. Wähle einen so genannten Modul $m \in \mathbb{N}$ und einen Faktor $a \in \mathbb{N}$, m, a geeignet und genügend groß.
2. Wähle einen Startwert $z_0 \in \{1, 2, \dots, m\}$.
3. Setze rekursiv

$$z_n := a \cdot z_{n-1} \bmod m \quad \text{für } n \in \mathbb{N}.$$

4. Setze $U_n := \frac{z_n}{m}$, $n \in \mathbb{Z}^+$.

Eine „geeignete“ Bestimmung der Parameter m und a erfordert tiefsinnige Überlegungen, die wesentlich mit Methoden der algebraischen Zahlentheorie arbeiten und daher hier nicht dargestellt werden können. Gebräuchlich sind etwa die Wahlen

$$m = 2^t \quad \text{mit } t \geq 3 \quad \text{und} \quad a \equiv 3 \pmod{8} \quad \text{oder} \quad a \equiv 5 \pmod{8}.$$

Man muss jedoch aufpassen, ob die so erzeugten Zahlenfolgen auch genügend „zufällig“ (im Sinne von stochastisch unabhängig) und „gleichmäßig verteilt“ sind, insbesondere wenn man aus den Zahlen Paare, Tripel oder höherdimensionale Zufallsvektoren bilden will.

Ein anderes Problem besteht darin, dass alle solchen rekursiven Verfahren eine Periode besitzen, d.h. nach einer gewissen Anzahl von Zahlen wiederholen sich diese regelmäßig. Die Periodenlänge des obigen Verfahrens beträgt etwa $2^{t-2} = \frac{m}{4}$.

Die meisten Programmiersprachen enthalten bereits eingebaute Befehle zur Erzeugung solcher Zufallszahlen, oft in der Form des RND-Befehls oder ähnlich.

II. Grundprinzipien stochastischer Modellierung

Mit Hilfe von Lemma 32 können wir nun auch die noch offene Rechnung zum Achensee-Paradox nachliefern. Wir nehmen an, dass der erste Spieler seine beiden Zahlen Z_1 und Z_2 unabhängig voneinander nach einer Verteilung mit stetiger und streng monotoner Verteilungsfunktion F erzeugt. Dann sind beide Zahlen fast sicher voneinander verschieden¹⁷. Der zweite Spieler ziehe nun seine „Ersatzzahl“ Z_3 auf dieselbe Weise, unabhängig vom ersten Spieler. Dann kann das Ereignis einer korrekten Vorhersage analog zu oben beschrieben werden als die Menge

$$\begin{aligned} A &:= (\{Z_1 < Z_2\} \cap \{Z_1 < Z_3\}) \oplus (\{Z_1 > Z_2\} \cap \{Z_1 > Z_3\}) \\ &= (\{F(Z_1) < F(Z_2)\} \cap \{F(Z_1) < F(Z_3)\}) \oplus (\{F(Z_1) > F(Z_2)\} \cap \{F(Z_1) > F(Z_3)\}), \end{aligned}$$

wobei jetzt nach Lemma 32 die Zufallsvariablen $U_i := F(Z_i)$, $i = 1, 2, 3$ stetig gleichverteilt und nach Lemma 28 ebenfalls stochastisch unabhängig sind. Es folgt

$$\begin{aligned} P(A) &= P(\{U_1 < U_2\} \cap \{U_1 < U_3\}) + P(\{U_1 > U_2\} \cap \{U_1 > U_3\}) \\ &= \int_0^1 \int_{u_1}^1 \int_{u_1}^1 du_2 du_3 du_1 + \int_0^1 \int_0^{u_1} \int_0^{u_1} du_2 du_3 du_1 = \int_0^1 \int_{u_1}^1 (1-u_1) du_3 du_1 + \int_0^1 \int_0^{u_1} u_1 du_3 du_1 \\ &= \int_0^1 (1-u_1)^2 du_1 + \int_0^1 u_1^2 du_1 = -\frac{(1-u_1)^3}{3} \Big|_0^1 + \frac{u_1^3}{3} \Big|_0^1 = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}, \end{aligned}$$

wie behauptet.

Der Rest dieses Abschnitts ist hauptsächlich den so genannten *Momenten* von Zufallsvariablen bzw. deren Verteilungen gewidmet. Eine herausragende Rolle spielt dabei der Begriff des *Erwartungswerts* einer Zufallsvariablen; er ist wesentlich für das wichtige *Gesetz der großen Zahlen*, das in einer einfachen Form bereits vor dem Jahr 1700 von Jakob Bernoulli (1654 – 1705) in seiner berühmten *Ars Conjectandi* (posthum im Jahre 1713 erschienen) formuliert wurde.

Definition 32 (Erwartungswert). Es sei X eine reellwertige Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Im Falle der Existenz heißt

$$E(X) := \int X dP$$

der *Erwartungswert* von X .

Bemerkung: Der Erwartungswert $E(X)$ existiert also genau dann, wenn X P -integrierbar ist, vgl. Definition 20. Insbesondere existiert $E(X)$ immer, wenn $X \geq 0$ oder $X \leq 0$ ist, ggf. mit dem Wert $+\infty$ oder $-\infty$. Wir unterscheiden sinngemäß auch hier zwischen *eigentlichem* und *uneigentlichem* Erwartungswert.

¹⁷ Man kann zeigen, dass in diesem Fall $P(Z_1 = Z_2) = P(Z_1 - Z_2 = 0) = 0$ ist.

II. Grundprinzipien stochastischer Modellierung

Eine Reihe von Eigenschaften des Erwartungswerts kann man sofort aus Lemma 20 sowie aus den Sätzen 18 und 19 ablesen. Wir listen diese hier nur noch einmal auf, der Beweis wurde ja jeweils schon in allgemeiner Form gegeben.

Lemma 33 (Eigenschaften des Erwartungswerts). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, X, Y und $\{X_n\}_{n \in \mathbb{N}}$ seien reellwertige Zufallsvariablen. Dann gilt:

- a) Ist $X = \mathbb{1}_A$ für ein $A \in \mathcal{A}$, so ist $E(X) = P(A)$.
- b) $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$ für alle $\alpha, \beta \in \mathbb{R}$ (im Falle der Existenz)
- c) $X \leq Y \Rightarrow E(X) \leq E(Y)$
- d) $|E(X)| \leq E(|X|)$
- e) $\forall A \in \mathcal{A}: P(A) = 0 \Rightarrow \int_A X dP = E(\mathbb{1}_A \cdot X) = 0$
- f) $X \geq 0$ und $E(X) = 0 \Rightarrow X = 0$ P -fast sicher
- g) $X = Y$ P -fast sicher $\Rightarrow E(X) = E(Y)$
- h) $\forall A \in \mathcal{A}: E(\mathbb{1}_A \cdot X) \leq E(\mathbb{1}_A \cdot Y) \Rightarrow X \leq Y$ P -fast sicher
- i) $\forall A \in \mathcal{A}: E(\mathbb{1}_A \cdot X) = E(\mathbb{1}_A \cdot Y) \Rightarrow X = Y$ P -fast sicher
- j) $\forall \varepsilon > 0: P(|X| > \varepsilon) \leq \frac{E(|X|)}{\varepsilon}$ (Markoff-Ungleichung).
- k) Ist die Folge $\{X_n\}_{n \in \mathbb{N}}$ nicht-negativ und schwach monoton wachsend, so gilt $\lim_{n \rightarrow \infty} E(X_n) = E\left(\lim_{n \rightarrow \infty} X_n\right)$.
- l) Ist die Folge $\{X_n\}_{n \in \mathbb{N}}$ punktweise konvergent gegen X und gilt $|X_n| \leq Y$ für alle $n \in \mathbb{N}$ mit $E(Y) < \infty$, so gilt ebenfalls $\lim_{n \rightarrow \infty} E(X_n) = E\left(\lim_{n \rightarrow \infty} X_n\right) = E(X)$ sowie $\lim_{n \rightarrow \infty} E(|X_n - X|) = 0$.

Satz 32 (Berechnung von Erwartungswerten). Es seien X und Y reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

- a) Ist X *diskret* verteilt, d.h. gibt es eine höchstens abzählbare Menge $T \subset \mathbb{R}$ mit $P^X(T) = 1$ ¹⁸ und $P(X = x) > 0$ für alle $x \in T$, so ist X genau dann eigentlich P -integrierbar, wenn

$$\sum_{x \in T} |x| P(X = x) < \infty$$

gilt. In diesem Fall lässt sich der Erwartungswert darstellen als

$$E(X) = \sum_{x \in T} x \cdot P(X = x).$$

¹⁸ Man nennt dann T auch den *Träger* der Verteilung P^X .

II. Grundprinzipien stochastischer Modellierung

Ist des Weiteren P selbst eine *diskrete Verteilung*, d.h. gilt $\{\omega\} \in \mathcal{A}$ für alle $\omega \in \Omega$ und $P(\{\omega\}) > 0$ für höchstens abzählbar viele $\omega \in \Omega$, so gilt analog

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

- b) Besitzt P^X eine Dichte bezüglich des Lebesgue-Maßes, so ist X genau dann eigentlich P -integrierbar, wenn

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty$$

gilt. In diesem Fall lässt sich der Erwartungswert darstellen als

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

- c) Bezeichnet F die Verteilungsfunktion von X und ist X eigentlich P -integrierbar, so gilt

$$E(X) = - \int_{-\infty}^0 F(x) dx + \int_0^{\infty} (1 - F(x)) dx.$$

- d) Ist X eigentlich P -integrierbar, so lässt sich der Erwartungswert auch darstellen als

$$E(X) = \int_0^1 F^{-1}(x) dx,$$

wobei F^{-1} wieder die Pseudo-Inverse von F bezeichne.

Beweis:

- a) Es sei $T^+ = \{x_1, x_2, \dots\}$ der (abzählbare) Träger von P^{X^+} und $T^- = \{y_1, y_2, \dots\}$ der (abzählbare) Träger von P^{X^-} . Wir setzen

$$X_n^+ = \sum_{k=1}^n x_k \mathbb{1}_{\{X^+ = x_k\}} \quad \text{und} \quad X_n^- = \sum_{k=1}^n y_k \mathbb{1}_{\{X^- = y_k\}}, \quad n \in \mathbb{N},$$

d.h. auf den Mengen $\{X^+ \in \{x_1, \dots, x_n\}\}$ bzw. $\{X^- \in \{y_1, \dots, y_n\}\}$ stimmen X_n^+ und X^+ bzw. X_n^- und X^- überein; ferner gilt jeweils

$$X_n^+ \uparrow X^+, \quad X_n^- \uparrow X^- \quad \text{mit } n \rightarrow \infty \text{ sowie}$$

$$E(X_n^+) = \int X_n^+ dP = \sum_{k=1}^n x_k \int \mathbb{1}_{\{X^+ = x_k\}} dP = \sum_{k=1}^n x_k P(X^+ = x_k)$$

II. Grundprinzipien stochastischer Modellierung

$$E(X_n^-) = \int X_n^- dP = \sum_{k=1}^n y_k \int \mathbb{1}_{\{X_n^- = y_k\}} dP = \sum_{k=1}^n y_k P(X_n^- = k).$$

Existiert nun im eigentlichen Sinn der Erwartungswert $E(X)$, so konvergieren nach Voraussetzung beide Reihen mit

$$E(X^+) = \lim_{n \rightarrow \infty} E(X_n^+) = \sum_{x \geq 0} x \cdot P(X_n^+ = x), \quad E(X^-) = \lim_{n \rightarrow \infty} E(X_n^-) = -\sum_{x \leq 0} x \cdot P(X_n^- = -x)$$

Daher konvergiert auch $\sum_{x \in T} |x| P(X = x)$ mit $E(|X|) = E(X^+) + E(X^-) = \sum_{x \in T} |x| P(X = x)$.

Ist umgekehrt die letztere Reihe konvergent, so konvergieren jeweils auch die Reihen $\sum_{x \geq 0} x \cdot P(X = x)$ und $\sum_{x \leq 0} x \cdot P(X = x)$, d.h. es existieren jeweils $E(X^+) = \sum_{x \geq 0} x \cdot P(X = x)$ und $E(X^-) = -\sum_{x \leq 0} x \cdot P(X = x)$, und damit auch $E(X) = E(X^+) - E(X^-) = \sum_{x \in T} x \cdot P(X = x)$.

Ist nun P diskret, so lässt sich wegen der absoluten Konvergenz der Reihe der Erwartungswert $E(X) = \sum_{x \in T} x \cdot P(X = x)$ auch folgendermaßen umordnen:

$$\begin{aligned} E(X) &= \sum_{x \in T} x \cdot P(X = x) = \sum_{x \in T} x \cdot P(\{\omega \in \Omega \mid X(\omega) = x\}) = \sum_{x \in T} x \sum_{\omega: X(\omega)=x} P(\{\omega\}) \\ &= \sum_{x \in T} \sum_{\omega: X(\omega)=x} X(\omega) P(\{\omega\}) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\}), \end{aligned}$$

wie behauptet.

- b) Der erste Teil folgt aus Definition 20 und den nachfolgenden Überlegungen. Es gilt nach den Sätzen 20 und 22, wenn id die Identitätsabbildung bezeichnet,

$$\begin{aligned} E(X^+) &= \int X^+ dP = \int id^+ dP^X = \int id^+ \cdot f d\mathfrak{m}^1 = \int_{-\infty}^{\infty} x^+ \cdot f(x) dx \\ E(X^-) &= \int X^- dP = \int id^- dP^X = \int id^- \cdot f d\mathfrak{m}^1 = \int_{-\infty}^{\infty} x^- \cdot f(x) dx \\ E(|X|) &= E(X^+) + E(X^-) = \int_{-\infty}^{\infty} (x^+ + x^-) \cdot f(x) dx = \int_{-\infty}^{\infty} |x| \cdot f(x) dx \\ E(X) &= E(X^+) - E(X^-) = \int_{-\infty}^{\infty} (x^+ - x^-) \cdot f(x) dx = \int_{-\infty}^{\infty} x \cdot f(x) dx. \end{aligned}$$

- c) Mit Hilfe des Satzes von Fubini (Satz 23) lässt sich der Beweis wie folgt durchführen:

II. Grundprinzipien stochastischer Modellierung

$$\begin{aligned} \int_0^{\infty} (1-F(x)) dx &= \int_0^{\infty} P(X > x) dx = \int_0^{\infty} \int_x^{\infty} P^X(dy) m^1(dx) = \int_0^{\infty} \int_0^y m^1(dx) P^X(dy) \\ &= \int_0^{\infty} y P^X(dy) = \int_{\mathbb{R}} y^+ P^X(dy) = \int_{\mathbb{R}} id^+ dP^X = \int X^+ dP = E(X^+) \end{aligned}$$

und

$$\begin{aligned} \int_{-\infty}^0 F(x) dx &= \int_{-\infty}^0 P(X \leq x) dx = \int_{-\infty}^0 \int_{-\infty}^x P^X(dy) m^1(dx) = \int_{-\infty}^0 \int_y^0 m^1(dx) P^X(dy) \\ &= - \int_{-\infty}^0 y P^X(dy) = \int_{\mathbb{R}} y^- P^X(dy) = \int_{\mathbb{R}} id^- dP^X = \int X^- dP = E(X^-), \end{aligned}$$

also

$$- \int_{-\infty}^0 F(x) dx + \int_0^{\infty} (1-F(x)) dx = -E(X^-) + E(X^+) = E(X),$$

wie behauptet.

d) Sei zunächst $X \geq 0$. Dann ist auch $F^{-1} \geq 0$, und man erhält

$$\int_0^1 F^{-1}(x) dx = \int_0^1 \int_0^{F^{-1}(x)} dy dx = \iint_{\substack{0 \leq y \leq F^{-1}(x) \\ 0 < x < 1}} dy dx = \iint_{\substack{F(y) \leq x < 1 \\ y > 0}} dx dy = \int_0^{\infty} (1-F(y)) dy = E(X).$$

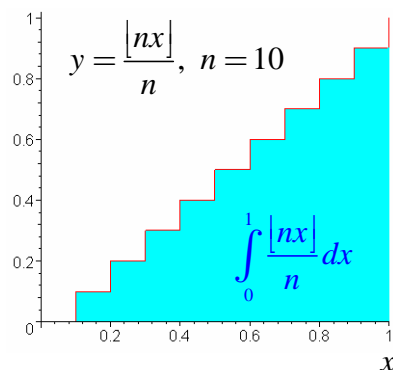
Im Fall $X \leq 0$ ergibt sich analog

$$\int_0^1 F^{-1}(x) dx = - \int_0^1 \int_{F^{-1}(x)}^0 dy dx = - \iint_{\substack{F^{-1}(x) \leq y \leq 0 \\ 0 < x < 1}} dy dx = - \iint_{\substack{0 < x \leq F(y) \\ y \leq 0}} dx dy = - \int_{-\infty}^0 F(y) dy = E(X).$$

Für beliebige X erhält man nun das Ergebnis durch die Zerlegung $X = X^+ - X^-$ mittels Teil c). Damit ist der Satz vollständig bewiesen. ■

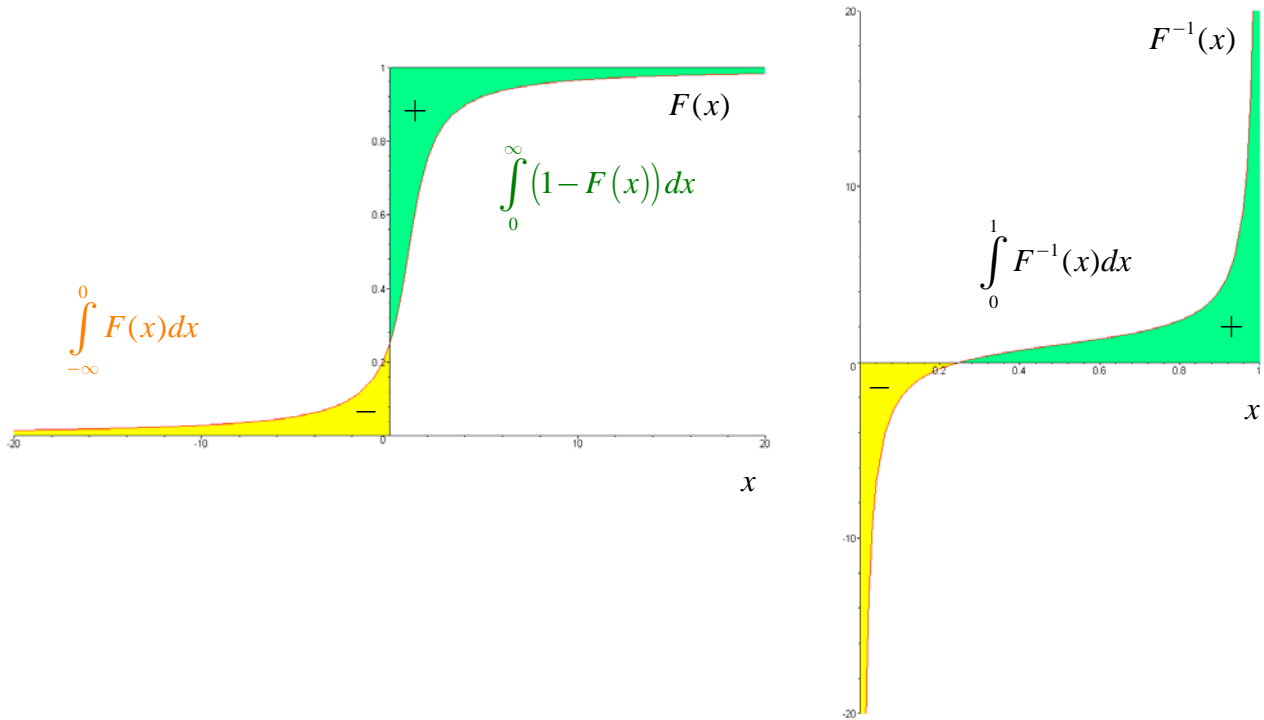
Für das auf S. 91 gebrachte Beispiel läßt sich der Erwartungswert $E(X)$ am besten mit Teil c) von Satz 32 berechnen:

$$\begin{aligned} E(X) &= \int_0^1 (1-F(x)) dx = 1 - \int_0^1 F(x) dx = 1 - \sum_{n=1}^{\infty} \frac{1}{2^n} \int_0^1 \frac{|nx|}{n} dx \\ &= 1 - \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \sum_{k=1}^n \frac{k-1}{n^2} = 1 - \sum_{n=1}^{\infty} \frac{n-1}{n2^{n+1}} = \frac{1+\ln 2}{2} \approx 0,8465... \end{aligned}$$



Die folgenden Graphiken zeigt den Zusammenhang zwischen den Teilen c) und d) von Satz 32:

II. Grundprinzipien stochastischer Modellierung



Bei Transformationen von Zufallsvariablen ist das folgende Ergebnis, das sich sofort aus den Sätzen 22 und 20 ergibt, oft recht nützlich.

Lemma 34. Es sei \mathbf{X} ein Zufallsvektor auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Werten in \mathbb{R}^d , $d \in \mathbb{N}$. Ferner sei $G: \mathbb{R}^d \rightarrow \mathbb{R}$ eine Borel-messbare Abbildung. Dann gilt: ist $G(\mathbf{X})$ P -integrierbar, so gilt

$$E(G(\mathbf{X})) = \int G(\mathbf{X}) dP = \int G dP^{\mathbf{X}}.$$

Besitzt darüber hinaus die Verteilung $P^{\mathbf{X}}$ eine Dichte f , so gilt auch

$$E(G(\mathbf{X})) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} G(x_1, \dots, x_d) \cdot f(x_1, \dots, x_d) dx_1 \cdots dx_d.$$

Als Anwendung dieses Lemmas ergibt sich sofort noch:

Lemma 35. Es seien X_1, \dots, X_d stochastisch unabhängige, reellwertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Sind dann alle X_i P -integrierbar, so auch $Y := \prod_{i=1}^d X_i$, und es gilt

$$E(Y) = E\left(\prod_{i=1}^d X_i\right) = \prod_{i=1}^d E(X_i).$$

II. Grundprinzipien stochastischer Modellierung

Beweis: Mit Hilfe des Satzes von Fubini (Satz 23) und Lemma 34 erhalten wir für

$$G(x_1, \dots, x_d) = \prod_{i=1}^d x_i :$$

$$\begin{aligned} E\left(\prod_{i=1}^d X_i\right) &= \int G dP^{\mathbf{X}} = \int G d\bigotimes_{i=1}^d P^{X_i} = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} G(x_1, \dots, x_d) P^{X_1}(dx_1) \dots P^{X_d}(dx_d) \\ &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \prod_{i=1}^d x_i P^{X_1}(dx_1) \dots P^{X_d}(dx_d) = \prod_{i=1}^d \int_{-\infty}^{\infty} x_i P^{X_i}(dx_i) = \prod_{i=1}^d E(X_i), \end{aligned}$$

womit das Lemma bewiesen ist. ■

Mit Hilfe von Lemma 32 kann damit alternativ auch Teil d) des Satzes 32 bewiesen werden, weil ja die Zufallsvariable X dort genau so verteilt ist wie $F^{-1}(U)$, wobei U eine über $[0,1]$ stetig gleichverteilte Zufallsvariable bezeichne. Es folgt dann mit Lemma 34

$$E(X) = E(F^{-1}(U)) = \int F^{-1} dP^U = \int_0^1 F^{-1}(x) dx.$$

Die folgende Ungleichung erweist sich in vielen Anwendungsfällen ebenfalls als sehr nützlich:

Lemma 36 (Jensen'sche Ungleichung). Es sei $\mathbf{X} = (X_1, \dots, X_d)$ ein d -dimensionaler Zufallsvektor auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Werten in einer konvexen Menge $\mathcal{M} \in \mathcal{B}^d$ mit $d \in \mathbb{N}$ und $G: \mathcal{M} \rightarrow \mathbb{R}$ eine messbare, konvexe (konkave) Abbildung. $G(\mathbf{X})$ sowie die Komponenten X_1, \dots, X_d von \mathbf{X} seien sämtlich P -integrierbar. Dann ist $(E(X_1), \dots, E(X_d)) \in \mathcal{M}$, und es gilt:

$$E(G(\mathbf{X})) \begin{cases} \geq G(E(X_1), \dots, E(X_m)), & \text{falls } G \text{ konvex} \\ \leq G(E(X_1), \dots, E(X_m)), & \text{falls } G \text{ konkav} \end{cases}$$

Ist G nicht-negativ und konkav, braucht dabei lediglich die Integrierbarkeit aller Komponenten von \mathbf{X} gefordert zu werden.

Beweis: Die erste Aussage des Lemmas wollen wir hier nur für den Fall beweisen, dass \mathcal{M} kompakt ist (einen vollständigen Beweis findet man z.B. in Gänsler & Stute (1977), Satz 5.4.1).

Es bezeichne $\mu = (\mu_1, \dots, \mu_d) = (E(X_1), \dots, E(X_d))$. Angenommen, μ liegt außerhalb von \mathcal{M} .

Dann existiert eine μ von \mathcal{M} strikt trennende Hyperebene, d.h. es existieren Zahlen $a_1, \dots, a_d \in \mathbb{R}$ mit

$$\sum_{i=1}^d a_i (x_i - \mu_i) > 0 \text{ für alle } x_1, \dots, x_d \in \mathcal{M}.$$

Insbesondere ist also $S = \sum_{i=1}^d a_i (X_i - \mu_i) > 0$ und damit auch $E(S) > 0$, denn anderenfalls ergäbe

II. Grundprinzipien stochastischer Modellierung

sich aus Monotoniegründen $E(S) = 0$ und daher mit Lemma 33 f) $S = 0$ P -fast sicher im Widerspruch zu $S > 0$. Damit folgt

$$0 < E(S) = \sum_{i=1}^d a_i (E(X_i) - \mu_i) = 0,$$

also ein erneuter Widerspruch, womit die Gültigkeit von $\mu \in \mathcal{M}$ bewiesen ist.

Zum Beweis der Ungleichungen für $E(G(\mathbf{X}))$ können wir annehmen, dass G konvex ist (sonst Übergang von G zu $-G$). Es existiert dann eine Stützhyperebene, die den Graphen von G im Punkt μ berührt, d.h. es existieren Zahlen $b_1, \dots, b_d \in \mathbb{R}$ mit

$$G(x_1, \dots, x_d) \geq \sum_{i=1}^d b_i (x_i - \mu_i) + G(\mu_1, \dots, \mu_d) \text{ mit } (x_1, \dots, x_d) \in \mathcal{M}.$$

Mit der Monotonie und Linearität des Erwartungswerts folgt daher

$$E(G(X_1, \dots, X_d)) \geq \sum_{i=1}^d b_i (E(X_i) - \mu_i) + G(\mu_1, \dots, \mu_d) = G(\mu_1, \dots, \mu_d),$$

wie behauptet. Die restliche Behauptung ergibt sich aus Lemma 33 c) sowie dem Umstand, dass bei Konkavität und Nichtnegativität von G analog Zahlen $c_1, \dots, c_d \in \mathbb{R}$ existieren mit

$$0 \leq G(X_1, \dots, X_d) \leq \sum_{i=1}^d c_i (X_i - \mu_i) + G(\mu_1, \dots, \mu_d).$$

Das Lemma ist damit vollständig bewiesen. ■

Eine Verschärfung der Jensen'schen Ungleichung ergibt sich für den Fall, dass G strikt konvex (strikt konkav) ist und das Maß P neben 0 und 1 noch mindestens einen weiteren Wert annimmt. In diesem Fall gilt

$$E(G(\mathbf{X})) \begin{cases} > G(E(X_1), \dots, E(X_m)), & \text{falls } G \text{ konvex} \\ < G(E(X_1), \dots, E(X_m)), & \text{falls } G \text{ konkav.} \end{cases}$$

Definition 33 (Varianz, Kovarianz und höhere Momente). Es seien X und Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) sowie $k \in \mathbb{N}$.

a) Ist X^2 P -integrierbar, so auch X ; in diesem Fall heißt

$$\text{Var}(X) := E\left(\left(X - E(X)\right)^2\right)$$

die *Varianz* von X bzw. von P^X .

b) Sind X, Y und $X \cdot Y$ jeweils P -integrierbar, so heißt

$$\text{Kov}(X, Y) := E\left[\left(X - E(X)\right)\left(Y - E(Y)\right)\right]$$

II. Grundprinzipien stochastischer Modellierung

die Kovarianz von X und Y . Ist $Kov(X, Y) = 0$, so heißen X und Y *unkorreliert*.

c) Ist $|X|^k$ P -integrierbar, so auch X ; in diesem Fall heißt

- $E(|X|^k)$ das k -te absolute Moment von X
- $E(X^k)$ das k -te Moment von X
- $E(|X - E(X)|^k)$ das k -te absolute zentrale Moment von X und
- $E((X - E(X))^k)$ das k -te zentrale Moment von X (bzw. P^X).

Man beachte dabei, dass die Zufallsvariable $|X|^k + 1$ für jedes $k \in \mathbb{N}$ eine P -integrierbare Majorante zu $|X|$ ist, wenn $|X|^k$ P -integrierbar ist. (vgl. Lemma 33 c) und d)).

Lemma 37 (Eigenschaften von Momenten). Es seien X und Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

a) Sind X^2, Y und $X \cdot Y$ P -integrierbar, so gilt auch

$$Var(X) = Kov(X, X) = E(X^2) - (E(X))^2$$

und

$$Kov(X, Y) = E(XY) - E(X)E(Y).$$

Für beliebige $c \in \mathbb{R}$ gilt darüber hinaus:

$$E((X - c)^2) = Var(X) + (E(X) - c)^2,$$

d.h. $E((X - c)^2)$ ist minimal für $c = E(X)$. Ferner gilt:

$$Var(X) = 0 \Leftrightarrow X = \text{const } P\text{-fast sicher.}$$

b) Sind X^2, Y und $X \cdot Y$ P -integrierbar, so gilt:

$$Var(aX + b) = a^2 \cdot Var(X)$$

$$Kov(aX + b, cY + d) = ac \cdot Kov(X, Y)$$

für alle $a, b, c, d \in \mathbb{R}$. Ferner ist

$$Var(X + Y) = Var(X) + Var(Y) + 2 Kov(X, Y).$$

Sind speziell X und Y unkorreliert, so gilt

$$Var(X + Y) = Var(X) + Var(Y).$$

II. Grundprinzipien stochastischer Modellierung

c) Ist X^2 P -integrierbar, so gilt

$$P(|X - E(X)| > \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2}$$

für alle $\varepsilon > 0$ (Tschebyscheff-Ungleichung).

d) Existiert für ein $k \in \mathbb{N}$ das k -te absolute Moment von X , so existieren auch alle m -ten Momente von X für $m \leq k$ (d.h. absolute und nicht-absolute, zentrale und nicht-zentrale), und es gilt

$$E((X - c)^m) \leq E(|X - c|^m) \leq (E(|X - c|^k))^{m/k} \leq 2^m (E(|X|^k) + |c|^k)^{m/k}$$

für alle $c \in \mathbb{R}$. Bezeichnet F die Verteilungsfunktion von X bzw. P^X , so gilt auch die Darstellung

$$E(X^k) = (-1)^k \int_{-\infty}^0 kx^{k-1} F(x) dx + \int_0^{\infty} kx^{k-1} (1 - F(x)) dx = \int_0^1 (F^{-1}(x))^k dx$$

$$E(|X|^k) = \int_{-\infty}^0 kx^{k-1} F(x) dx + \int_0^{\infty} kx^{k-1} (1 - F(x)) dx = \int_0^1 |F^{-1}(x)|^k dx.$$

e) Sind $|X|^p$ und $|Y|^q$ P -integrierbar für $p, q > 1$ mit $\frac{1}{p} + \frac{1}{q} = 1$, so ist auch $|XY|$ P -integrierbar, und es gilt

$$E(|X \cdot Y|) \leq (E(|X|^p))^{1/p} \cdot (E(|Y|^q))^{1/q} \quad (\text{Hölder-Ungleichung}).$$

Ist speziell $p = q = 2$, so gilt auch

$$|\text{Kov}(X, Y)| \leq \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}$$

mit Gleichheit genau dann, wenn X und Y P -f.s. linear voneinander abhängen, d.h. wenn Zahlen $a, b, c \in \mathbb{R}$, a, b nicht beide Null, existieren mit $aX + bY = c$ P -fast sicher.

Beweis:

a) Die Beziehung zwischen Varianz und Kovarianz ergibt sich unmittelbar aus Definition 33 a) und b) durch Auflösen der Klammern und der Linearität des Erwartungswertes:

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

II. Grundprinzipien stochastischer Modellierung

$$\begin{aligned} \text{Kov}(X, Y) &= E[(X - E(X))(Y - E(Y))] = E[XY - YE(X) - XE(Y) + E(X)E(Y)] \\ &= E(XY) - 2E(X)E(Y) + E(X)E(Y) = E(XY) - E(X)E(Y). \end{aligned}$$

Entsprechend gilt mit $\mu = E(X)$:

$$\begin{aligned} E((X - c)^2) &= E(X^2 - 2cX + c^2) \\ &= E(X^2 - 2\mu X + \mu^2) + 2(\mu - c)E(X) + c^2 - \mu^2 \\ &= \text{Var}(X) + 2(\mu - c)\mu + c^2 - \mu^2 \\ &= \text{Var}(X) + c^2 - 2\mu c + \mu^2 \\ &= \text{Var}(X) + (E(X) - c)^2. \end{aligned}$$

Ferner gilt nach Lemma 33 f)

$$\text{Var}(X) = 0 \Leftrightarrow E(X - E(X))^2 = 0 \Leftrightarrow (X - E(X))^2 = 0 \Leftrightarrow X = E(X) \text{ P-fast sicher.}$$

Dies bedeutet aber gerade $\text{Var}(X) = 0 \Leftrightarrow X = \text{const}$ P-fast sicher.

b) Es gilt

$$\begin{aligned} \text{Var}(aX + b) &= E[(aX + b - E\{aX + b\})^2] = E[(aX - aE(X))^2] \\ &= a^2 \cdot E[(X - E(X))^2] = a^2 \cdot \text{Var}(X), \\ \text{Kov}(aX + b, cY + d) &= E[(aX + b - \{aE(X) + b\}) \cdot (cY + d - \{cE(Y) + d\})] \\ &= E[a(X - E(X)) \cdot c(Y - E(Y))] = ac \cdot E[(X - E(X)) \cdot (Y - E(Y))] \\ &= ac \cdot \text{Kov}(X, Y). \end{aligned}$$

Entsprechend gilt mit $\mu = E(X)$ und $\nu = E(Y)$:

$$\begin{aligned} \text{Var}(X + Y) &= E((X + Y)^2) - (\mu + \nu)^2 \\ &= E(X^2 + 2XY + Y^2) - (\mu^2 + 2\mu\nu + \nu^2) \\ &= E(X^2) - \mu^2 + E(Y^2) - \nu^2 + 2[E(XY) - \mu\nu] \\ &= \text{Var}(X) + \text{Var}(Y) + 2 \text{Kov}(X, Y), \end{aligned}$$

wie behauptet. Sind X und Y zusätzlich unkorreliert, so ist $\text{Kov}(X, Y) = 0$, womit $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ folgt.

c) Die Tschebyscheff-Ungleichung ist ein Spezialfall der Markoff-Ungleichung:

II. Grundprinzipien stochastischer Modellierung

$$P(|X - E(X)| > \varepsilon) = P(|X - E(X)|^2 > \varepsilon^2) \leq \frac{1}{\varepsilon^2} E((X - E(X))^2) = \frac{\text{Var}(X)}{\varepsilon^2}.$$

d) Es reicht, auf Grund von Lemma 33 c) und d) nur die Beziehung

$$E((X - c)^m) \leq E(|X - c|^m) \leq \left(E(|X - c|^k)\right)^{m/k} \leq 2^m \left(E(|X|^k) + |c|^k\right)^{m/k} \quad \text{für alle } c \in \mathbb{R}$$

nachzuweisen. Diese folgt aber für $m < k$ aus der Hölder-Ungleichung Teil e) für die Wahl $p = \frac{k}{m}$, $q = \frac{k}{k-m}$ und $Y \equiv 1$, wenn dort X durch $|X|^m$ ersetzt wird sowie aus der für alle reellen Zahlen a, b gültigen Abschätzung

$$|a - b|^k \leq (|a| + |b|)^k \leq (2 \max(|a|, |b|))^k \leq 2^k (|a|^k + |b|^k).$$

Der Beweis für die Darstellung von

$$E(X^k) = (-1)^k \int_{-\infty}^0 kx^{k-1} F(x) dx + \int_0^{\infty} kx^{k-1} (1 - F(x)) dx = \int_0^1 (F^{-1}(x))^k dx$$

und

$$E(|X|^k) = \int_{-\infty}^0 kx^{k-1} F(x) dx + \int_0^{\infty} kx^{k-1} (1 - F(x)) dx = \int_0^1 |F^{-1}(x)|^k dx$$

ergibt sich analog zum Beweis zu Satz 32 c) und d).

e) Wir benutzen die Jensen'sche Ungleichung aus Lemma 36 für die konkave Abbildung

$$G(u, v) = u^{\frac{1}{p}} v^{\frac{1}{q}}, \quad u, v > 0.$$

Die Konkavität ergibt sich z.B. aufgrund von

$$\begin{aligned} \frac{\partial^2 G}{\partial u^2}(u, v) &= -\frac{p-1}{p^2} u^{\frac{1}{p}-2} v^{\frac{1}{q}} < 0, & \frac{\partial^2 G}{\partial v^2}(u, v) &= -\frac{q-1}{q^2} u^{\frac{1}{p}} v^{\frac{1}{q}-2} < 0, \\ \frac{\partial^2 G}{\partial u \partial v}(u, v) &= \frac{1}{pq} u^{\frac{1}{p}-1} v^{\frac{1}{q}-1} > 0, \end{aligned}$$

d.h. die Matrix $\begin{pmatrix} \frac{\partial^2 G}{\partial u^2} & \frac{\partial^2 G}{\partial u \partial v} \\ \frac{\partial^2 G}{\partial u \partial v} & \frac{\partial^2 G}{\partial v^2} \end{pmatrix}$ ist negativ-semidefinit auf der konvexen Menge

$\mathcal{M} = (0, \infty) \times (0, \infty)$ (mit den Eigenwerten 0 und $-u^{\frac{1}{p}-2} v^{\frac{1}{q}-2} \frac{(p-1)qv^2 + (q-1)pu^2}{p^2q^2} < 0$). Nach

Lemma 35 folgt also

II. Grundprinzipien stochastischer Modellierung

$$E(|XY|) = E(G(|X|^p |Y|^q)) \leq G(E(|X|^p), E(|Y|^q)) = (E(|X|^p))^{1/p} \cdot (E(|Y|^q))^{1/q},$$

wie behauptet. Die Ungleichung

$$|Kov(X, Y)| \leq \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}$$

ergibt sich nun aus der Hölder-Ungleichung, wenn man $p = q = 2$ wählt und X durch $X - E(X)$ sowie Y durch $Y - E(Y)$ ersetzt.

Seien nun $a, b, c \in \mathbb{R}$, a, b nicht beide Null, etwa o.B.d.A. $a \neq 0$, mit $aX + bY = c$, dann gilt nach Teil b)

$$\begin{aligned} |a| \cdot |Kov(X, Y)| &= |Kov(aX, Y)| \stackrel{aX+bY=c}{=} |Kov(c - bY, Y)| = |b| \text{Var}(Y) \\ &= \sqrt{b^2 \text{Var}(Y)} \sqrt{\text{Var}(Y)} = \sqrt{\text{Var}(aX) \text{Var}(Y)} \\ &= |a| \sqrt{\text{Var}(X) \text{Var}(Y)}, \end{aligned}$$

also $|Kov(X, Y)| = \sqrt{\text{Var}(X) \cdot \text{Var}(Y)}$.

Ist umgekehrt $Kov(X, Y) = \pm \sqrt{\text{Var}(X) \text{Var}(Y)}$ und nicht $\sqrt{\text{Var}(X)} = \sqrt{\text{Var}(Y)} = 0$, so ergibt sich für alle $a, b \in \mathbb{R}$ die Gleichung

$$\begin{aligned} \text{Var}(aX + bY) &= a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \cdot Kov(X, Y) \\ &= a^2 \text{Var}(X) + b^2 \text{Var}(Y) \pm 2ab \sqrt{\text{Var}(X) \text{Var}(Y)} \\ &= \left(a \sqrt{\text{Var}(X)} \pm b \sqrt{\text{Var}(Y)} \right)^2, \end{aligned}$$

also speziell $\text{Var}(aX + bY) = 0$ für die Wahl $a = \sqrt{\text{Var}(Y)}$, $b = \mp \sqrt{\text{Var}(X)}$. Nach Teil a) ist dann aber $aX + bY = \text{const}$ P -fast sicher, wie behauptet.

Für $\sqrt{\text{Var}(X)} = \sqrt{\text{Var}(Y)} = 0$ sind X und Y P -fast sicher konstant, so dass sich in diesem Fall a und b stets so wählen lassen, dass $aX + bY = \text{const}$ P -fast sicher ist, ohne dass $a = b = 0$ gilt. Damit ist das Lemma vollständig bewiesen. ■

Definition 34 (Korrelation). Es seien X und Y Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit $0 < \text{Var}(X), \text{Var}(Y) < \infty$. Dann heißt

$$\text{Korr}(X, Y) = \frac{Kov(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}}$$

die Korrelation zwischen X und Y .

II. Grundprinzipien stochastischer Modellierung

Bemerkung: Die Korrelation ist also ein Maß für die lineare Abhängigkeit zwischen X und Y ; sie nimmt nach obigem Werte zwischen -1 und 1 an mit $\text{Korr}(X, Y) = \pm 1$ genau dann, wenn Zahlen $a, b, c \in \mathbb{R}$, a, b nicht beide Null, existieren mit $aX + bY = c$ P -fast sicher. Ist $\text{Korr}(X, Y) > 0$, so sagt man auch, X und Y seien *positiv* korreliert; ist $\text{Korr}(X, Y) < 0$, so heißen X und Y *negativ* korreliert.

Oft wird die Korrelation auch mit dem Buchstaben ρ bezeichnet.

Lemma 37 b) lässt sich sofort auch auf mehr als zwei Summanden X_1, \dots, X_n erweitern; es gilt dann entsprechend

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq n} \text{Kov}(X_i, X_j).$$

Lemma 35 impliziert, dass stochastisch unabhängige Zufallsvariablen X_1, \dots, X_d auch (paarweise) unkorreliert sind. Die Umkehrung ist aber i.Allg. nicht richtig, wie das folgende Beispiel zeigt:

Wir betrachten das zweimalige unabhängige Werfen einer fairen (Laplace-)Münze. Das Ergebnis „Kopf“ identifizieren wir mit „ -1 “, „Zahl“ mit „ 1 “. X bezeichne das Ergebnis im ersten, Y das im zweiten Wurf. Dann sind die Zufallsvariablen $S := X + Y$ und $D := X - Y$ unkorreliert, denn es gilt

$$E(S \cdot D) = E(X^2 - Y^2) = E(X^2) - E(Y^2) = 0,$$

weil X und Y dieselbe (Rand-)Verteilung besitzen mit $X^2 = Y^2 = 1$, also auch $E(X^2) = E(Y^2) = 1$.

Wegen $E(X) = E(Y) = -1 \cdot \frac{1}{2} + 1 \cdot \frac{1}{2} = 0$ folgt also $\text{Kov}(S, D) = E(S \cdot D) - E(S)E(D) = 0$. S und D sind aber nicht stochastisch unabhängig, weil etwa

$$\begin{aligned} P(S = 2, D = 0) &= P(X = Y = 1) = P(X = 1) \cdot P(Y = 1) = \frac{1}{4} \neq \frac{1}{8} = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{4} \cdot \left(\frac{1}{4} + \frac{1}{4}\right) \\ &= P(X = Y = 1) \cdot (P(\{X = Y = 1\} \oplus \{X = Y = -1\})) = P(S = 2) \cdot P(D = 0) \end{aligned}$$

gilt.

Der Erwartungswertbegriff ist im übrigen zentral für *faire* Spiele. Ein Spiel heißt *fair*, wenn der Einsatz gleich der erwarteten Auszahlung, der erwartete Gewinn also Null ist. Beim *chuck-a-luck* (S. 70 – 72) ergibt sich etwa aus der Tabelle auf S. 72 oben, für alle $k \in \{1, \dots, 6\}$:

$$E(X_k) = \sum_{x \in \mathfrak{X}} x \cdot P(X_k = x) = \frac{-126 + 75 + 30 + 3}{216} = -\frac{17}{216} \approx -0,079.$$

Dieser Wert ist negativ, was bedeutet, dass das Spiel unfair ist, und zwar zu Lasten des Spielers. Er verliert auf Dauer im Mittel ca. 8 Cent (auf Grund des Gesetzes der Großen Zahlen, siehe Abschnitt II.6.); siehe auch BARTH UND HALLER (1998), S. 168.

II. Grundprinzipien stochastischer Modellierung

Von zentraler Bedeutung in der Stochastik sind ferner (die Verteilungen von) Summen unabhängiger Zufallsvariablen, so etwa beim so genannten *Gesetz der großen Zahlen* und beim *Zentralen Grenzwertsatz*, die in Abschnitt II.6 und II.7 gesondert behandelt werden.

Definition 35 (Faltung). Es seien X und Y reellwertige, stochastisch unabhängige Zufallsvariablen. Dann heißt die Verteilung der Summe $X + Y$ die Faltung der Verteilungen von X und Y , in Zeichen:

$$P^X * P^Y := P^{X+Y}.$$

Die Faltung von Verteilungen kann auf verschiedene Weise dargestellt werden:

Lemma 38. Es seien X und Y stochastisch unabhängige, reellwertige Zufallsvariablen. Ihre Verteilungsfunktionen seien mit F_X und F_Y bezeichnet, bei stetigen Verteilungen ferner ihre Dichten entsprechend mit f_X und f_Y . Die Verteilungsfunktion der Faltung $P^X * P^Y$ sei mit $F_X * F_Y$ bezeichnet, entsprechend im Fall der Existenz die Dichte mit $f_X * f_Y$. Dann gilt:

$$\begin{aligned} F_X * F_Y(z) &= F_Y * F_X(z) = \int_{-\infty}^{\infty} F_X(z-y) P^Y(dy) = \int_{-\infty}^{\infty} F_Y(z-x) P^X(dx) \\ F_X * F_Y(z) &= F_Y * F_X(z) = \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy = \int_{-\infty}^{\infty} F_Y(z-x) f_X(x) dx \\ f_X * f_Y(z) &= f_Y * f_X(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy = \int_{-\infty}^{\infty} f_Y(z-x) f_X(x) dx, \quad z \in \mathbb{R}. \end{aligned}$$

Beweis: Es ist

$$\begin{aligned} F_X * F_Y(z) &= P(X + Y \leq z) = \iint_{x+y \leq z} P^{(X,Y)}(dx, dy) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} P^X(dx) P^Y(dy) = \int_{-\infty}^{\infty} P(X \leq z-y) P^Y(dy) \\ &= \int_{-\infty}^{\infty} F_X(z-y) P^Y(dy) = \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy \end{aligned}$$

(die letzte Gleichung im Fall der Existenz von Dichten), woraus durch Differentiation folgt

$$f_X * f_Y(z) = \frac{d}{dz} F_X * F_Y(z) = \frac{d}{dz} \int_{-\infty}^{\infty} F_X(z-y) f_Y(y) dy = \int_{-\infty}^{\infty} f_X(z-v) f_Y(v) dv. \quad \blacksquare$$

Für diskrete Verteilungen über \mathbb{Z} (oder Teilmengen davon) gilt folgende, analoge Aussage:

Lemma 39. Es seien X und Y stochastisch unabhängige, reellwertige Zufallsvariablen mit Werten in \mathbb{Z} . Dann gilt:

$$\begin{aligned} F_X * F_Y(n) &= P(X + Y \leq n) = \sum_{k \in \mathbb{Z}} F_Y(n-k) \cdot P(X = k) = \sum_{k \in \mathbb{Z}} F_X(n-k) \cdot P(Y = k) \\ P(X + Y = n) &= \sum_{k \in \mathbb{Z}} P(X = k) \cdot P(Y = n-k) = \sum_{k \in \mathbb{Z}} P(Y = k) \cdot P(X = n-k) \end{aligned}$$

für alle $n \in \mathbb{Z}$.

II. Grundprinzipien stochastischer Modellierung

Beweis: Direktes Nachrechnen und Anwenden der Unabhängigkeitsannahme liefert

$$P(X + Y = n) = P\left(\bigoplus_{k \in \mathbb{Z}} \{X = k\} \cap \{Y = n - k\}\right) = \sum_{k \in \mathbb{Z}} P(X = k) \cdot P(Y = n - k), \quad n \in \mathbb{Z}.$$

Hieraus folgt unmittelbar

$$\begin{aligned} F_X * F_Y(n) &= P(X + Y \leq n) = \sum_{k=-\infty}^n P(X + Y = k) = \sum_{k=-\infty}^n \sum_{j \in \mathbb{Z}} P(X = j) \cdot P(Y = k - j) \\ &= \sum_{j \in \mathbb{Z}} P(X = j) \sum_{k=-\infty}^n P(Y = k - j) = \sum_{j \in \mathbb{Z}} F_Y(n - j) \cdot P(X = j), \quad n \in \mathbb{Z}. \end{aligned}$$

Die übrigen Aussagen ergeben sich hieraus durch Vertauschen der Rollen von X und Y . ■

Für den Fall, dass die Zufallsvariablen in Lemma 38 und 39 [fast sicher] nicht-negativ sind, muss nicht über ganz \mathbb{R} bzw. über ganz \mathbb{Z} summiert werden. Man erhält dann folgende Spezialfälle:

Im Fall von Lemma 38:

$$\begin{aligned} F_X * F_Y(z) &= F_Y * F_X(z) = \int_0^z F_X(z - y) P^Y(dy) = \int_0^z F_Y(z - y) P^X(dy) \\ F_X * F_Y(z) &= F_Y * F_X(z) = \int_0^z F_X(z - y) f_Y(y) dy = \int_0^z F_Y(z - x) f_X(x) dx \\ f_X * f_Y(z) &= f_Y * f_X(z) = \int_0^z f_X(z - y) f_Y(y) dy = \int_0^z f_Y(z - x) f_X(x) dx, \quad z \in \mathbb{R} \end{aligned}$$

und im Fall von Lemma 39:

$$\begin{aligned} F_X * F_Y(n) &= P(X + Y \leq n) = \sum_{k=0}^n F_Y(n - k) \cdot P(X = k) = \sum_{k=0}^n F_X(n - k) \cdot P(Y = k) \\ P(X + Y = n) &= \sum_{k=0}^n P(X = k) \cdot P(Y = n - k) = \sum_{k=0}^n P(Y = k) \cdot P(X = n - k) \end{aligned}$$

für alle $n \in \mathbb{Z}^+$.

Natürlich kann man auf entsprechende Weise auch die Summe *abhängiger* Zufallsvariablen behandeln. Wir zeigen dies an einem Beispiel, das zugleich noch einmal alternative Berechnungsmöglichkeiten für Erwartungswerte, Varianzen und Korrelationen aufzeigt.

II. Grundprinzipien stochastischer Modellierung

Beispiel: Wir betrachten einen Zufallsvektor $Z = (X, Y)$ mit Werten in dem Dreieck $D := \{(x, y) | 0 \leq x \leq y \leq 1\}$. Es sei angenommen, dass die Verteilung von (X, Y) eine Lebesgue-Dichte besitzt; diese muss aufgrund der Vorgabe des Wertebereichs dann nur auf D spezifiziert werden. Die Dichte möge von folgender Form sein:

$$f_Z(x, y) = c \cdot (x + y) \text{ für } (x, y) \in D$$

mit einer noch passend zu bestimmenden positiven Konstanten c . Diese erhalten wir aus der Bedingung

$$\begin{aligned} 1 = P(Z \in D) &= \iint_D f_Z(x, y) dx dy = c \int_0^1 \int_0^y x + y dx dy \\ &= c \int_0^1 \left[\frac{x^2}{2} + xy \right]_{x=0}^{x=y} dy = c \int_0^1 \frac{3}{2} y^2 dy = c \left[\frac{y^3}{2} \right]_{y=0}^{y=1} = \frac{c}{2}, \end{aligned}$$

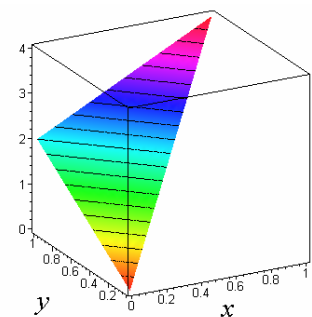
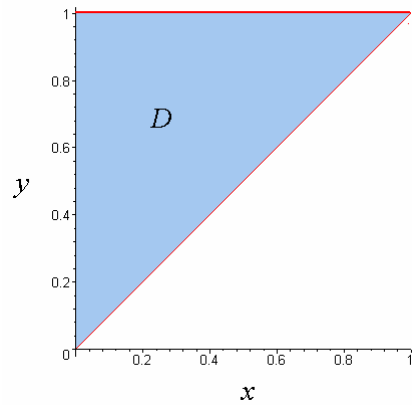
also $c = 2$. Man muss hier übrigens im Gegensatz zum abstrakten Doppelintegral bei der Festlegung der Integrationsgrenzen in der iterativen Auflösung auf die Reihenfolge der Integration achten. Man hätte die Konstante nämlich auch aus der Rechnung

$$\begin{aligned} 1 = P(Z \in D) &= \iint_D f_Z(x, y) dy dx = c \int_0^1 \int_x^1 x + y dy dx \\ &= c \int_0^1 \left[xy + \frac{y^2}{2} \right]_{y=x}^{y=1} dx = c \int_0^1 \left(\frac{1}{2} + x - \frac{3}{2} x^2 \right) dx = c \left[\frac{x}{2} + \frac{x^2}{2} - \frac{x^3}{2} \right]_{x=0}^{x=1} = \frac{c}{2} \end{aligned}$$

erhalten können. Die jeweiligen Integrationsgrenzen ergeben sich hierbei aus der das Dreieck charakterisierenden Ungleichung $0 \leq x \leq y \leq 1$. Integriert man zuerst nach x , zieht die Ungleichung $0 \leq x \leq y$; diese legt dann die Integrationsgrenzen für das innere Integral fest. Nach „Herausintegrieren“ von x bleibt dann nur noch die Ungleichung $0 \leq y \leq 1$, welche die Grenzen für das äußere Integral festlegt. Für das zweite Doppelintegral zieht dagegen zuerst die Ungleichung $x \leq y \leq 1$ und dann die Ungleichung $0 \leq x \leq 1$.

Die Dichte von Z lautet somit:

$$f_Z(x, y) = 2(x + y), \quad (x, y) \in D \Leftrightarrow 0 \leq x \leq y \leq 1.$$



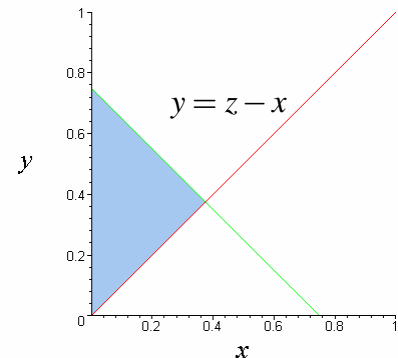
Graph der Dichte von Z

II. Grundprinzipien stochastischer Modellierung

Für die Berechnung der Summendichte für $S := X + Y$ gehen wir wie im Beweis von Lemma 38 vor, allerdings müssen wir jetzt auf Grund der speziellen Situation einer Dichte über einem Dreieck eine Fallunterscheidung machen.

Fall 1: $0 \leq z \leq 1$: Es sind die Ungleichungen $0 \leq x \leq y \leq 1$ und $x + y \leq z$ zu beachten. Diese beschreiben das rechts blau markierte Teildreieck, das äquivalent bestimmt ist durch die Ungleichungen

$$0 \leq x \leq \frac{z}{2}, \quad x \leq y \leq z - x.$$



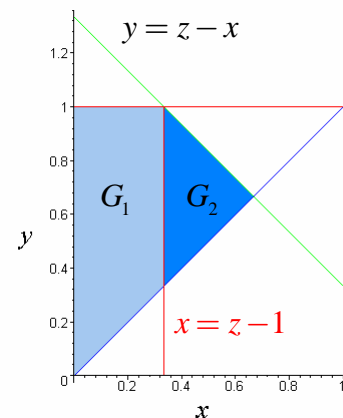
Die einfachste innere Integration ist also die nach y mit

$$\begin{aligned} F_S(z) &= P(X + Y \leq z) = \iint_{\substack{x+y \leq z \\ 0 \leq x \leq y \leq 1}} f_Z(x, y) dy dx = 2 \int_0^{z/2} \int_x^{z-x} x + y dy dx = 2 \int_0^{z/2} xy + \frac{y^2}{2} \Big|_{y=x}^{y=z-x} dx \\ &= \int_0^{z/2} z^2 - 4x^2 dx = z^2 x - \frac{4}{3} x^3 \Big|_{x=0}^{x=z/2} = \left(\frac{1}{2} - \frac{1}{6} \right) z^3 = \frac{z^3}{3}. \end{aligned}$$

Fall 2: $1 \leq z \leq 2$: Es sind wieder die Ungleichungen $0 \leq x \leq y \leq 1$ und $x + y \leq z$ zu beachten. Diese beschreiben diesmal die Vereinigung der beiden rechts blau markierten Gebiete G_1 und G_2 , die äquivalent bestimmt sind durch die Ungleichungen

$$0 \leq x \leq z - 1, \quad x \leq y \leq 1 \text{ für } G_1$$

$$z - 1 \leq x \leq \frac{z}{2}, \quad x \leq y \leq z - x \text{ für } G_2.$$



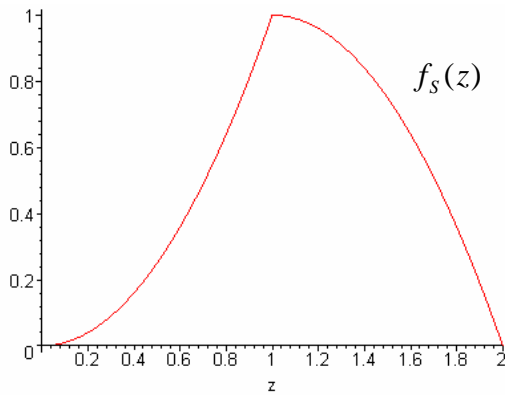
Die einfachste innere Integration ist also jeweils wieder die nach y mit

$$\begin{aligned} F_S(z) &= P(X + Y \leq z) = 2 \int_0^{z-1} \int_x^1 x + y dy dx + 2 \int_{z-1}^{z/2} \int_x^{z-x} x + y dy dx \\ &= 2 \int_0^{z-1} \frac{1}{2} + x - \frac{3}{2} x^2 dx + 2 \int_{z-1}^{z/2} z^2 - 4x^2 dx = x + x^2 - x^3 \Big|_0^{z-1} + \left(2z^2 x - \frac{8}{3} x^3 \Big|_{z-1}^{z/2} \right) = -\frac{1}{3} + z^2 - \frac{z^3}{3}. \end{aligned}$$

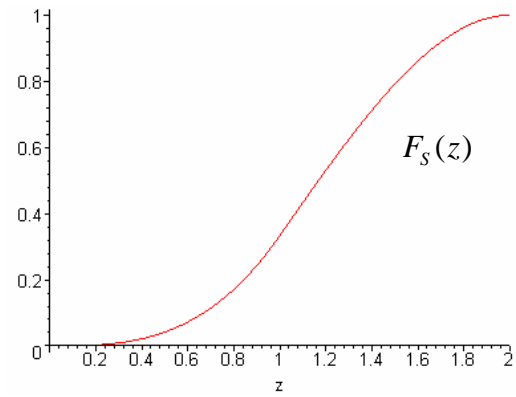
Durch Differenzieren erhalten wir damit auch noch gleich die Summendichte:

$$f_S(z) = \begin{cases} z^2, & 0 \leq z \leq 1 \\ 2z - z^2, & 1 \leq z \leq 2 \end{cases}$$

II. Grundprinzipien stochastischer Modellierung



Dichte von S



Verteilungsfunktion von S

Weiterhin erhalten wir noch nach Lemma 34, mit $G(x, y) = x$:

$$E(X) = \iint_D x \cdot f_Z(x, y) dy dx = 2 \int_0^1 \int_0^y x^2 + xy dx dy = 2 \int_0^1 \frac{x^3}{3} + \frac{x^2}{2} y \Big|_{x=0}^{x=y} dy = \frac{5}{3} \int_0^1 y^3 dy = \frac{5}{12}.$$

Eine analoge Rechnung mit $G(x, y) = y$ liefert

$$E(Y) = \iint_D y \cdot f_Z(x, y) dy dx = 2 \int_0^1 \int_0^y xy + y^2 dx dy = \int_0^1 x^2 y + 2xy^2 \Big|_{x=0}^{x=y} dy = \int_0^1 3y^3 dy = \frac{3}{4}.$$

Diese beiden Größen hätte man natürlich auch nach Satz 32 b) über die Randdichten berechnen können:

$$f_X(x) = \int_0^1 f_Z(x, y) dy = 2 \int_x^1 x + y dy = 2xy + y^2 \Big|_{y=x}^{y=1} = 1 + 2x - 3x^2, \quad 0 \leq x \leq 1$$

$$E(X) = \int_0^1 x \cdot f_X(x) dx = \int_0^1 x + 2x^2 - 3x^3 dx = \frac{x^2}{2} + \frac{2}{3}x^3 - \frac{3}{4}x^4 \Big|_0^1 = \frac{1}{2} + \frac{2}{3} - \frac{3}{4} = \frac{5}{12}$$

$$f_Y(y) = \int_0^y f_Z(x, y) dx = 2 \int_0^y x + y dx = x^2 + 2xy \Big|_{x=0}^{x=y} = 3y^2, \quad 0 \leq y \leq 1$$

$$E(Y) = \int_0^1 y \cdot f_Y(y) dy = \int_0^1 3y^3 dy = \frac{3}{4}y^4 \Big|_0^1 = \frac{3}{4}.$$

II. Grundprinzipien stochastischer Modellierung

Für die Kovarianz / Korrelation zwischen X und Y ergibt sich analog:

$$E(XY) = \iint_D xy \cdot f_Z(x, y) dy dx = 2 \int_0^1 \int_0^y x^2 y + xy^2 dx dy = \int_0^1 \frac{2}{3} x^3 y + x^2 y^2 \Big|_{x=0}^{x=y} dy = \int_0^1 \frac{5}{3} y^4 dy = \frac{1}{3}$$

$$Kov(X, Y) = E(XY) - E(X)E(Y) = \frac{1}{3} - \frac{15}{48} = \frac{1}{48}$$

$$E(X^2) = \iint_D x^2 \cdot f_Z(x, y) dy dx = 2 \int_0^1 \int_0^y x^3 + x^2 y dx dy = \int_0^1 \frac{1}{2} x^4 + \frac{2}{3} x^3 y \Big|_{x=0}^{x=y} dy = \int_0^1 \frac{7}{6} y^4 dy = \frac{7}{30}$$

$$E(Y^2) = \iint_D y^2 \cdot f_Z(x, y) dy dx = 2 \int_0^1 \int_0^y xy^2 + y^3 dx dy = \int_0^1 x^2 y^2 + 2xy^3 \Big|_{x=0}^{x=y} dy = \int_0^1 3y^4 dy = \frac{3}{5}$$

$$Var(X) = E(X^2) - (E(X))^2 = \frac{7}{30} - \frac{25}{144} = \frac{43}{720}$$

$$Var(Y) = E(Y^2) - (E(Y))^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$$

$$Korr(X, Y) = \frac{Kov(X, Y)}{\sqrt{Var(X)Var(Y)}} = \frac{\frac{1}{48}}{\sqrt{\frac{43}{720} \cdot \frac{3}{80}}} = \frac{5}{\sqrt{129}} = 0,4402\dots$$

Abschließend folgt noch:

$$E(S) = E(X) + E(Y) = \frac{5}{12} + \frac{3}{4} = \frac{7}{6}$$

$$E(S^2) = \int_0^2 z^2 f_S(z) dz = \int_0^1 z^4 dz + \int_1^2 2z^3 - z^4 dz = \frac{1}{5} + \left(\frac{1}{2} z^4 - \frac{1}{5} z^5 \Big|_1^2 \right) = \frac{3}{2}$$

$$Var(S) = E(S^2) - (E(S))^2 = \frac{5}{36} = \frac{43}{720} + \frac{3}{80} + \frac{1}{24} = Var(X) + Var(Y) + 2Kov(X, Y).$$

II.4. Erzeugende Funktionen

Erzeugende Funktionen sind traditionell ein effektives Hilfsmittel in der Kombinatorik. Sie wurden schon im 18. Jahrhundert von Laplace eingeführt. Hier werden sie uns helfen, eine Reihe wichtiger Ergebnisse zur Berechnung von Momenten oder zur Berechnung der Verteilung von Summen unabhängiger Zufallsvariablen zu beweisen.

Definition 36 (erzeugende Funktionen). Es sei X eine reellwertige Zufallsvariable derart, dass für eine Teilmenge $I \subseteq \mathbb{R}$ der Ausdruck

$$\psi_X(t) := E(e^{tX}), \quad t \in I$$

für alle $t \in I$ endlich ist. Dann heißt die auf I definierte Abbildung ψ_X die *momenterzeugende Funktion* zu X bzw. zu der Verteilung P^X .

Die durch

$$\varphi_X(s) := \psi_X(\ln s) = E(s^X), \quad s \in e^I := \{e^t \mid t \in I\}$$

definierte Funktion heißt die *wahrscheinlichkeitserzeugende Funktion* zu X bzw. zu der Verteilung P^X .

Die momenterzeugende Funktion ψ_X charakterisiert die Verteilung P^X eindeutig, wenn die Menge I ein Intervall $[-\delta, \delta]$ mit $\delta > 0$ enthält. Der zugehörige Beweis wird üblicherweise mit Hilfsmitteln der Fourier-Analyse geführt; wir verweisen deshalb hier auf die einschlägige Literatur zur Wahrscheinlichkeitstheorie (z.B. BILLINGSLEY (1986), Theorem 30.1).

Satz 33. Es sei X eine reellwertige Zufallsvariable derart, dass für eine Teilmenge $I \subseteq \mathbb{R}$ die momenterzeugende Funktion ψ_X existiert. Dann gilt:

- a) Es ist stets $\psi_X(0) = \varphi_X(1) = 1$. Existiert ferner $\psi_X(t)$ für ein $t = t^* > 0$ bzw. $t = t_* < 0$, so auch für alle $t \in [0, t^*]$ bzw. $t \in [t_*, 0]$. Bezeichnet speziell

$$t^+ := \sup \{t \in \mathbb{R} \mid \psi_X(t) < \infty\}, \quad t^- := \inf \{t \in \mathbb{R} \mid \psi_X(t) < \infty\},$$

so existiert $\psi_X(t)$ für alle $t \in (t^-, t^+)$, und $\varphi_X(s)$ existiert für alle $s \in (e^{t^-}, e^{t^+})$ (mit der Konvention $e^{-\infty} = 0$, $e^{\infty} = \infty$). Gilt insbesondere $X \geq 0$ bzw. $X \leq 0$ (mit Wahrscheinlichkeit 1), so ist $t^- = -\infty$ bzw. $t^+ = +\infty$.

II. Grundprinzipien stochastischer Modellierung

- b) Es sei $0 < \delta < \min\{t^+, -t^-\}$. Dann existieren sämtliche Momente $E(|X|^k)$, $k \in \mathbb{N}$, ψ_X ist im Nullpunkt beliebig oft differenzierbar, und es gilt

$$\begin{aligned}\psi_X^{(k)}(0) &= E(X^k), \quad k \in \mathbb{N} \text{ und} \\ \psi_X(t) &= \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k, \quad |t| \leq \delta.\end{aligned}$$

Insbesondere ist

$$E(X) = \psi_X'(0), \quad \text{Var}(X) = \psi_X''(0) - \{\psi_X'(0)\}^2.$$

Ferner ist $\varphi_X(s)$ für $s=1$ differenzierbar, und es gilt

$$\begin{aligned}\varphi_X^{(k)}(1) &= E\left(\prod_{i=0}^{k-1} (X-i)\right), \quad k \in \mathbb{N} \text{ und} \\ \varphi_X(s) &= \sum_{k=0}^{\infty} \frac{E\left(\prod_{i=0}^{k-1} (X-i)\right)}{k!} (s-1)^k, \quad |s-1| \leq 1 - e^{-\delta}.\end{aligned}$$

Insbesondere ist

$$E(X) = \varphi_X'(1), \quad \text{Var}(X) = \varphi_X''(1) + \varphi_X'(1) \{1 - \varphi_X'(1)\}.$$

- c) Gilt $P(X \in \mathbb{Z}^+) = 1$ mit $\mathbb{Z}^+ := \{0, 1, 2, \dots\}$, so lässt sich φ_X fortsetzen, d.h. es existiert $\varphi_X(s) = E(s^X)$ auch für alle $|s| \leq 1$, und es gilt

$$\begin{aligned}\frac{\varphi_X^{(k)}(0)}{k!} &= P(X = k), \quad k \in \mathbb{N} \text{ und} \\ \varphi_X(s) &= \sum_{k=0}^{\infty} P(X = k) s^k, \quad |s| \leq 1.\end{aligned}$$

- d) Sind X und Y stochastisch unabhängige, reellwertige Zufallsvariablen mit momenterzeugenden Funktionen ψ_X und ψ_Y , die beide in derselben Menge $I \subseteq \mathbb{R}$ existieren, so besitzt dort auch die Zufallsvariable $Z = X + Y$ eine momenterzeugende Funktion, und es gilt

$$\psi_{X+Y}(t) = \psi_X(t) \cdot \psi_Y(t), \quad t \in I$$

bzw. auch

$$\varphi_{X+Y}(s) = \varphi_X(s) \cdot \varphi_Y(s), \quad s \in e^I.$$

II. Grundprinzipien stochastischer Modellierung

Beweis:

- a) Die erste Aussage ist wegen $e^0 = 1$ trivial. Mit der *Hölder-Ungleichung* erhält man ferner für $0 \leq t \leq t^*$

$$\psi_X(t) = E(e^{tX}) \leq \left\{ E(e^{t^*X}) \right\}^{t/t^*} = \left\{ \psi_X(t^*) \right\}^{t/t^*}$$

(mit der Wahl $p = \frac{t^*}{t}$ und $Y \equiv 1$), wenn man dort X durch e^{tX} ersetzt. Mit analoger Argumentation folgt für $t_* \leq t \leq 0$

$$\psi_X(t) = E(e^{tX}) = E(e^{(-t)(-X)}) \leq \left\{ E(e^{(-t_*)(-X)}) \right\}^{t/t_*} = \left\{ \psi_X(t_*) \right\}^{t/t_*}.$$

Somit existiert $\psi_X(t)$ für alle $t \in (t^-, t^+)$, also auch $\varphi_X(s)$ für alle $s \in (e^-, e^{t^+})$. Für $X \geq 0$ (mit Wahrscheinlichkeit 1) ist stets $e^{tX} \leq 1$ für $t \leq 0$ (mit Wahrscheinlichkeit 1), d.h. es folgt $t^- = -\infty$. Analog folgt $t^+ = +\infty$, falls $X \leq 0$ (mit Wahrscheinlichkeit 1).

- b) Für $|t| \leq \delta$ seien die Abbildungen $G_n(t, \bullet)$ definiert durch

$$G_n(t; x) := \sum_{k=0}^n \frac{x^k}{k!} t^k, \quad x \in \mathbb{R}, \quad n \in \mathbb{N}.$$

Für $n \in \mathbb{N}$ und $|t| \leq \delta$ ist dann

$$|G_n(t; X)| \leq e^{|tX|} \leq e^{\delta X} + e^{-\delta X} =: Z,$$

d.h. Z ist integrierbar mit $E(Z) = \psi_X(\delta) + \psi_X(-\delta)$. Ferner ist

$$|X|^n \leq \frac{n!}{\delta^n} G_n(\delta; |X|) \leq \frac{n!}{\delta^n} Z \quad \text{für alle } n \in \mathbb{N},$$

d.h. es existieren sämtliche Momente von X . Wegen

$$e^{tX} = \lim_{n \rightarrow \infty} G_n(t; X) \quad \text{für } |t| \leq \delta$$

sind die Voraussetzungen des Satzes von der *majorisierten Konvergenz* erfüllt, also gilt

II. Grundprinzipien stochastischer Modellierung

$$\begin{aligned}\psi_X(t) &= E(e^{tX}) = \lim_{n \rightarrow \infty} E(G_n(t; X)) = \lim_{n \rightarrow \infty} E\left(\sum_{k=0}^n \frac{X^k}{k!} t^k\right) \\ &= \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{E(X^k)}{k!} t^k = \sum_{k=0}^{\infty} \frac{E(X^k)}{k!} t^k, \quad |t| \leq \delta,\end{aligned}$$

wie behauptet. Die Differenzierbarkeit von ψ_X im Nullpunkt ergibt sich sofort durch gliedweise Differentiation dieser Reihe, mit $\psi_X^{(k)}(0) = E(X^k)$, $k \in \mathbb{N}$.

Die letzte Behauptung ergibt sich analog unter Beachtung der Reihenentwicklung (verallgemeinerte binomische Formel)

$$s^x = (1 + (s-1))^x = \sum_{k=0}^{\infty} \frac{\prod_{i=0}^{k-1} (x-i)}{k!} (s-1)^k, \quad |s-1| < 1.^{19}$$

c) Unter der angegebenen Bedingung lässt sich φ_X darstellen als

$$\varphi_X(s) = E(s^X) = \sum_{k=0}^{\infty} s^k P(X = k) = \sum_{k=0}^{\infty} \frac{\varphi_X^{(k)}(0)}{k!} s^k,$$

wobei die Reihe wegen $\sum_{k=0}^{\infty} P(X = k) = 1$ in jedem Fall für $|s| \leq 1$ (absolut) konvergiert. Hieraus folgt die Behauptung.

d) Mit X und Y sind auch die Zufallsvariablen e^{tX} und e^{tY} für alle $t \in \mathbb{R}$ stochastisch unabhängig, woraus sofort

$$\psi_{X+Y}(t) = E(e^{t(X+Y)}) = E(e^{tX} \cdot e^{tY}) = E(e^{tX}) E(e^{tY}) = \psi_X(t) \psi_Y(t), \quad t \in I$$

folgt. Die andere Aussage ergibt sich völlig analog. ■

Bemerkung: Die wahrscheinlichkeitserzeugende Funktion φ_X für diskrete Zufallsvariablen X wird in der Literatur meist direkt über die Beziehung in c) definiert.

¹⁹ Für ganzzahlige x ist die Reihe endlich, da $\prod_{i=0}^{k-1} (x-i) = 0$ ist für $k > x$.

II.5. Grundlegende Verteilungen

Wir listen hier zunächst für viele Anwendungen wichtige Verteilungen tabellarisch auf und kommentieren einige von ihnen danach etwas genauer.

Wir beginnen mit einigen diskreten Verteilungen.

P^X	Name	Zähldichte $f(k) = P(X = k)$	$\varphi_X(s)$	$E(X)$	$Var(X)$
\mathcal{U}_n	Gleichverteilung	$\frac{1}{n}, k = 1, \dots, n \in \mathbb{N}$	$\frac{s \cdot s^n - 1}{n \cdot s - 1}$	$\frac{n+1}{2}$	$\frac{n^2 - 1}{12}$
$B(n, p)$	Binomialverteilung	$\binom{n}{k} p^k (1-p)^{n-k}, k = 0, \dots, n; p \in [0, 1]$	$(1-p + ps)^n$	np	$np(1-p)$
$NB(\beta, p)$	negative Binomialverteilung	$\binom{\beta+k-1}{k} p^\beta (1-p)^k, k \in \mathbb{Z}^+; \beta > 0, p \in (0, 1]$	$\left(\frac{p}{1-(1-p)s}\right)^\beta$	$\beta \frac{1-p}{p}$	$\beta \frac{1-p}{p^2}$
$\mathcal{P}(\lambda)$	Poisson-Verteilung	$e^{-\lambda} \frac{\lambda^k}{k!}, k \in \mathbb{Z}^+; \lambda > 0$	$e^{\lambda(s-1)}$	λ	λ
$\mathcal{LS}(p)$	Log-Series-Verteilung	$\frac{p^k}{L(p) \cdot k}, k \in \mathbb{N}; p \in [0, 1), L(p) := -\ln(1-p)$	$\frac{\ln(1-pt)}{\ln(1-p)}$	$\frac{p}{(1-p)L(p)}$	$\frac{p(L(p)-p)}{(1-p)^2 L^2(p)}$

Man beachte, dass die negative Binomialverteilung $NB(\beta, p)$ auch für nicht-ganzzahlige Parameter $\beta > 0$ definiert ist; es gilt dann

$$\binom{\beta+k-1}{k} = \frac{(\beta+k-1)(\beta+k-2)\cdots\beta}{k!} = \binom{-\beta}{k}, k = 0, 1, 2, \dots$$

Als Spezialfall für $\beta = 1$ ergibt sich hieraus die *geometrische Verteilung* $\mathcal{G}(p)$.

Die hier erwähnte diskrete Gleichverteilung (*Laplace-Verteilung*) über der Menge $\{1, 2, \dots, n\}$ ist ebenfalls ein Spezialfall, vgl. Definition 7.

Die *Binomialverteilung* entsteht bei $n \in \mathbb{N}$ unabhängigen Versuchswiederholungen mit dem jeweiligen Ausgang "Treffer" (entsprechend der 1) oder "Niete" (entsprechend der 0). Zählt man hier die Gesamtzahl aller "Treffer", so ist die entsprechende Zufallsvariable X binomialverteilt, wobei der Parameter $p \in [0, 1]$ der Trefferwahrscheinlichkeit im Einzelversuch entspricht.

Die *negative Binomialverteilung* mit Parametern $p \in (0, 1]$ und $\beta \in \mathbb{N}$ entsteht ebenfalls bei solchen Experimenten (mit nicht vorher festgelegtem Versuchsumfang), wenn man die Anzahl der "Nieten" vor dem β -ten "Treffer" zählt. Potenziell führt dies zu unendlich langen Versuchsserien; im Licht von Satz 27 bedeutet dies, dass man zur korrekten stochastischen Modellierung selbst dieser so einfach anmutenden Situation schon einen *überabzählbaren* Wahrscheinlichkeitsraum benötigt!

Die *Log-Series-Verteilung* verdankt ihren Namen der Taylorreihen-Entwicklung der Funktion

$$L(p) = -\ln(1-p) = \sum_{k=1}^{\infty} \frac{p^k}{k} \text{ für } p \in [0, 1).$$

II. Grundprinzipien stochastischer Modellierung

Bemerkenswert ist bei der geometrischen Verteilung $\mathcal{G}(p)$ ihre so genannte *Gedächtnislosigkeit*:

Lemma 40. Ist X eine $\mathcal{G}(p)$ -verteilte Zufallsvariable mit $0 < p < 1$, so gilt:

$$P(X = m | X \geq n) = P(X = m - n) \text{ für alle } 0 \leq n \leq m \in \mathbb{Z}^+.$$

Beweis: Es ist

$$\begin{aligned} P(X = m | X \geq n) &= \frac{P(X = m)}{P(X \geq n)} = \frac{p(1-p)^m}{\sum_{k=n}^{\infty} p(1-p)^k} = \frac{p(1-p)^m}{(1-p)^n \underbrace{\sum_{j=0}^{\infty} p(1-p)^j}_{=1}} \\ &= p(1-p)^{m-n} = P(X = m - n) \end{aligned}$$

für alle $0 \leq n \leq m \in \mathbb{Z}^+$ (beachte für die Berechnung der bedingten Wahrscheinlichkeit: unter dieser Voraussetzung ist $\{X = m\} \subseteq \{X \geq n\}$, also $\{X = m\} \cap \{X \geq n\} = \{X = m\}$.) ■

Dies bedeutet: interpretiert man – wie oben allgemeiner bei der negativen Binomialverteilung – die Zufallsvariable X als Anzahl der "Nieten" vor dem *ersten* "Treffer", so führt die Information $\{X \geq n\}$, d.h. das Ereignis "in den ersten n Versuchen gab es keinen Treffer" zur gleichen Wahrscheinlichkeit für das (bedingte) Ereignis "vor dem ersten Treffer liegen m Nieten" (d.h. es folgen noch weitere $m - n$ Nieten), wie für das (unbedingte) Ereignis "vor dem ersten Treffer liegen $m - n$ Nieten". Dies zeigt erneut – auf andere Weise –, dass in einer unabhängigen Folge von Versuchen die Vergangenheit "keinen Einfluss" auf die Zukunft hat; die Kenntnis der Vergangenheit ändert die Trefferwahrscheinlichkeit für zukünftige Versuche nicht.

Diese Eigenschaft der Gedächtnislosigkeit ist sogar charakteristisch für die geometrische Verteilung, d.h. man kann zeigen, dass eine Zufallsvariable X mit Werten in \mathbb{Z}^+ , die die Gleichung in Lemma 40 für alle $0 \leq n \leq m \in \mathbb{Z}^+$ erfüllt, notwendig geometrisch verteilt ist.

Die *Poisson-Verteilung* lässt sich in gewisser Weise als "Grenzfall" der Binomial- und der negativen Binomialverteilung auffassen, wie der folgende Satz zeigt.

Satz 34 (Poisson 1837). Es seien $\lambda, \nu > 0$ und $\{p_n\}_{n \in \mathbb{N}}, \{q_n\}_{n \in \mathbb{N}} \subseteq (0,1)$ Folgen reeller Zahlen mit $\lim_{n \rightarrow \infty} n \cdot p_n = \lambda, \lim_{n \rightarrow \infty} n \cdot (1 - q_n) = \nu$. Ferner seien X, Y, X_n, Y_n Zufallsvariablen mit $P^X = \mathcal{P}(\lambda), P^Y = \mathcal{P}(\nu), P^{X_n} = B(n, p_n), P^{Y_n} = NB(n, q_n)$ für alle $n \in \mathbb{N}$. Dann gilt:

$$P(X = k) = \lim_{n \rightarrow \infty} P(X_n = k), \quad P(Y = k) = \lim_{n \rightarrow \infty} P(Y_n = k) \quad \text{für alle } k \in \mathbb{Z}^+.$$

Beweis: Es gilt

$$P(X_n = k) = \binom{n}{k} p_n^k (1 - p_n)^{n-k} = \frac{n!}{(n-k)! k!} \frac{p_n^k (1 - p_n)^{n-k}}{(1 - p_n)^k} = \frac{n!}{(n-k)! k!} \left(1 - \frac{\lambda}{n} \cdot \frac{n p_n}{\lambda}\right)^n \frac{(n p_n)^k}{k!} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

und

II. Grundprinzipien stochastischer Modellierung

$$P(Y_n = k) = \binom{n+k-1}{k} q_n^n (1-q_n)^k = \frac{(n+k-1)!}{(n-1)! n^k} \left(1 + \frac{1}{n} \cdot \frac{n(1-q_n)}{q_n}\right)^{-n} \frac{(n(1-q_n))^k}{k!} \rightarrow e^{-\nu} \frac{\nu^k}{k!}$$

für $n \rightarrow \infty$ und alle $k \in \mathbb{Z}^+$ (beachte: aus der Voraussetzung folgt $\lim_{n \rightarrow \infty} \frac{n(1-q_n)}{q_n} = \nu$). ■

Man kann das Ergebnis des Satzes in vereinfachter Form kurz auch so ausdrücken:

$$B\left(n, \frac{\lambda}{n}\right) \sim \mathcal{P}(\lambda) \quad \text{und} \quad NB\left(n, 1 - \frac{\nu}{n}\right) \sim \mathcal{P}(\nu) \quad \text{für } n \rightarrow \infty.$$

Siméon Denis Poisson (1781 – 1840) publizierte den ersten Sachverhalt, der auch als das *Gesetz der seltenen Ereignisse* bekannt ist, in seinem berühmten Buch *Recherches sur la probabilité des jugements en matières criminelles et matière civile* (1837), wo er unter anderem die Wahrscheinlichkeitsrechnung bei Vaterschaftsprozessen einsetzte.

Die angegebenen Momente dieser Verteilungen lassen sich übrigens sehr leicht direkt aus Satz 33 b) herleiten. Wir zeigen dies exemplarisch am Beispiel der Binomialverteilung: Es ist zunächst

$$\varphi_X(s) = E(s^X) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} s^k = \sum_{k=0}^n \binom{n}{k} (ps)^k (1-p)^{n-k} = (1-p+ps)^n \quad \text{für } s \in \mathbb{R}$$

mit

$$\frac{d}{ds} \varphi_X(s) = np(1-p+ps)^{n-1}, \quad \frac{d^2}{ds^2} \varphi_X(s) = n(n-1)p^2(1-p+ps)^{n-2},$$

also

$$E(X) = \varphi_X'(1) = np, \quad \text{Var}(X) = \varphi_X''(1) + \varphi_X'(1)\{1 - \varphi_X'(1)\} = n(n-1)p^2 + np(1-np) = np(1-p).$$

Die obigen Verteilungen besitzen darüber hinaus folgende sofort einsichtige Faltungseigenschaften:

$$\begin{aligned} B(n, p) * B(m, p) &= B(n+m, p) && [n, m \in \mathbb{N}, 0 < p < 1] \\ NB(\beta, p) * NB(\gamma, p) &= NB(\beta + \gamma, p) && [\beta, \gamma > 0, 0 < p < 1] \\ \mathcal{P}(\lambda) * \mathcal{P}(\mu) &= \mathcal{P}(\lambda + \mu) && [\lambda, \mu > 0] \end{aligned}$$

Dies erkennt man am einfachsten über Satz 33 d) durch Vergleich der entsprechenden wahrscheinlichkeitserzeugenden Funktionen, da die zur Faltung gehörige erzeugende Funktion bei Unabhängigkeit gerade das Produkt der einzelnen erzeugenden Funktionen ist.

II. Grundprinzipien stochastischer Modellierung

Die folgende Tabelle enthält einige wichtige stetige Verteilungen.

p^x	Name	Dichte $f(x)$	$\psi_x(t)$	$E(X)$	$Var(X)$
$\mathcal{U}[a, b]$	Gleichverteilung	$\frac{1}{b-a}, a \leq x \leq b$	$\frac{e^{bt} - e^{at}}{t(b-a)}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
$\mathcal{E}(\lambda)$	Exponentialverteilung	$\lambda e^{-\lambda x}, x \geq 0; \lambda > 0$	$\frac{\lambda}{\lambda - t}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
$\Gamma(\alpha, \lambda)$	Gamma-Verteilung	$\lambda^\alpha \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-\lambda x}, x > 0; \alpha, \lambda > 0$	$\left(\frac{\lambda}{\lambda - t}\right)^\alpha$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$\mathcal{N}(\mu, \sigma^2)$	Normalverteilung	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x, \mu \in \mathbb{R}, \sigma > 0$	$\exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right)$	μ	σ^2

Die Exponentialverteilung $\mathcal{E}(\lambda)$ ist dabei ein Spezialfall der Gamma-Verteilung: es ist nämlich $\mathcal{E}(\lambda) = \Gamma(1, \lambda)$. Für ganzzahlige Werte n von α heißt die $\Gamma(n, \lambda)$ -Verteilung auch *Erlang-Verteilung*; sie ist die Verteilung einer Summe n unabhängiger jeweils $\mathcal{E}(\lambda)$ -verteilter Zufallsvariablen. Sie wurde historisch zum ersten Mal von dem dänischen Mathematiker und Ingenieur *Agner Krarup Erlang* (1878 – 1929) zur Modellierung und Optimierung von Telefonzentralen verwendet und spielt heute in ganz unterschiedlichen Anwendungen (Warteschlangenmodelle, Rechnernetze, Kommunikations- und Sicherheitssysteme, Versicherungsmathematik, Risikotheorie) eine bedeutende Rolle. Seine grundlegende Arbeit *The Theory of Probabilities and Telephone Conversations* stammt aus dem Jahr 1909 (!).

Die Exponentialverteilung ist in gewisser Weise das stetige Gegenstück zur geometrischen Verteilung, weil auch sie durch die *Gedächtnislosigkeit* charakterisiert ist:

Lemma 41. Ist X eine $\mathcal{E}(\lambda)$ -verteilte Zufallsvariable mit $\lambda > 0$, so gilt:

$$P(X > z | X > x) = P(X > z - x) \text{ für alle } 0 \leq x \leq z \in \mathbb{R}^+.$$

Beweis: Es ist

$$P(X > z | X > x) = \frac{P(X > z)}{P(X > x)} = \frac{e^{-\lambda z}}{e^{-\lambda x}} = e^{-\lambda(z-x)} = P(X > z - x) \text{ für alle } 0 \leq x \leq z \in \mathbb{R}^+$$

(beachte für die Berechnung der bedingten Wahrscheinlichkeit: unter dieser Voraussetzung ist $\{X > z\} \subseteq \{X > x\}$, also $\{X > z\} \cap \{X > x\} = \{X > z\}$.) ■

Anschaulich kann man diesen Sachverhalt vielleicht so verdeutlichen: wenn wir annehmen, dass die Dauer eines einzelnen Telefongesprächs eine $\mathcal{E}(\lambda)$ -verteilte Zufallsvariable ist und wir bei Anruf zum Zeitpunkt $x > 0$ ein Besetztzeichen hören (das ist das Ereignis $\{X > x\}$: das Gespräch dauert an), so ist die Wahrscheinlichkeitsverteilung der Restdauer des Gesprächs genau dieselbe wie die der gesamten ursprünglichen Gesprächsdauer!

II. Grundprinzipien stochastischer Modellierung

Auch hier ist die Gedächtnislosigkeit charakteristisch für die Exponentialverteilung. Wir zeigen dies in folgendem

Satz 35. Ist X eine Zufallsvariable mit Werten in \mathbb{R}^+ und einer *stetigen* Verteilungsfunktion F , und gilt

$$P(X > z | X > x) = P(X > z - x) \text{ für alle } 0 \leq x \leq z \in \mathbb{R}^+,$$

so ist notwendig X exponentialverteilt mit einem Parameter $\lambda > 0$.

Beweis: Unter den gemachten Annahmen gilt

$$P(X > z | X > x) = \frac{P(X > z)}{P(X > x)} = \frac{1 - F(z)}{1 - F(x)} = P(X > z - x) = 1 - F(z - x) \text{ für alle } 0 \leq x \leq z \in \mathbb{R}^+.$$

Setzen wir zur Abkürzung $G(x) := 1 - F(x)$, $x \geq 0$, so ist diese Gleichung äquivalent zu

$$G(x + y) = G(x) \cdot G(y) \text{ für alle } x, y \geq 0,$$

mit der Nebenbedingung $G(0) = 1 - F(0) = P(X > 0) = P(X \geq 0) = 1$ (wegen der Stetigkeitsannahme für F ist $P(\{0\}) = 0$.) Induktiv folgt hieraus zunächst allgemeiner

$$G\left(\sum_{i=1}^n x_i\right) = \prod_{i=1}^n G(x_i) \text{ für alle } x_i \geq 0, i = 1, \dots, n, n \in \mathbb{N}. \quad (*)$$

Wir setzen zur Abkürzung $\alpha := G(1) \in [0, 1]$. Aus (*) folgt dann wieder induktiv

$$G(n) = \alpha^n \text{ sowie } \alpha = G(1) = \left\{ G\left(\frac{1}{n}\right) \right\}^n, \text{ also } G\left(\frac{1}{n}\right) = \alpha^{1/n} \text{ für alle } n \in \mathbb{N}.$$

Es folgt

$$G\left(\frac{k}{n}\right) = \left\{ G\left(\frac{1}{n}\right) \right\}^k = \alpha^{k/n} \text{ für alle } k, n \in \mathbb{N}$$

und damit

$$G(x) = \alpha^x \text{ für alle } x \in \mathbb{Q}^+.$$

Wegen der Stetigkeit von F (und damit auch der Stetigkeit von G) heißt das aber gerade:

$$G(x) = \alpha^x \text{ für alle } x \in \mathbb{R}^+.$$

Hierbei ist notwendig $0 < \alpha < 1$, denn $\alpha = 0$ führt zu $F(x) = 1 - G(x) = 1$ für alle $x > 0$, was wegen $F(0) = 0$ zu einem Widerspruch führt, und $\alpha = 1$ bedeutet $F(x) = 1 - G(x) = 0$ für alle $x > 0$, was wegen $\lim_{z \rightarrow \infty} F(z) = 1$ zu einem Widerspruch führt. Mit der Setzung $\lambda := -\ln \alpha > 0$ ergibt dies

$$F(x) = 1 - G(x) = 1 - e^{-\lambda x} \text{ für alle } x \in \mathbb{R}^+,$$

und das war ja gerade zu beweisen. ■

II. Grundprinzipien stochastischer Modellierung

Einen fundamentalen Zusammenhang zwischen geometrischer und Exponentialverteilung beschreibt

Lemma 42. Ist X eine $\mathcal{E}(\lambda)$ -verteilte Zufallsvariable mit $\lambda > 0$, so ist $Y := \lfloor X \rfloor$ geometrisch verteilt mit Parameter $p = 1 - e^{-\lambda}$. Dabei bezeichne $\lfloor z \rfloor := \max \{m \in \mathbb{Z} \mid m \leq z\}$ für alle $z \in \mathbb{R}$ (größte ganze Zahl unterhalb von z , "Gauß-Klammer"). Ist umgekehrt $\{X_n\}_{n \in \mathbb{N}}$ eine Folge geometrisch verteilter Zufallsvariablen mit Parametern $p_n = 1 - e^{-\lambda/n}$, so ist $Y_n := \frac{X_n}{n}$ asymptotisch $\mathcal{E}(\lambda)$ -verteilt, d.h. hier:

$$\lim_{n \rightarrow \infty} P(Y_n \leq x) = 1 - e^{-\lambda x} \quad \text{für alle } x \geq 0.$$

Beweis: Zunächst gilt (mit F_X als Verteilungsfunktion von X):

$$\begin{aligned} P(Y = n) &= P(\lfloor X \rfloor = n) = P(n \leq X < n+1) = F_X(n+1) - F_X(n) \\ &= (1 - e^{-\lambda(n+1)}) - (1 - e^{-\lambda n}) = e^{-\lambda n} (1 - e^{-\lambda}) \end{aligned}$$

für alle $n \in \mathbb{Z}^+$, woraus der erste Teil des Lemmas folgt.

Für den zweiten Teil gilt:

$$P(Y_n > x) = P(X_n > nx) = P(X_n > \lfloor nx \rfloor) = \sum_{k=\lfloor nx \rfloor+1}^{\infty} p_n (1-p_n)^k = (1-p_n)^{\lfloor nx \rfloor+1} = \exp\left(-\frac{\lambda}{n} \cdot (\lfloor nx \rfloor+1)\right)$$

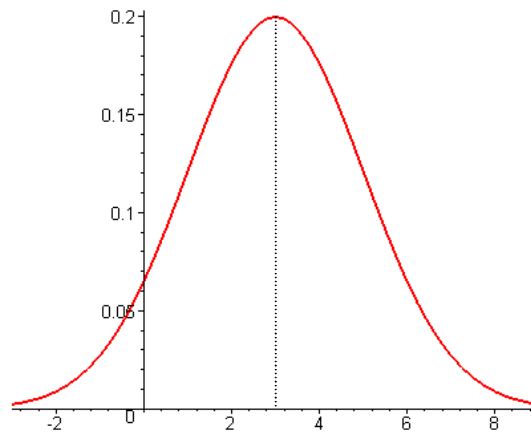
mit Grenzwert $e^{-\lambda x}$ für alle $x \geq 0$, woraus die Behauptung folgt. ■

Die *Normalverteilung* ist vielleicht die bekannteste Wahrscheinlichkeitsverteilung überhaupt; sie wurde von *Carl-Friedrich Gauß* (1777 – 1855) intensiv studiert und unter anderem für die Verteilung von Messfehlern und die lineare Ausgleichsrechnung benutzt. Gauß zu Ehren war die Dichte der Normalverteilung zusammen mit einem (spiegelverkehrten!) Bild von ihm auf dem letzten deutschen 10-DM-Schein abgebildet.



II. Grundprinzipien stochastischer Modellierung

Die Anfänge der Normalverteilung gehen aber schon auf Studien von *Abraham de Moivre* (1667 – 1754) zurück, dem aufgefallen war, dass viele Histogramme aus realen Daten eine symmetrische Glocken-Form besitzen. Er formulierte als erster eine spezielle Version des *Zentralen Grenzwertsatzes* (→ Abschnitt II.7.) für binomialverteilte Zufallsvariablen; ein exakter Beweis dafür gelang aber erst *Pierre Simon Laplace* (1749 – 1827), der sogar 6 Wochen lang unter Napoleon französischer Innenminister war, aber wieder entlassen wurde, weil er "den Geist des unendlich Kleinen in die Regierungsgeschäfte einbrachte". Der sich hierauf beziehende *Lokale Zentrale Grenzwertsatz von de Moivre – Laplace* ist in der Regel auch Gegenstand der Schulmathematik, siehe etwa BARTH UND HALLER (1998), die sogar einen vollständigen Beweis des Satzes bringen.



Dichte der Normalverteilung $\mathcal{N}(3, 2)$

Historisch hat sich eingebürgert, für die Dichte und die Verteilungsfunktion der *Standard-Normalverteilung* $\mathcal{N}(0, 1)$ spezielle Symbole zu verwenden, nämlich

$$\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \quad x \in \mathbb{R} \quad \text{für die Dichte}$$

$$\Phi(x) := \int_{-\infty}^x \varphi(u) du, \quad x \in \mathbb{R} \quad \text{für die Verteilungsfunktion.}$$

Warum der ominöse Faktor $\frac{1}{\sqrt{2\pi}}$ in der Dichte auftritt, lässt sich elegant mit Hilfe der Polarkoordinaten-Transformation zeigen (siehe Satz 26; der Beweis stammt ursprünglich von Gauß; er ist auch in dem Schulbuch von BARTH UND HALLER (1998) explizit enthalten): es gilt nämlich unter Verwendung des Satzes von Fubini:

$$\begin{aligned} \left(\int_{-\infty}^{\infty} e^{-z^2/2} dz \right)^2 &= \left(\int_{-\infty}^{\infty} e^{-x^2/2} dx \right) \cdot \left(\int_{-\infty}^{\infty} e^{-y^2/2} dy \right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-(x^2+y^2)/2} dx dy = \int_0^{2\pi} \int_0^{\infty} e^{-r^2/2} r dr d\phi \\ &\quad \uparrow \\ &\quad \begin{matrix} x=r \cdot \cos\phi \\ y=r \cdot \sin\phi \end{matrix} \\ &= 2\pi \int_0^{\infty} e^{-r^2/2} r dr = -2\pi \cdot e^{-r^2/2} \Big|_0^{\infty} = 2\pi, \quad \text{also} \quad \int_{-\infty}^{\infty} e^{-z^2/2} dz = \sqrt{2\pi}. \end{aligned}$$

II. Grundprinzipien stochastischer Modellierung

Lemma 43. Ist X eine $\mathcal{N}(0,1)$ -verteilte Zufallsvariable, so ist $Y := \mu + \sigma X$ mit $\mu \in \mathbb{R}$, $\sigma > 0$ $\mathcal{N}(\mu, \sigma^2)$ -verteilt. Ist umgekehrt Y $\mathcal{N}(\mu, \sigma^2)$ -verteilt, so ist $X := \frac{Y - \mu}{\sigma}$ $\mathcal{N}(0,1)$ -verteilt.

Beweis: Es gilt im ersten Teil

$$P(Y \leq x) = P(\mu + \sigma X \leq x) = P\left(X \leq \frac{x - \mu}{\sigma}\right) = \Phi\left(\frac{x - \mu}{\sigma}\right) = \int_{-\infty}^{\frac{x - \mu}{\sigma}} \varphi(u) du = \frac{1}{\sigma} \int_{-\infty}^x \varphi\left(\frac{v - \mu}{\sigma}\right) du$$

\uparrow
 $v = \mu + \sigma u$
 $u = \frac{v - \mu}{\sigma}$

für alle $x \in \mathbb{R}$, woraus durch Differenzieren die Behauptung folgt. Der zweite Teil ergibt sich analog. ■

Die für beliebige Zufallsvariablen Y mit reellem Erwartungswert $E(Y) = \mu$ und Varianz $Var(Y) = \sigma^2 > 0$ stets mögliche Transformation $X := \frac{Y - \mu}{\sigma}$ nennt man in der Statistik übrigens *z-Standardisierung*. Sie führt dazu, dass danach $E(X) = 0$ und $Var(X) = 1$ gilt.

Wir wollen hier exemplarisch noch die momenterzeugende Funktion der Normalverteilung herleiten:

Lemma 44. Für eine $\mathcal{N}(\mu, \sigma^2)$ -verteilte Zufallsvariable X gilt:

$$\psi_X(t) = \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right), \quad t \in \mathbb{R}; \quad y \in \mathbb{R}, \quad \sigma > 0.$$

Beweis: Wir nehmen zunächst $\mu = 0$, $\sigma = 1$ an. Dann gilt:

$$\psi_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} \cdot \varphi(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} \cdot e^{-x^2/2} dx = e^{t^2/2} \int_{-\infty}^{\infty} \underbrace{\frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}}_{\text{Dichte der } N(t,1)\text{-Verteilung}} dx = e^{t^2/2}, \quad t \in \mathbb{R}.$$

Für den allgemeinen Fall können wir $Y = \mu + \sigma X$ mit $\mu \in \mathbb{R}$, $\sigma > 0$ annehmen, mit standard-normalverteiltem X . Dann gilt:

$$\psi_Y(t) = E(e^{tY}) = E(e^{t(\mu + \sigma X)}) = e^{t\mu} \cdot E(e^{\sigma t X}) = e^{t\mu} \cdot \psi_X(\sigma t), \quad t \in \mathbb{R},$$

woraus die Behauptung folgt. ■

II. Grundprinzipien stochastischer Modellierung

Lemma 44 kann wieder elegant benutzt werden, um mit Hilfe von Satz 33 d) Erwartungswert und Varianz einer $\mathcal{N}(\mu, \sigma^2)$ -verteilten Zufallsvariablen X herzuleiten:

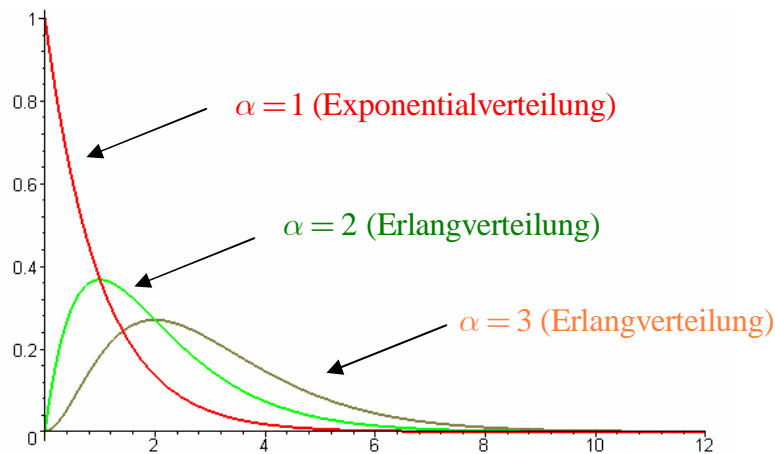
$$E(X) = \psi'_X(0) = \frac{d}{dt} \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) \Big|_{t=0} = (\mu + \sigma^2 t) \cdot \psi_X(t) \Big|_{t=0} = \mu,$$

$$\text{Var}(X) = \psi''_X(0) - \{\psi'_X(0)\}^2 = \frac{d^2}{dt^2} \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) \Big|_{t=0} - \mu^2 = (\sigma^2 + (\mu + \sigma^2 t)^2) \cdot \psi_X(t) \Big|_{t=0} - \mu^2 = \sigma^2.$$

Ähnlich wie bei den obigen diskreten Verteilungen besitzen auch einige der hier genannten stetigen Verteilungen folgende sofort einsichtige Faltungseigenschaften:

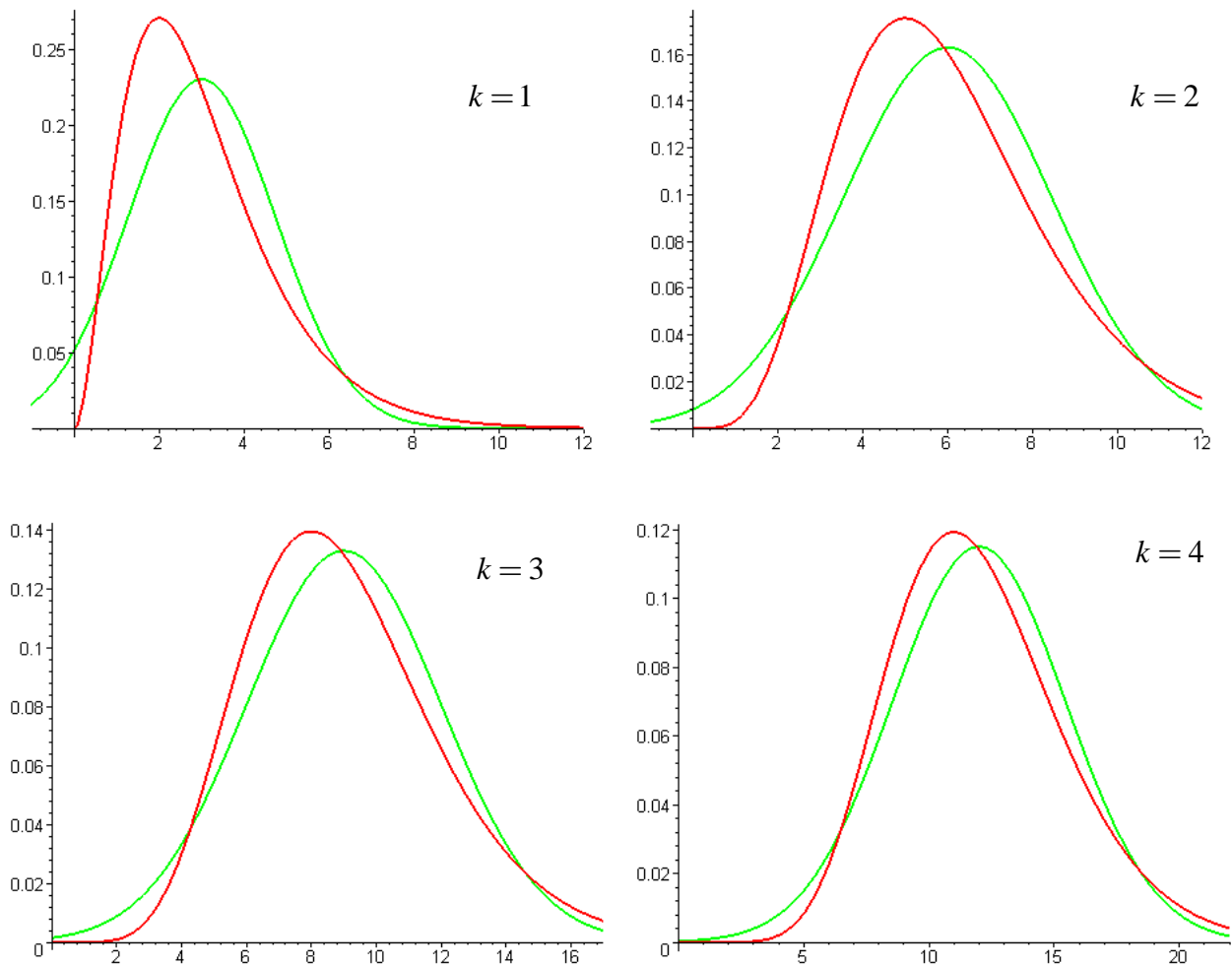
$$\begin{aligned} \Gamma(\alpha, \lambda) * \Gamma(\delta, \lambda) &= \Gamma(\alpha + \delta, \lambda) & [\alpha, \delta, \lambda > 0] \\ \mathcal{N}(\mu, \sigma^2) * \mathcal{N}(\nu, \tau^2) &= \mathcal{N}(\mu + \nu, \sigma^2 + \tau^2) & [\mu, \nu \in \mathbb{R}, \sigma^2, \tau^2 > 0]. \end{aligned}$$

Dies erkennt man am einfachsten wieder über Satz 33 d) durch Vergleich der entsprechenden momenterzeugenden Funktionen, da die zur Faltung gehörige erzeugende Funktion bei Unabhängigkeit auch hier das Produkt der einzelnen erzeugenden Funktionen ist.



Dichten verschiedener $\Gamma(\alpha, 1)$ -Verteilungen

II. Grundprinzipien stochastischer Modellierung



Im Vergleich: Dichten der $\Gamma(3k, 1)$ -Verteilungen (rot)
mit Dichten der $\mathcal{N}(3k, \sqrt{3k})$ -Verteilungen (grün) für $k = 1, 2, 3, 4$

Die auffallende Ähnlichkeit der Dichten mit wachsendem k ist nicht zufällig, sondern ist durch den *Zentralen Grenzwertsatz* begründet, der in Abschnitt II.7. genauer besprochen wird.

Wir setzen die tabellarische Auflistung wichtiger stetiger Wahrscheinlichkeitsverteilungen fort.

P^X	Name	Dichte $f(x)$, $x \in \mathbb{R}$	$E(X)$	$Var(X)$
$\chi_n^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$	χ^2 -Verteilung ($n \in \mathbb{N}$)	$\frac{x^{(n-2)/2}}{\sqrt{2^n} \Gamma(n/2)} e^{-x/2}$, $x > 0$	n	$2n$
$\mathcal{B}(\alpha, \beta)$	Beta-Verteilung ($\alpha, \beta > 0$)	$\frac{x^{\alpha-1} (1-x)^{\beta-1}}{B(\alpha, \beta)}$, $0 \leq x \leq 1$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$
$\mathcal{LN}(\mu, \sigma^2)$	Log-Normalverteilung ($\mu \in \mathbb{R}$, $\sigma > 0$)	$\frac{1}{x\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right)$, $x > 0$	$e^{\mu + \sigma^2/2}$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$

Bemerkung: die Beta-Funktion ist definiert durch $B(\alpha, \beta) = \frac{\Gamma(\alpha) \cdot \Gamma(\beta)}{\Gamma(\alpha + \beta)}$ für $\alpha, \beta > 0$.

II. Grundprinzipien stochastischer Modellierung

\mathcal{P}^x	Name	Dichte $f(x)$, $x \in \mathbb{R}$	$E(X)$	$Var(X)$
\mathcal{L}	Logistische Verteilung	$\frac{e^{-x}}{(1+e^{-x})^2}$	0	$\frac{\pi^2}{3}$
\mathcal{LL}	Log-Logistische Verteilung ($\alpha > 0$)	$\frac{\alpha x^{\alpha-1}}{(1+x^\alpha)^2}, x > 0$	$\frac{\pi}{\alpha \sin \frac{\pi}{\alpha}}$	$\frac{\pi \left(\sin \frac{\pi}{\alpha} - \frac{\pi}{\alpha} \cos \frac{\pi}{\alpha} \right)}{\alpha \sin^3 \frac{\pi}{\alpha} \cos \frac{\pi}{\alpha}}$
\mathcal{C}	Cauchy-Verteilung	$\frac{1}{\pi} \frac{1}{1+x^2}$	–	–
$\mathcal{IN}(\mu, \sigma^2)$	Inverse Gauß-Verteilung ($\mu, \sigma > 0$)	$\frac{1}{\sqrt{2\pi\sigma^2 x^3}} \exp\left(-\frac{(x-\mu)^2}{2\mu^2\sigma^2 x}\right), x > 0$	μ	$\mu^3\sigma^2$
$\mathcal{IG}(\alpha)$	Inverse Gamma-Verteilung ($\alpha > 0$)	$\frac{e^{-1/x}}{x^{\alpha+1}\Gamma(\alpha)}, x > 0$	$\frac{1}{\alpha-1}$	$\frac{1}{(\alpha-1)^2(\alpha-2)}$
$\mathcal{Pa}(\alpha)$	Pareto-Verteilung ($\alpha > 0$)	$\frac{\alpha}{(1+x)^{\alpha+1}}, x > 0$	$\frac{1}{\alpha-1}$	$\frac{\alpha}{(\alpha-1)^2(\alpha-2)}$
\mathcal{G}	Gumbel-Verteilung	$e^{-x} \cdot \exp(-e^{-x})$	γ	$\frac{\pi^2}{6}$
$\mathcal{F}(\alpha)$	Fréchet-Verteilung ($\alpha > 0$)	$\frac{\alpha}{x^{\alpha+1}} \exp(-x^{-\alpha}), x > 0$	$\Gamma\left(1-\frac{1}{\alpha}\right)$	$\Gamma\left(1-\frac{2}{\alpha}\right) - \Gamma^2\left(1-\frac{1}{\alpha}\right)$
$\mathcal{W}(\alpha)$	Weibull-Verteilung ($\alpha > 0$)	$\alpha x^{\alpha-1} \exp(-x^\alpha), x > 0$	$\Gamma\left(1+\frac{1}{\alpha}\right)$	$\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)$

Bemerkung: Bei der Gumbel-Verteilung bezeichnet γ die *Euler'sche Konstante*, die durch

$$\gamma := \lim_{n \rightarrow \infty} \left(\sum_{k=1}^n \frac{1}{k} - \ln n \right) \approx 0,577216$$

definiert ist.

Für einige der genannten Verteilungen existieren auch explizite Darstellungen der Verteilungsfunktionen:

II. Grundprinzipien stochastischer Modellierung

P^X	Name	Verteilungsfunktion $F(x)$, $x \in \mathbb{R}$
\mathcal{L}	Logistische Verteilung	$\frac{1}{1+e^{-x}}$
\mathcal{LL}	Log-Logistische Verteilung	$\frac{1}{1+x^{-\alpha}}$, $x > 0$ ($\alpha > 0$)
\mathcal{C}	Cauchy-Verteilung	$\frac{1}{\pi} \arctan(x) + \frac{1}{2}$
$\mathcal{Pa}(\alpha)$	Pareto-Verteilung	$1 - \frac{1}{(1+x)^\alpha}$, $x > 0$ ($\alpha > 0$)
\mathcal{G}	Gumbel-Verteilung	$\exp(-e^{-x})$
$\mathcal{F}(\alpha)$	Fréchet-Verteilung	$\exp(-x^{-\alpha})$, $x > 0$ ($\alpha > 0$)
$\mathcal{W}(\alpha)$	Weibull-Verteilung	$1 - \exp(-x^\alpha)$, $x > 0$ ($\alpha > 0$)

Einige der genannten Verteilungen entstehen durch einfache Transformation von X :

P^X	Transformation	P^Y	P^X	Transformation	P^Y
$\mathcal{U}[0,1]$	$Y = a + (b-a)X$	$\mathcal{U}[a,b]$	$\mathcal{U}[0,1]$	$Y = X^{-1/\alpha} - 1$	$\mathcal{Pa}(\alpha)$
$\mathcal{U}[0,1]$	$Y = -\frac{1}{\lambda} \ln(X)$	$\mathcal{E}(\lambda)$	$\mathcal{U}[0,1]$	$Y = -\ln(-\ln X)$	\mathcal{G}
$\mathcal{N}(\mu, \sigma^2)$	$Y = e^X$	$\mathcal{LN}(\mu, \sigma^2)$	$\mathcal{E}(1)$	$Y = -\ln(X)$	\mathcal{G}
$\mathcal{U}[0,1]$	$Y = \ln\left(\frac{X}{1-X}\right)$	\mathcal{L}	\mathcal{G}	$Y = e^{X/\alpha}$	$\mathcal{F}(\alpha)$
\mathcal{L}	$Y = e^{X/\alpha}$	\mathcal{LL}	$\mathcal{F}(\alpha)$	$Y = \frac{1}{X}$	$\mathcal{W}(\alpha)$
$\mathcal{U}[0,1]$	$Y = -\cot(\pi X)$	\mathcal{C}	$\mathcal{E}(1)$	$Y = X^{1/\alpha}$	$\mathcal{W}(\alpha)$
$\Gamma(\alpha, 1)$	$Y = \frac{1}{X}$	$\mathcal{IG}(\alpha)$	$\mathcal{W}(\alpha)$	$Y = \frac{1}{X}$	$\mathcal{F}(\alpha)$

Allgemeine Berechnung von Dichten bei Transformation (z.B. über Satz 26 oder direkt mit der Verteilungsfunktion und anschließender Differentiation):

Dichte von P^X	Bereich	Transformation	Dichte von P^Y	Bereich
$f(x)$	$x > 0$	$Y = \frac{1}{X}$	$\frac{1}{x^2} f\left(\frac{1}{x}\right)$	$x > 0$
$f(x)$	$x > 0$	$Y = X^\alpha$, $\alpha > 0$	$\frac{1}{\alpha} x^{1/\alpha-1} f(x^{1/\alpha})$	$x > 0$
$f(x)$	$x > 0$	$Y = \ln(X)$	$e^x \cdot f(e^x)$	$x \in \mathbb{R}$
$f(x)$	$x \in \mathbb{R}$	$Y = e^X$	$\frac{1}{x} f(\ln x)$	$x > 0$

II.6. Das Gesetz der Großen Zahlen

Gegenstand dieses und des nächsten Abschnitts ist das asymptotische Verhalten des arithmetischen Mittels unabhängiger oder paarweise unkorrelierter Zufallsvariablen bzw. als Spezialfall hiervon der relativen Häufigkeiten des Eintretens eines Ereignisses in einer Folge unabhängiger Versuchswiederholungen. Gerade das letzte Ergebnis ist das zentrale Bindeglied zwischen "Theorie" und "Praxis", weil es erlaubt, Wahrscheinlichkeiten für Ereignisse aus unabhängigen Beobachtungen zu schätzen, wobei die Präzision der Schätzung mit wachsendem Beobachtungsumfang zunimmt.

Da es sich hier um asymptotische Aussagen handelt, sind vorab einige Konvergenzbegriffe, die in der Stochastik verwendet werden, zu klären.

Definition 37 (Stochastische Konvergenz): Eine Folge $\{X_n\}_{n \in \mathbb{N}}$ reeller Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) heißt *stochastisch konvergent* gegen eine reelle Zufallsvariable X , wenn gilt

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0 \quad \text{für alle } \varepsilon > 0,$$

in Zeichen: $P\text{-}\lim_{n \rightarrow \infty} X_n = X$ oder $X_n \xrightarrow{P} X$ ($n \rightarrow \infty$). Die Folge heißt *gleichmäßig stochastisch konvergent* gegen eine reelle Zufallsvariable X , wenn gilt

$$\lim_{n \rightarrow \infty} P\left(\sup_{m \geq n} |X_m - X| > \varepsilon\right) = 0 \quad \text{für alle } \varepsilon > 0.$$

Man macht sich leicht klar, dass im Falle stochastischer Konvergenz die Grenzzufallsvariable X fast sicher eindeutig bestimmt ist, d.h. ist Y eine weitere reelle Zufallsvariable mit $P\text{-}\lim_{n \rightarrow \infty} X_n = Y$, so ist notwendig $P(X = Y) = 1$, also $X = Y$ fast sicher. Für beliebiges $\varepsilon > 0$ und $n \in \mathbb{N}$ ist nämlich

$$\begin{aligned} P(|X - Y| > \varepsilon) &= P(|(X - X_n) - (Y - X_n)| > \varepsilon) \leq P\left(\left\{|X - X_n| > \frac{\varepsilon}{2}\right\} \cup \left\{|Y - X_n| > \frac{\varepsilon}{2}\right\}\right) \\ &\leq P\left(\left\{|X - X_n| > \frac{\varepsilon}{2}\right\}\right) + P\left(\left\{|Y - X_n| > \frac{\varepsilon}{2}\right\}\right); \end{aligned}$$

die rechte Seite strebt nach Voraussetzung aber gegen 0, so dass $P(|X - Y| > \varepsilon) = 0$ gilt für alle $\varepsilon > 0$, woraus die Behauptung folgt.

II. Grundprinzipien stochastischer Modellierung

Definition 38 (fast sichere Konvergenz): Eine Folge $\{X_n\}_{n \in \mathbb{N}}$ reeller Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) heißt *fast sicher konvergent* gegen eine reelle Zufallsvariable X , wenn gilt

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = P\left(\bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{|X_m - X| \leq \frac{1}{k}\right\}\right) = 1,$$

in Zeichen: $\lim_{n \rightarrow \infty} X_n = X$ f.s. oder $X_n \xrightarrow{\text{f.s.}} X$ ($n \rightarrow \infty$).

Auch hier ist (unmittelbar) klar, dass im Falle fast sicherer Konvergenz die Grenzzufallsvariable X fast sicher eindeutig bestimmt ist.

Um den Zusammenhang zwischen beiden Konvergenzarten darstellen zu können, benötigen wir noch folgende Begriffsbildung.

Definition 39: Es sei (Ω, \mathcal{A}) ein Messraum und $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$ eine beliebige Folge von Ereignissen. Dann heißt

$$\begin{aligned} \limsup_{n \rightarrow \infty} A_n &:= \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m \quad \text{der Limes Superior der Mengenfolge,} \\ \liminf_{n \rightarrow \infty} A_n &:= \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} A_m \quad \text{der Limes Inferior der Mengenfolge.} \end{aligned}$$

Anschaulich gesprochen enthält $\limsup_{n \rightarrow \infty} A_n$ genau diejenigen $\omega \in \Omega$, die in *unendlich vielen* dieser Ereignisse liegen, und $\liminf_{n \rightarrow \infty} A_n$ genau diejenigen $\omega \in \Omega$, die in *fast allen* dieser Ereignisse liegen.

Beispiel: Wir betrachten den Borel-Raum $(\Omega, \mathcal{A}) = (\mathbb{R}^1, \mathcal{B}^1)$ und darin die Folge

$$A_n := \begin{cases} \left[-\frac{1}{n}, n\right] & \text{falls } n \text{ ungerade} \\ \left[-n, \frac{1}{n}\right] & \text{falls } n \text{ gerade} \end{cases} \quad \text{für } n \in \mathbb{N}.$$

Dann gilt $\bigcup_{m=n}^{\infty} A_m = \mathbb{R}$ und $\bigcap_{m=n}^{\infty} A_m = \{0\}$ für alle $n \in \mathbb{N}$, also $\limsup_{n \rightarrow \infty} A_n = \mathbb{R}$ und $\liminf_{n \rightarrow \infty} A_n = \{0\}$.

Den angekündigte Zusammenhang zwischen stochastischer und fast sicherer Konvergenz behandelt

II. Grundprinzipien stochastischer Modellierung

Satz 36: Es sei $\{X, X_n\}_{n \in \mathbb{N}}$ eine Folge reeller Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Dann gelten folgende Aussagen:

a) $\lim_{n \rightarrow \infty} X_n = X$ f.s. $\Rightarrow P\text{-}\lim_{n \rightarrow \infty} X_n = X$,

d.h. *fast sichere* impliziert *stochastische* Konvergenz.

b) $P\text{-}\lim_{n \rightarrow \infty} \sup_{m \geq n} |X_m - X| = 0 \Rightarrow \lim_{n \rightarrow \infty} X_n = X$ f.s.,

d.h. *gleichmäßige stochastische* Konvergenz impliziert *fast sichere* Konvergenz.

Den letzten Sachverhalt kann man auch so ausdrücken: *fast sichere* und *gleichmäßige stochastische* Konvergenz sind äquivalent.

Beweis:

a) Für jedes $\varepsilon > 0$ ergibt sich folgende Ungleichungskette:

$$\begin{aligned} 0 &\leq \lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) \leq \lim_{n \rightarrow \infty} P\left(\bigcup_{m=n}^{\infty} \{|X_m - X| > \varepsilon\}\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|X_m - X| > \varepsilon\}\right) \\ &= P\left(\limsup_{n \rightarrow \infty} \{|X_n - X| > \varepsilon\}\right) = P(|X_n - X| > \varepsilon \text{ für unendlich viele } n) \leq P\left(\left\{\lim_{n \rightarrow \infty} X_n = X\right\}^c\right) = 0 \end{aligned}$$

und damit $\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$, was gerade zu zeigen war.

b) Es bezeichne $D_n := \sup_{m \geq n} |X_m - X|$. Dann ist die Folge $\{D_n\}_{n \in \mathbb{N}}$ schwach monoton fallend und nach unten durch 0 beschränkt, also punktweise (und damit insbesondere fast sicher) konvergent gegen eine endliche Zufallsvariable $D \geq 0$. Nach Voraussetzung und Teil a) des Satzes ist dann $D = P\text{-}\lim_{n \rightarrow \infty} D_n = 0$ f.s., womit alles gezeigt ist. ■

Allgemeiner haben wir damit sogar gezeigt, dass für eine punktweise *monotone* Folge $\{X_n\}_{n \in \mathbb{N}}$ (wachsend oder fallend) stochastische und fast sichere Konvergenz gegen X äquivalent sind.

Das folgende Beispiel zeigt, dass die Aussage aus Teil a) des letzten Satzes nicht ohne weiteres umkehrbar ist.

II. Grundprinzipien stochastischer Modellierung

Beispiel: Es sei $(\Omega, \mathcal{A}) = (\mathbb{R}^1, \mathcal{B}^1)$ und P die stetige Gleichverteilung $\mathcal{U}[0,1]$ über dem Intervall $[0,1]$. Für die durch

$$X_{n^2+k} := \mathbb{1}_{\left[\frac{k-1}{2n+1}, \frac{k}{2n+1}\right]} \quad \text{für } 1 \leq k \leq 2n+1, \quad n \in \mathbb{Z}^+$$

definierten Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ gilt dann

$$P(|X_n| > \varepsilon) = P(|X_n| = 1) \leq \frac{1}{2\sqrt{n}-1} \quad \text{für alle } \varepsilon > 0 \text{ und } n \in \mathbb{N},$$

d.h. es gilt $P\text{-}\lim_{n \rightarrow \infty} X_n = X$, aber die Folge $\{X_n\}_{n \in \mathbb{N}}$ konvergiert nicht fast sicher, denn es gilt

$$\limsup_{n \rightarrow \infty} X_n(\omega) = 1, \quad \liminf_{n \rightarrow \infty} X_n(\omega) = 0 \quad \text{für alle } \omega \in (0,1).$$

(Man mache sich anschaulich klar, dass bei dieser Folge das "Rechteck" $\mathbb{1}_{\left[\frac{k-1}{2n+1}, \frac{k}{2n+1}\right]}$ immer wieder über das Einheitsintervall "rutscht", wobei die Breite immer kleiner wird.)

Wir sind jetzt in der Lage, das so genannte *schwache Gesetz der Großen Zahlen* zu formulieren und zu beweisen.

Satz 37 (schwaches Gesetz der Großen Zahlen): Es sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge paarweise unkorrelierter quadratisch integrierbarer Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit

$$E(X_n) = \mu \in \mathbb{R} \quad \text{für alle } n \in \mathbb{N} \quad \text{und} \quad \lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) = 0.$$

Dann gilt für die arithmetischen Mittel $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$ dieser Folge:

$$P\text{-}\lim_{n \rightarrow \infty} \bar{X}_n = \mu.$$

Beweis: Es ist $E(\bar{X}_n) = \frac{1}{n} \sum_{k=1}^n E(X_k) = \mu$ für alle $n \in \mathbb{N}$ sowie, nach Lemma 37 b) bzw. der Bemerkung auf S. 111

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{k=1}^n \text{Var}(X_k) \quad \text{für alle } n \in \mathbb{N}.$$

Eine einfache Anwendung der Tschebyscheff-Ungleichung (Lemma 37 c)) ergibt nun sofort:

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \sum_{k=1}^n \text{Var}(X_k) \rightarrow 0 \quad \text{für } n \rightarrow \infty,$$

womit der Satz bewiesen ist. ■

II. Grundprinzipien stochastischer Modellierung

Bemerkung: Ein einfacher Spezialfall des schwachen Gesetzes der Großen Zahlen ist gegeben, wenn

$$\text{Var}(X_n) = \sigma^2 \text{ für alle } n \in \mathbb{N}$$

gilt (oder allgemeiner: die Varianzen nach oben beschränkt sind), also insbesondere, wenn alle X_n dieselbe Verteilung besitzen. Man sagt dann, dass die X_n identisch verteilt sind.

Wählt man für die Zufallsvariablen noch spezieller *Indikatorfunktionen*, so erhält man eine Version des Gesetzes der Großen Zahlen, die schon von dem schweizerischen Mathematiker und Physiker *Jakob Bernoulli* (1654 – 1705) vermutlich um das Jahr 1690 herum gefunden und posthum in seiner berühmten *Ars Conjectandi* 1713 veröffentlicht wurde.

Lemma 45: Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $A \in \mathcal{A}$ ein Ereignis mit Eintrittswahrscheinlichkeit $p := P(A) \in (0, 1)$. Betrachtet wird ein Experiment mit unabhängiger Versuchswiederholung, d.h. der Produkt-Wahrscheinlichkeitsraum $(\Omega^*, \mathcal{A}^*, P^*) = \left(\prod_{n \in \mathbb{N}} \Omega, \otimes_{n \in \mathbb{N}} \mathcal{A}, \otimes_{n \in \mathbb{N}} P \right)$ mit der stochastisch unabhängigen Ereignisfolge $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}^*$, definiert durch

$$A_n := \prod_{i=1}^{n-1} \Omega \times A \times \prod_{i=n+1}^{\infty} \Omega \text{ für alle } n \in \mathbb{N},$$

d.h. A_n beschreibt das Eintreten von A im n -ten Experiment. Ferner sei $X_n := \mathbb{1}_{A_n}$ für alle $n \in \mathbb{N}$.

Dann beschreibt $H_n := \frac{1}{n} \sum_{k=1}^n X_k$ die *relative Häufigkeit* des Eintretens des Ereignisses A in den ersten n Experimenten, und es gilt

$$P\text{-}\lim_{n \rightarrow \infty} H_n = p.$$

Beweis: Folgt direkt aus Satz 37 wegen $E(X_n) = p$ und $\text{Var}(X_n) = p(1-p)$ für alle $n \in \mathbb{N}$. (Beachte: X_n ist $B(1, p)$ -verteilt für alle $n \in \mathbb{N}$.) ■

Diese Version des Gesetzes der Großen Zahlen (allerdings ohne den strengen mathematischen Formalismus) findet sich auch in praktisch allen Schulbüchern zur Stochastik der gymnasialen Oberstufe einschließlich des Beweises über die Tschebyscheff-Ungleichung, vgl. etwa BARTH UND HALLER (1998), Satz 249.1.

Das schwache Gesetz der Großen Zahlen gilt in ähnlicher Form auch mit fast sicherer Konvergenz. Zum Beweis benötigen wir allerdings das folgende Hilfsresultat.

II. Grundprinzipien stochastischer Modellierung

Lemma 46 (Borel-Cantelli): Es sei $\{A_n\}_{n \in \mathbb{N}}$ eine Folge von Ereignissen in einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Dann gilt:

a) $\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P\left(\limsup_{n \rightarrow \infty} A_n\right) = 0.$

b) Ist die Folge $\{A_n\}_{n \in \mathbb{N}}$ zusätzlich stochastisch unabhängig, so gilt

$$\sum_{n=1}^{\infty} P(A_n) = \infty \Rightarrow P\left(\limsup_{n \rightarrow \infty} A_n\right) = 1.$$

Beweis:

a) Nach Satz 5 gilt $P\left(\bigcup_{k=n}^m A_k\right) \leq \sum_{k=n}^m P(A_k) \leq \sum_{k=n}^{\infty} P(A_k)$ für alle $n \leq m$, $n, m \in \mathbb{N}$ und somit aus

Monotoniegründen auch $P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k)$ für alle $n \in \mathbb{N}$. Damit ist

$$P\left(\limsup_{n \rightarrow \infty} A_n\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} A_m\right) = \lim_{n \rightarrow \infty} P\left(\bigcup_{m=n}^{\infty} A_m\right) \leq \lim_{n \rightarrow \infty} \sum_{k=n}^{\infty} P(A_k) = 0$$

wegen der vorausgesetzten Konvergenz der Reihe.

b) Es ist

$$\begin{aligned} P\left(\limsup_{n \rightarrow \infty} A_n\right) &= 1 - P\left(\liminf_{n \rightarrow \infty} A_n^c\right) = 1 - P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) = 1 - \lim_{n \rightarrow \infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \\ &= 1 - \lim_{n \rightarrow \infty} \prod_{k=n}^{\infty} (1 - P(A_k)). \end{aligned}$$

Dieser Ausdruck ist sicher 1, wenn $\limsup_{n \rightarrow \infty} P(A_n) = 1$ ist. Im anderen Fall $\limsup_{n \rightarrow \infty} P(A_n) < 1$ gilt für genügend große $n \in \mathbb{N}$

$$\ln \left[\prod_{k=n}^m (1 - P(A_k)) \right] = \sum_{k=n}^m \ln(1 - P(A_k)) \leq -\sum_{k=n}^m P(A_k) \text{ für } n \leq m \in \mathbb{N},$$

woraus nach Grenzübergang

$$\prod_{k=n}^{\infty} (1 - P(A_k)) = \lim_{m \rightarrow \infty} \exp \left(\ln \left[\prod_{k=n}^m (1 - P(A_k)) \right] \right) = 0$$

und damit die Aussage folgt. ■

II. Grundprinzipien stochastischer Modellierung

Satz 38 (starkes Gesetz der Großen Zahlen; Etemadi 1981): Es sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge paarweise unkorrelierter quadratisch integrierbarer Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit

$$E(X_n) = \mu \in \mathbb{R} \text{ für alle } n \in \mathbb{N} \text{ und } V := \sup_{n \in \mathbb{N}} \{Var(X_n)\} < \infty.$$

Dann gilt für die arithmetischen Mittel $\bar{X}_n := \frac{1}{n} \sum_{k=1}^n X_k$ dieser Folge:

$$\lim_{n \rightarrow \infty} \bar{X}_n = \mu \text{ fast sicher.}$$

Beweis: Wir können o.B.d.A. für die Erwartungswerte $E(X_n) = 0$ annehmen, sonst gehen wir zu den Zufallsvariablen $\{X_n - \mu\}_{n \in \mathbb{N}}$ über, die diese Eigenschaft besitzen. Wir zeigen unter dieser Annahme zunächst, dass $\lim_{n \rightarrow \infty} \bar{X}_{n^2} = 0$ fast sicher gilt. Mit der Tschebyscheff-Ungleichung (Lemma 37c)) erhalten wir für beliebiges $\varepsilon > 0$

$$P(|\bar{X}_{n^2}| > \varepsilon) \leq \frac{Var(\bar{X}_{n^2})}{\varepsilon^2} \leq \frac{V}{n^2 \varepsilon^2} \text{ für alle } n \in \mathbb{N}.$$

Dies impliziert $\sum_{n=1}^{\infty} P(|\bar{X}_{n^2}| > \varepsilon) < \infty$ und damit nach Lemma 46 a)

$$P\left(\limsup_{n \rightarrow \infty} \{|\bar{X}_{n^2}| > \varepsilon\}\right) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \{|\bar{X}_{m^2}| > \varepsilon\}\right) = 0.$$

Daraus folgt aber

$$\begin{aligned} P\left(\lim_{n \rightarrow \infty} \bar{X}_{n^2} \neq 0\right) &= P\left(\left\{\bigcap_{k=1}^{\infty} \bigcup_{n=1}^{\infty} \bigcap_{m=n}^{\infty} \left\{|\bar{X}_{m^2}| \leq \frac{1}{k}\right\}\right\}^c\right) \\ &= P\left(\bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \left\{|\bar{X}_{m^2}| > \frac{1}{k}\right\}\right) \leq \sum_{k=1}^{\infty} P\left(\bigcap_{n=1}^{\infty} \bigcup_{m=n}^{\infty} \left\{|\bar{X}_{m^2}| > \frac{1}{k}\right\}\right) = 0, \end{aligned}$$

und das bedeutet gerade $P\left(\lim_{n \rightarrow \infty} \bar{X}_{n^2} = 0\right) = 1$, also

$$\lim_{n \rightarrow \infty} \bar{X}_{n^2} = 0 \text{ fast sicher.} \quad (*)$$

Es bleibt zu zeigen, dass die Konvergenz für die *gesamte* Folge (und nicht nur für die oben ausgewählte Teilfolge) gilt. Dazu wählen wir für $m \in \mathbb{N}$ den Index $n = n_m := \lfloor \sqrt{m} \rfloor$. Mit einer erneuten Anwendung der Tschebyscheff-Ungleichung erhalten wir für die Folge $\{S_k\}_{k \in \mathbb{N}}$ mit

$$S_k := k \bar{X}_k = \sum_{i=1}^k X_i$$

II. Grundprinzipien stochastischer Modellierung

$$P\left(|S_m - S_{n_m^2}| \geq \varepsilon n_m^2\right) \leq \frac{\text{Var}\left(\sum_{n_m^2 < i \leq m} X_i\right)}{\varepsilon^2 n_m^4} \leq \frac{(m - n_m^2)V}{\varepsilon^2 n_m^4} \text{ für alle } m \in \mathbb{N}$$

und somit

$$\begin{aligned} \sum_{m=1}^{\infty} P\left(|S_m - S_{n_m^2}| \geq \varepsilon n_m^2\right) &\leq \frac{V}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{m - \lfloor \sqrt{m} \rfloor^2}{\lfloor \sqrt{m} \rfloor^4} = \frac{V}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{\sqrt{m^2} - \lfloor \sqrt{m} \rfloor^2}{\lfloor \sqrt{m} \rfloor^4} \\ &= \frac{V}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{\overbrace{(\sqrt{m} - \lfloor \sqrt{m} \rfloor)}^{\leq 1} \cdot \overbrace{(\sqrt{m} + \lfloor \sqrt{m} \rfloor)}^{\leq 2\sqrt{m}}}{\lfloor \sqrt{m} \rfloor^4} \leq \frac{2V}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{\sqrt{m}}{\lfloor \sqrt{m} \rfloor^4} \leq \frac{18V}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{\sqrt{m}}{\sqrt{m}^4} \leq \frac{18V}{\varepsilon^2} \sum_{m=1}^{\infty} \frac{1}{m^{3/2}} < \infty \end{aligned}$$

wegen $\frac{m^2}{\lfloor \sqrt{m} \rfloor^4} \leq 9$ für alle $m \in \mathbb{N}$ mit Gleichheit für $m = 3$. Wiederum nach Lemma 46 a) folgt

nun, analog zu der Argumentation oben:

$$\lim_{m \rightarrow \infty} \left(\frac{S_m}{\lfloor \sqrt{m} \rfloor^2} - \bar{X}_{\lfloor \sqrt{m} \rfloor^2} \right) = 0 \text{ fast sicher.}$$

Nach (*) oben folgt also auch $\lim_{m \rightarrow \infty} \frac{S_m}{\lfloor \sqrt{m} \rfloor^2} = 0$ fast sicher und somit letztendlich

$$0 \leq \lim_{m \rightarrow \infty} |\bar{X}_m| = \lim_{m \rightarrow \infty} \frac{|S_m|}{m} \leq \lim_{m \rightarrow \infty} \frac{|S_m|}{\lfloor \sqrt{m} \rfloor^2} = 0 \text{ fast sicher}$$

und damit das gewünschte Ergebnis. ■

Bemerkung: Satz 38 bleibt auch ohne die Varianzbedingung /quadratische Integrierbarkeit unter der alleinigen Annahme gültig, dass die $\{X_n\}_{n \in \mathbb{N}}$ integrierbar und identisch verteilt sind, vgl. BAUER (2002) oder GEORGII (2002). Diese Version geht ursprünglich auf Andrej Nikolajewitsch Kolmogoroff (1903 – 1987) zurück, der diesen Sachverhalt 1930 in seiner Arbeit *Sur la loi forte des grands nombres* erläuterte. Dabei kann man die Voraussetzungen sogar noch auf die *paarweise Unabhängigkeit* reduzieren, wie N. Etemadi 1981 in seiner Originalarbeit zeigte.

Émile Borel (1781 – 1956) hat als Erster im Jahre 1909 ein starkes Gesetz großer Zahlen formuliert, und zwar für eine Situation ähnlich der des Lemmas 45, mit $p = \frac{1}{2}$. Er betrachtete dazu die dyadische Entwicklung der irrationalen Zahlen im Intervall $[0,1]$, die für jede solche Zahl x zu einer eindeutigen unendlichen Folge von Einsen ("Treffer") und Nullen ("Niete") äquivalent ist. Bezüglich des Lebesgue-Maßes (der stetigen Gleichverteilung über $[0,1]$ entsprechend) kann man dann das Ziehen einer solchen Zahl bzw. der sie charakterisierenden 0-1-Folge als ein Zufallsexperiment im Sinne von Lemma 45 auffassen. Leider enthielt der von Borel angegebene "Beweis" eine Lücke,

II. Grundprinzipien stochastischer Modellierung

die erst 1914 von *Felix Hausdorff* (1868 – 1942) in seinem Buch *Grundzüge der Mengenlehre* geschlossen wurde.

Offensichtlich sind unter den Bedingungen von Lemma 45 auch die Bedingungen von Satz 38 erfüllt, so dass wir die von Borel und Hausdorff betrachtete Situation schärfer formulieren können:

Lemma 47: Unter den Bedingungen von Lemma 45 gilt auch:

$$\lim_{n \rightarrow \infty} H_n = p \text{ fast sicher.}$$

Bemerkung: Die Aussage in Satz 38 kann auch so gefasst werden:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n (X_k - \mu) = 0 \text{ fast sicher.}$$

Hier stellt sich natürlicherweise die Frage nach der *Konvergenzgeschwindigkeit*. Eine Antwort hierauf gibt das folgende Resultat von A.N. Kolmogoroff (1929), das wir hier aber nicht beweisen können; siehe etwa BILLINGSLEY (1986), Theorem 9.5.

Satz 39 (Gesetz vom iterierten Logarithmus): Es sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge unabhängiger, identisch verteilter, quadratisch integrierbarer Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit

$$E(X_n) = \mu \in \mathbb{R} \text{ für alle } n \in \mathbb{N} \text{ und } \text{Var}(X_n) = \sigma^2 > 0.$$

Dann gilt, mit $\sigma = \sqrt{\sigma^2} > 0$:

$$\limsup_{n \rightarrow \infty} \frac{\sum_{k=1}^n (X_k - \mu)}{\sqrt{2n \ln \ln n}} = \sigma, \quad \liminf_{n \rightarrow \infty} \frac{\sum_{k=1}^n (X_k - \mu)}{\sqrt{2n \ln \ln n}} = -\sigma \text{ fast sicher.}$$

II.7. Der Zentrale Grenzwertsatz

Das Wachstum des Terms $\ln \ln n$ im Gesetz vom iterierten Logarithmus ist außerordentlich langsam; so gilt etwa $\ln \ln n \leq 3$ für $n \leq 5 \cdot 10^8$. Man kann sich daher zu Recht fragen, was asymptotisch

mit dem Term $\frac{1}{\sqrt{n}} \sum_{k=1}^n (X_k - \mu)$ für $n \rightarrow \infty$ passiert. Tatsächlich ergibt sich hierfür auch eine Art Konvergenz, die als *schwache Konvergenz* bezeichnet wird.

II. Grundprinzipien stochastischer Modellierung

Definition 40 (schwache Konvergenz): Es seien $\{Q_n\}_{n \in \mathbb{N}}$ und Q Verteilungen auf der Borel'schen σ -Algebra \mathcal{B}^1 . Die Folge $\{Q_n\}_{n \in \mathbb{N}}$ heißt *schwach* (oder auch *verteilungs-*)*konvergent* gegen Q , wenn gilt:

$$\lim_{n \rightarrow \infty} \int g dQ_n = \int g dQ \text{ für alle } g \in CB(\mathbb{R}),$$

wobei $CB(\mathbb{R})$ die Menge der reellen, auf \mathbb{R} stetigen und beschränkten Funktionen bezeichne; in Zeichen: $Q = w\text{-}\lim Q_n$ oder $Q_n \xrightarrow{w} Q$ ($n \rightarrow \infty$).

Das folgende Resultat zeigt, dass man die Funktionenklasse $CB(\mathbb{R})$ dabei sogar noch weiter einschränken kann.

Lemma 48: Es seien $\{Q_n\}_{n \in \mathbb{N}}$ und Q Verteilungen auf der Borel'schen σ -Algebra \mathcal{B}^1 . Die Folge $\{Q_n\}_{n \in \mathbb{N}}$ ist genau dann schwach konvergent gegen Q , wenn gilt:

$$\lim_{n \rightarrow \infty} \int g dQ_n = \int g dQ \text{ für alle } g \in UCB(\mathbb{R}),$$

wobei $UCB(\mathbb{R})$ die Menge der reellen, auf \mathbb{R} *gleichmäßig* stetigen und beschränkten Funktionen bezeichne (aus dem Englischen: **u**niformly **c**ontinuous and **b**ounded).

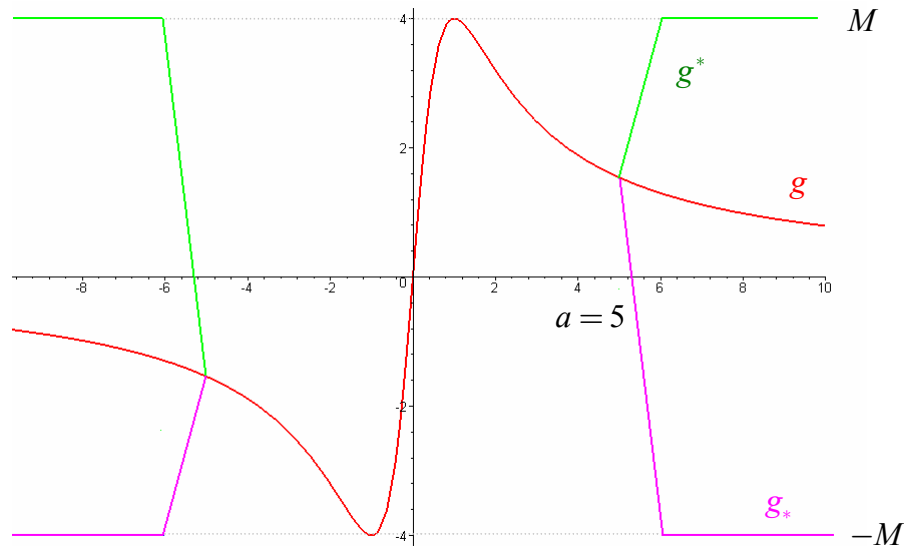
Beweis: Wegen $UCB(\mathbb{R}) \subset CB(\mathbb{R})$ genügt es, das Hinreichen der angegebenen Bedingung zu zeigen. Sei dazu g stetig und beschränkt, etwa mit $|g| \leq M \in \mathbb{R}$, sowie $\varepsilon > 0$. Wähle eine Zahl $a > 0$, so dass

$$Q\left(\left(-a, a\right)^c\right) = \int_{|x| \geq a} Q(dx) \leq \frac{\varepsilon}{2M}$$

ist sowie stetige Funktionen g_*, g^* mit den Eigenschaften

$$\begin{aligned} g_*(x) &= g^*(x) = g(x), & |x| &\leq a \\ -M &\leq g_*(x) \leq g(x), & a &\leq |x| \leq a+1 \\ g(x) &\leq g^*(x) \leq M, & a &\leq |x| \leq a+1 \\ g_*(x) &= -M, & |x| &> a+1 \\ g^*(x) &= M, & |x| &> a+1. \end{aligned}$$

II. Grundprinzipien stochastischer Modellierung



Veranschaulichung der Funktionen g_* und g^*

Dann sind g_* und g^* auf \mathbb{R} gleichmäßig stetig (da diese Funktionen auf dem kompakten Intervall $[-a-1, a+1]$ stetig und damit dort auch gleichmäßig stetig sind und außerhalb dieses Intervalls konstant), mit der Eigenschaft

$$\int (g^* - g_*) dQ \leq 2M \int_{|x| \geq a} Q(dx) \leq \varepsilon.$$

Es folgt

$$\begin{aligned} \liminf_{n \rightarrow \infty} \int g dQ_n &\geq \liminf_{n \rightarrow \infty} \int g_* dQ_n = \int g_* dQ = \int g dQ - \int (g - g_*) dQ \\ &\geq \int g dQ - \int (g^* - g_*) dQ \geq \int g dQ - \varepsilon \end{aligned}$$

sowie analog

$$\limsup_{n \rightarrow \infty} \int g dQ_n \leq \int g dQ + \varepsilon.$$

Da $\varepsilon > 0$ beliebig war, folgt also $\lim_{n \rightarrow \infty} \int g dQ_n = \int g dQ$, was zu beweisen war. ■

Bemerkung: Sind die Verteilungen Q_n und Q speziell Verteilungen von Zufallsvariablen, d.h. ist $Q_n = P^{X_n}$, $Q = P^X$, so bedeutet die schwache Konvergenz wegen Lemma 34 in diesem Falle gerade

$$\lim_{n \rightarrow \infty} E(g(X_n)) = E(g(X)) \text{ für alle } g \in [U]CB(\mathbb{R}).$$

Wir schreiben dann auch $X = w\text{-}\lim_{n \rightarrow \infty} X_n$, allerdings ist der "Grenzwert" X als messbare Abbildung in diesem Fall in der Regel nicht einmal fast sicher eindeutig bestimmt, sondern nur seine Verteilung.

Einen Zusammenhang zwischen schwacher und stochastischer Konvergenz gibt der folgende

II. Grundprinzipien stochastischer Modellierung

Satz 40: Es sei $\{X, X_n\}_{n \in \mathbb{N}}$ eine Folge reeller Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Dann gelten folgende Aussagen:

a) $P\text{-}\lim_{n \rightarrow \infty} X_n = X \Rightarrow w\text{-}\lim_{n \rightarrow \infty} X_n = X,$

d.h. *stochastische* impliziert *schwache* Konvergenz.

b) Ist X fast sicher konstant, gilt hiervon auch die Umkehrung:

$$w\text{-}\lim_{n \rightarrow \infty} X_n = X \Rightarrow P\text{-}\lim_{n \rightarrow \infty} X_n = X.$$

Beweis:

a) Gemäß Lemma 48 können wir uns auf *gleichmäßig* stetige Funktionen g beschränken, etwa mit $|g| \leq M \in \mathbb{R}$. Wegen der gleichmäßigen Stetigkeit gibt es zu jedem $\varepsilon > 0$ ein $\delta > 0$ mit der Eigenschaft

$$|x - y| \leq \delta \Rightarrow |g(x) - g(y)| \leq \varepsilon \text{ für alle } x, y \in \mathbb{R}.$$

Für $n \in \mathbb{N}$ bezeichne $A_n(\delta) := \{|X_n - X| \leq \delta\}$. Dann folgt für alle $n \in \mathbb{N}$:

$$\begin{aligned} |E(g(X_n)) - E(g(X))| &\leq \int_{A_n(\delta)} |g(X_n) - g(X)| dP + \int_{A_n(\delta)^c} |g(X_n) - g(X)| dP \\ &\leq \varepsilon + 2M \int_{A_n(\delta)^c} dP \leq \varepsilon + 2M \cdot P(A_n(\delta)^c) = \varepsilon + 2M \cdot P(|X_n - X| > \delta), \end{aligned}$$

so dass nach Grenzübergang wegen der vorausgesetzten stochastischen Konvergenz folgt:

$$\limsup_{n \rightarrow \infty} |E(g(X_n)) - E(g(X))| \leq \varepsilon \text{ für alle } \varepsilon > 0,$$

also $\lim_{n \rightarrow \infty} E(g(X_n)) = E(g(X))$ gilt, was diesen Teil des Satzes beweist.

b) Sei $X = a \in \mathbb{R}$ fast sicher. Für $\varepsilon > 0$ wählen wir

$$g_\varepsilon(x) := \min \left\{ 1, \frac{|x - a|}{\varepsilon} \right\}, \quad x \in \mathbb{R}.$$

Dann ist $g_\varepsilon \in UCB(\mathbb{R})$ mit

$$P(|X_n - X| > \varepsilon) = E(\mathbb{1}_{\{|X_n - X| > \varepsilon\}}) \leq E(g_\varepsilon(X_n)) \rightarrow E(g_\varepsilon(X)) = g_\varepsilon(a) = 0 \text{ für } n \rightarrow \infty,$$

also gilt $P\text{-}\lim_{n \rightarrow \infty} X_n = X$. ■

II. Grundprinzipien stochastischer Modellierung

Die Aussage in Teil a) des letzten Satzes kann nicht ohne weiteres umgekehrt werden, dies zeigt das folgende

Beispiel: Es sei $(\Omega, \mathcal{A}) = (\mathbb{R}^1, \mathcal{B}^1)$ und P die stetige Gleichverteilung $\mathcal{U}[0,1]$ über dem Intervall $[0,1]$ und $0 < p < \frac{1}{2}$. Für die durch

$$X_n := \begin{cases} \mathbb{1}_{[0,p]} & \text{falls } n \text{ ungerade} \\ \mathbb{1}_{[1-p,1]} & \text{falls } n \text{ gerade} \end{cases} \quad \text{für } n \in \mathbb{N}$$

definierten Zufallsvariablen $\{X_n\}_{n \in \mathbb{N}}$ gilt dann $P^{X_n} = B(1, p)$ für alle $n \in \mathbb{N}$, d.h. trivialerweise also auch $B(1, p) = w\text{-}\lim_{n \in \mathbb{N}} P^{X_n}$. Die Folge $\{X_n\}_{n \in \mathbb{N}}$ konvergiert aber nicht stochastisch. Denn gäbe es eine Zufallsvariable X mit $X = P\text{-}\lim_{n \in \mathbb{N}} P^{X_n}$, dann wäre nach Satz 40 a) $P^X = B(1, p)$, d.h. es würde eine Borel-Menge $A \subseteq [0,1]$ existieren mit $X = \mathbb{1}_A$ und $P(A) = p$. Für $0 < \varepsilon < 1$ wäre dann

$$\{|X_n - X| > \varepsilon\} = \{X_n \neq X\} = \begin{cases} ((p,1] \cap A) \oplus ([0,p] \cap A^c) & \text{falls } n \text{ ungerade} \\ ([0,1-p) \cap A) \oplus ([1-p,1] \cap A^c) & \text{falls } n \text{ gerade} \end{cases} \quad \text{für alle } n \in \mathbb{N}.$$

Da diese Mengen nicht von n abhängen, muss für stochastische Konvergenz also notwendig $P((p,1] \cap A) \oplus ([0,p] \cap A^c) = 0$ und $P([0,1-p) \cap A) \oplus ([1-p,1] \cap A^c) = 0$ gelten, woraus durch Addition der Terme und Zusammenfassung der Mengen $0 \leq P(A) \leq P((p,1] \cap A) + P([0,1-p) \cap A) = 0$ \Leftarrow folgen würde.

Das folgende Resultat veranschaulicht einige weitere wichtige Eigenschaften der schwachen Konvergenz.

Satz 41. Es seien $\{Q_n\}_{n \in \mathbb{N}}$ und Q Verteilungen auf der Borel'schen σ -Algebra \mathcal{B}^1 . Die zugehörigen Verteilungsfunktionen seien mit F_n , $n \in \mathbb{N}$ bzw. F bezeichnet; $C(F)$ bzw. $C(F^{-1})$ mögen die Mengen der reellen Stetigkeitspunkte von F bzw. der im Intervall $(0,1)$ gelegenen Stetigkeitspunkte der Pseudoinversen F^{-1} bezeichnen. Die folgenden drei Aussagen sind dann äquivalent:

- $Q = w\text{-}\lim Q_n$
- $F(x) = \lim_{n \rightarrow \infty} F_n(x)$ für alle $x \in C(F)$
- $F^{-1}(u) = \lim_{n \rightarrow \infty} F_n^{-1}(u)$ für alle $u \in C(F^{-1})$

Ferner gilt: Ist U eine über dem Intervall $(0,1)$ stetig gleichverteilte Zufallsvariable und gilt $Q = w\text{-}\lim Q_n$, so konvergiert die Folge $\{X_n\}_{n \in \mathbb{N}}$ mit

$$X_n := F_n^{-1}(U), \quad n \in \mathbb{N},$$

fast sicher gegen $X := F^{-1}(U)$. Dabei ist $P^X = Q$ und $P^{X_n} = Q_n$ für alle $n \in \mathbb{N}$.

II. Grundprinzipien stochastischer Modellierung

Beweis: Wir machen einen Ringschluss:

a) \Rightarrow b): Sei $x \in C(F)$ und $\varepsilon > 0$. Dann existiert ein $\delta > 0$ mit der Eigenschaft

$$|x - y| \leq \delta \quad \Rightarrow \quad |F(x) - F(y)| \leq \varepsilon \quad \text{für alle } y \in \mathbb{R}.$$

Wähle nun stetige Funktionen $0 \leq g_*, g^* \leq 1$ mit

$$g_*(y) = \begin{cases} 1, & \text{falls } y \leq x - \delta \\ 0, & \text{falls } y \geq x \end{cases} \quad \text{und} \quad g^*(y) = \begin{cases} 1, & \text{falls } y \leq x \\ 0, & \text{falls } y \geq x + \delta \end{cases} \quad \text{für } y \in \mathbb{R}.$$

Es folgt

$$\begin{aligned} \liminf_{n \rightarrow \infty} F_n(x) &= \liminf_{n \rightarrow \infty} \int \mathbb{1}_{(-\infty, x]} dQ_n \geq \liminf_{n \rightarrow \infty} \int g_*(y) Q_n(dy) = \int g_*(y) Q_n(dy) \\ &\geq \int \mathbb{1}_{(-\infty, x-\delta]} dQ = F(x - \delta) \geq F(x) - \varepsilon. \end{aligned}$$

Analog erhält man

$$\limsup_{n \rightarrow \infty} F_n(x) \leq F(x) + \varepsilon$$

und damit, da $\varepsilon > 0$ beliebig war, $F(x) = \lim_{n \rightarrow \infty} F_n(x)$.

b) \Rightarrow c): Auf Grund der schwachen Monotonie von F und F^{-1} sind die Mengen $C(F)$ und $C(F^{-1})$ höchstens abzählbar, d.h. $C(F)$ liegt dicht in \mathbb{R} und $C(F^{-1})$ dicht im Intervall $(0, 1)$.

Sei nun $\varepsilon > 0$ und $u \in C(F^{-1})$. Dann existieren $x, z \in C(F)$ mit

$$x < F^{-1}(u) < z \quad \text{und} \quad z - x < \varepsilon. \quad (*)$$

Dann muss auch $F(x) < u$ gelten; denn sonst wäre $F(x) \geq u$ und damit nach Definition der Pseudo-Inversen $x \geq F^{-1}(u)$ [das ist ja das "kleinste" reelle x mit dieser Eigenschaft]. Nach Lemma 31 b) ist ferner $u \leq F(F^{-1}(u)) \leq F(z)$. Dann ist aber sogar $u < F(z)$; anderenfalls wäre nämlich $u = F(z)$ und damit für jedes $\delta \in (0, 1 - u)$

$$F^{-1}(u + \delta) = \inf \{y \in \mathbb{R} \mid F(y) \geq u + \delta\} > z,$$

zugleich aber $F^{-1}(u) < z$ im Widerspruch zur Stetigkeit von F^{-1} in u . Damit erhält man insgesamt

$$F(x) < u < F(z).$$

Nach dem vorausgesetzten Teil b) des Lemmas existiert nun eine Zahl $n_\varepsilon \in \mathbb{N}$ mit

$$F_n(x) < u < F_n(z) \quad \text{für alle } n \geq n_\varepsilon,$$

was seinerseits nach Definition der Pseudo-Inversen

II. Grundprinzipien stochastischer Modellierung

$$x \leq F_n^{-1}(u) \leq z \text{ für alle } n \geq n_\varepsilon$$

nach sich zieht. Zusammen mit (*) oben führt dies zu

$$\left| F_n^{-1}(u) - F^{-1}(u) \right| \leq z - x \leq \varepsilon \text{ für alle } n \geq n_\varepsilon,$$

womit die Gültigkeit von c) bewiesen ist.

c) \Rightarrow a): Hierzu benutzen wir den letzten Teil des Lemmas, den wir vorab beweisen. Nach Lemma 32 gilt zunächst $P^X = Q$ und $P^{X_n} = Q_n$ für alle $n \in \mathbb{N}$. Wegen der Abzählbarkeit von $N := C(F^{-1})^c$ gilt außerdem $P(U^{-1}(N)) = P^U(N) = 0$ (beachte: P^U entspricht dem auf das Intervall $(0,1)$ eingeschränkte Lebesgue-Maß), d.h. nach c) konvergiert die Folge $\{F_n^{-1}(U(\omega))\}_{n \in \mathbb{N}}$ gegen $F^{-1}(U(\omega))$ für alle $\omega \in U^{-1}(C(F^{-1}))$, mit $P(U^{-1}(C(F^{-1}))) = P^U(N^c) = 1$. Also gilt $X = F^{-1}(U) = \lim_{n \rightarrow \infty} F_n^{-1}(U) = \lim_{n \rightarrow \infty} X_n$ fast sicher, wie im letzten Teil des Satzes behauptet.

Nach den Sätzen 36 und 40 gilt nun also auch $Q = w\text{-}\lim Q_n$, was dem zu zeigenden Teil a) entspricht.

Damit ist Satz 41 vollständig bewiesen. ■

Das bemerkenswerte an der letzten Aussage in Satz 41 ist, dass sie in gewisser Weise doch die Umkehrung von Satz 40 a) impliziert, wenn man dies so versteht, dass man immer eine geeignete Konstruktion von Zufallsvariablen finden kann, die fast sichere (und damit nach Satz 38 auch stochastische) Konvergenz erzeugt, wobei die Verteilungen der Zufallsvariablen den vorgegebenen Wahrscheinlichkeitsmaßen aus der schwachen Konvergenz entsprechen. Dieser Sachverhalt ist auch als *Einbettungssatz von Skohorod* bekannt.

Als letzte Vorbereitung auf das zentrale Ergebnis dieses Abschnitts benötigen wir noch einen Zusammenhang zwischen der schwachen Konvergenz und der Konvergenz momenterzeugender Funktionen.

Satz 42: Es sei $\{X, X_n\}_{n \in \mathbb{N}}$ eine Folge reeller Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit zugehörigen momenterzeugenden Funktionen $\{\psi_X, \psi_{X_n}\}_{n \in \mathbb{N}}$, die sämtlich in einem Intervall $[-\delta, \delta]$ für ein $\delta > 0$ existieren mögen. Dann gilt:

$$\lim_{n \rightarrow \infty} \psi_{X_n}(t) = \psi_X(t) \text{ für alle } |t| \leq \delta \quad \Rightarrow \quad P^X = w\text{-}\lim P^{X_n},$$

d.h. die Konvergenz der erzeugenden Funktionen impliziert die schwache Konvergenz der Verteilungen, und es gilt ferner

$$\lim_{n \rightarrow \infty} E(X_n^k) = E(X^k) \text{ für alle } k \in \mathbb{N},$$

d.h. die Konvergenz sämtlicher Momente.

II. Grundprinzipien stochastischer Modellierung

Beweis: Der Beweis dieses Satzes beruht zunächst auf der Überlegung, dass die Menge der "Testfunktionen" g für die Überprüfung der schwachen Konvergenz noch weiter eingeschränkt werden kann, nämlich auf solche Funktionen $g \in UCB(\mathbb{R})$, die im Nullpunkt beliebig häufig differenzierbar sind, auf ganz \mathbb{R} durch ihre Taylor-Reihe

$$g(x) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} x^k, \quad x \in \mathbb{R}$$

dargestellt werden und für die die Majorante

$$\bar{g}(x) = \sum_{k=0}^{\infty} \frac{|g^{(k)}(0)|}{k!} |x|^k, \quad x \in \mathbb{R}$$

ebenfalls beschränkt ist. Aus der Konvergenz der Momente sowie der absoluten Momente (die sich direkt aus der Konvergenz der momenterzeugenden Funktionen ergibt) folgt dann nämlich auch die Konvergenz von

$$E(g(X_n)) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} E(X_n^k) \quad \text{gegen} \quad E(g(X)) = \sum_{k=0}^{\infty} \frac{g^{(k)}(0)}{k!} E(X^k), \quad (*)$$

womit die schwache Konvergenz gezeigt ist. Im Einzelnen sieht man das so: Die (messbaren) Abbildungen $|g(X_n)|$ und $|g(X)|$ sind für alle $n \in \mathbb{N}$ durch dieselbe Konstante – etwa $M > 0$ – beschränkt. Nach dem Lebesgue'schen Satz von der majorisierten Konvergenz (Satz 19) ergibt sich also die (absolute) Konvergenz der Reihen in (*) mit dem jeweils angegebenen Grenzwert. Setzt man jetzt noch

$$s_n(k) := \sup_{m \geq n} |E(X_m^k) - E(X^k)|, \quad n, k \in \mathbb{N},$$

so folgt

$$\sum_{k=0}^{\infty} \frac{|g^{(k)}(0)|}{k!} s_1(k) \leq 2M < \infty.$$

Sei nun $\varepsilon > 0$. Dann existiert eine Zahl $N(\varepsilon) \in \mathbb{N}$ mit

$$\sum_{k > N(\varepsilon)} \frac{|g^{(k)}(0)|}{k!} s_1(k) \leq \varepsilon.$$

Auf Grund der Konvergenz der Momente ergibt sich aber

$$\lim_{n \rightarrow \infty} s_n(k) = 0 \quad \text{für} \quad k = 1, 2, \dots, N(\varepsilon),$$

so dass

$$\lim_{n \rightarrow \infty} |E(g(X_n)) - E(g(X))| \leq \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \frac{|g^{(k)}(0)|}{k!} s_n(k) \leq \lim_{n \rightarrow \infty} \sum_{k > N(\varepsilon)} \frac{|g^{(k)}(0)|}{k!} s_n(k) \leq \sum_{k > N(\varepsilon)} \frac{|g^{(k)}(0)|}{k!} s_1(k) \leq \varepsilon$$

folgt. Da $\varepsilon > 0$ beliebig war, ergibt sich also die Konvergenzaussage in (*). ■

II. Grundprinzipien stochastischer Modellierung

Satz 43 (Zentraler Grenzwertsatz): Es sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge stochastisch unabhängiger, identisch verteilter, quadratisch integrierbarer Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit

$$E(X_n) = \mu \in \mathbb{R} \quad \text{und} \quad \text{Var}(X_n) = \sigma^2 > 0 \quad \text{für alle } n \in \mathbb{N}.$$

Die (normalisierten) Zufallsvariablen $\{T_n\}_{n \in \mathbb{N}}$ seien definiert durch

$$T_n := \frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n (X_k - \mu) \quad \text{für alle } n \in \mathbb{N}.$$

Dann gilt:

$$\mathcal{N}(0,1) = w\text{-}\lim P^{T_n}, \quad \text{d.h. insbesondere} \quad \lim_{n \rightarrow \infty} P\left(\frac{1}{\sigma\sqrt{n}} \sum_{k=1}^n (X_k - \mu) \leq x\right) = \Phi(x) \quad \text{für alle } x \in \mathbb{R},$$

unabhängig von der den $\{X_n\}_{n \in \mathbb{N}}$ zu Grunde liegenden Verteilung.

Beweis: Wir zeigen den Satz hier nur unter der stärkeren Annahme, dass die momenterzeugende Funktion ψ der $\{X_n\}_{n \in \mathbb{N}}$ in einem Intervall $[-\delta, \delta]$ mit einem $\delta > 0$ existiert. Dazu bezeichne $Y_n := \frac{X_n - \mu}{\sigma}$, $n \in \mathbb{N}$. Dann gilt $E(Y_n) = 0$ und $\text{Var}(Y_n) = 1$ für alle $n \in \mathbb{N}$, und es existieren nach Satz 33 alle Momente von X_n und Y_n sowie die momenterzeugende Funktion ψ^* der Y_n , die durch

$$\psi^*(t) = \psi\left(\frac{t}{\sigma}\right) \cdot e^{-\frac{\mu t}{\sigma}}, \quad |t| \leq \delta\sigma$$

gegeben ist. Bezeichnet ferner ψ_n die momenterzeugende Funktion der $\frac{Y_i}{\sqrt{n}}$, $i, n \in \mathbb{N}$, so erhält man weiter für alle $i \in \mathbb{N}$ und $|t| \leq \delta\sigma$:

$$\psi_n(t) = \psi^*\left(\frac{t}{\sqrt{n}}\right) = \sum_{k=0}^{\infty} \frac{t^k}{k! \sqrt{n}^k} E(Y_i^k) = 1 + \frac{t^2}{2n} + R_n(t)$$

mit

$$R_n(t) = \sum_{k=3}^{\infty} \frac{t^k}{k! \sqrt{n}^k} E(Y_i^k) \leq \frac{1}{n\sqrt{n}} \sum_{k=3}^{\infty} \frac{t^k}{k!} E(|Y_i^k|) \leq \frac{1}{n\sqrt{n}} (\psi^*(\delta) + \psi^*(-\delta)) = \frac{M}{n\sqrt{n}} \quad \text{für alle } n \in \mathbb{N}$$

(vgl. den Beweis zu Satz 33). Damit ergibt sich für die momenterzeugende Funktion ψ_{T_n} von T_n für $|t| \leq \delta\sigma$:

$$\psi_{T_n}(t) = (\psi_n(t))^n = \left(1 + \frac{t^2}{2n} + R_n(t)\right)^n \rightarrow \exp\left(\frac{t^2}{2}\right) \quad \text{für } n \rightarrow \infty. \quad (*)$$

Die Grenzfunktion auf der rechten Seite ist aber gerade die momenterzeugende Funktion einer $\mathcal{N}(0,1)$ -verteilten Zufallsvariablen, so dass mit Satz 42 und Satz 41 b) die Aussage folgt. ■

II. Grundprinzipien stochastischer Modellierung

Bemerkung: Wenn man in der momenterzeugenden Funktion zum imaginären Argument it mit der imaginären Einheit $i = \sqrt{-1}$ übergeht, erhält man die so genannte *Fourier-Transformierte* oder *charakteristische Funktion* der Zufallsvariablen bzw. deren Verteilung. Mit dieser Funktion kann der Beweis wörtlich übertragen werden, d.h. man erhält die Konvergenz der charakteristischen Funktionen der T_n gegen diejenige der Standard-Normalverteilung. In diesem Fall benötigt man keine einschränkenden Bedingungen, weil die Fourier-Transformierte für alle $t \in \mathbb{R}$ existiert (das beruht auf der Abschätzung $|e^{it}| = 1$ für alle $t \in \mathbb{R}$). Satz 42 gilt dann entsprechend auch ohne Einschränkungen für die Fourier-Transformierten. Diesen (klassischen) Beweisgang findet man z.B. in BAUER (2002). GEORGII (2002) gibt einen alternativen, eher elementaren Beweis in seinem Satz (5.28), der aber den Kern des Arguments nicht deutlich macht.

In seiner frühesten Form geht der Beweis des Zentralen Grenzwertsatzes schon auf *Alexander Michailovitsch Ljapunoff* (1857 – 1918) aus dem Jahr 1901 zurück.

Die obigen Voraussetzungen in Satz 43 können natürlich abgeschwächt werden, solange eine Grenzwertaussage ähnlich (*) erhalten bleibt. Eine solche Abschwächung macht der

Satz 44 (Zentraler Grenzwertsatz von Lindeberg): Es sei $\{X_n\}_{n \in \mathbb{N}}$ eine Folge stochastisch unabhängiger quadratisch integrierbarer Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit

$$E(X_n) = \mu \in \mathbb{R} \quad \text{und} \quad \text{Var}(X_n) = \sigma_n^2 > 0 \quad \text{für alle } n \in \mathbb{N}.$$

Die (normalisierten) Zufallsvariablen $\{T_n\}_{n \in \mathbb{N}}$ seien definiert durch

$$T_n := \frac{1}{\sigma^{[n]}} \sum_{k=1}^n (X_k - \mu) \quad \text{für alle } n \in \mathbb{N}$$

mit

$$\sigma^{[n]} = \sqrt{\text{Var}\left(\sum_{k=1}^n X_k\right)} = \sqrt{\sum_{k=1}^n \sigma_k^2}.$$

Dann gilt $\mathcal{N}(0,1) = w\text{-}\lim P^{T_n}$, wenn die so genannte *Lindeberg-Bedingung* erfüllt ist:

$$\lim_{n \rightarrow \infty} \frac{1}{(\sigma^{[n]})^2} \sum_{k=1}^n \int_{\{|x-\mu| \geq \varepsilon \sigma^{[n]}\}} (x-\mu)^2 P^{X_k}(dx) = 0 \quad \text{für ein } \varepsilon > 0.$$

II. Grundprinzipien stochastischer Modellierung

Bemerkung: Die Lindeberg-Bedingung ist in folgenden Situationen erfüllt (und damit der Zentrale Grenzwertsatz gültig):

a) Die Folge $\{X_n\}_{n \in \mathbb{N}}$ ist *identisch verteilt*. Dies entspricht der Situation des Satzes 43.

b) Die Folge $\{X_n\}_{n \in \mathbb{N}}$ erfüllt die so genannte *Ljapunoff-Bedingung*:

$$\lim_{n \rightarrow \infty} \frac{1}{(\sigma^{[n]})^{2+\delta}} \sum_{k=1}^n E(|X_k - \mu|^{2+\delta}) = 0.$$

c) Die Folge $\{X_n\}_{n \in \mathbb{N}}$ ist gleichmäßig beschränkt und es gilt $\lim_{n \rightarrow \infty} \sigma^{[n]} = \infty$. (In diesem Fall ist nämlich die Ljapunoff-Bedingung erfüllt.)

Für weitergehende Diskussionen zum Zentralen Grenzwertsatz siehe BAUER (2002).

II.8. Bedingte Verteilungen

In diesem Abschnitt kommen wir noch einmal auf die mit Definition 27 begonnene Begriffsbildung der bedingten Wahrscheinlichkeiten zurück, allerdings in wesentlich größerer Allgemeinheit. Wir beginnen dazu mit der Situation aus Satz 28, wo in dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) eine disjunkte Zerlegung $\Omega = \bigoplus_{i \in I} B_i$ mit $\{B_i\}_{i \in I} \subseteq \mathcal{A}$ gegeben ist, mit einer höchstens abzählbaren Indexmenge I . (Gemäß der Bemerkung nach Satz 28 brauchen hier keine einschränkenden Voraussetzungen an die $P(B_i)$, $i \in I$ gemacht zu werden.) Nach Satz 3 ist dann

$$\mathcal{C} := \left\{ \bigoplus_{i \in J} B_i \mid J \subseteq I \right\}$$

eine σ -Algebra über Ω , und zwar genau die von $\{B_i\}_{i \in I}$ über Ω erzeugte. \mathcal{C} ist also insbesondere eine Teil σ -Algebra von \mathcal{A} ; wir wollen dies durch die Schreibweise

$$\mathcal{C} \sqsubseteq \mathcal{A}$$

zum Ausdruck bringen. Definiert man für $A \in \mathcal{A}$ die Zufallsvariable $P(A|\mathcal{C})$ durch

$$P(A|\mathcal{C})(\omega) = \sum_{i \in I} P(A|B_i) \cdot \mathbb{1}_{B_i}(\omega) \text{ für alle } \omega \in \Omega,$$

so ist diese \mathcal{C} -messbar, da sie auf den Erzeugermengen $\{B_i\}_{i \in I}$ jeweils konstant ist. Wegen $\mathcal{C} \sqsubseteq \mathcal{A}$ ist sie natürlich auch \mathcal{A} -messbar. Andererseits erhalten wir durch Integration

$$P(A \cap B) = \sum_{i \in I_B} P(A|B_i) \cdot P(B_i) = \int_B P(A|\mathcal{C})(\omega) P(d\omega) = \int_B P(A|\mathcal{C}) dP \text{ für alle } A \in \mathcal{A}, B \in \mathcal{C}$$

II. Grundprinzipien stochastischer Modellierung

mit der (eindeutigen) Darstellung $B = \bigoplus_{i \in I_B} B_i$, d.h. $I_B \subseteq I$ besteht aus genau denjenigen Indices, die für die disjunkte Darstellung von B benötigt werden. Diese Darstellung entspricht exakt derjenigen aus Satz 28 a) für die Wahl $B = \Omega$. Bezeichnen wir mit P_C die Einschränkung von P auf die Teil- σ -Algebra \mathcal{C} , so bedeutet die obige Integraldarstellung nach Definition 21 gerade, dass für jedes (feste) $A \in \mathcal{A}$ $P(A|\mathcal{C})$ eine P_C -Dichte des Spurmaßes $P(\cdot \cap A)$ auf \mathcal{C} ist. Insofern ist $P(A|\mathcal{C})$ auch nur P_C -fast sicher eindeutig bestimmt.

Definition 41 (allgemeine bedingte Wahrscheinlichkeit): Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\mathcal{C} \sqsubseteq \mathcal{A}$ eine beliebige Teil- σ -Algebra. Jede \mathcal{C} -messbare Zufallsvariable $P(A|\mathcal{C})$ mit $A \in \mathcal{A}$ und den folgenden Eigenschaften

- $P(A|\mathcal{C})(\omega) \in [0,1]$ für alle $\omega \in \Omega$
- $P(\emptyset|\mathcal{C}) = 0$, $P(\Omega|\mathcal{C}) = 1$
- $P(A \cap B) = \int_B P(A|\mathcal{C}) dP$ für alle $B \in \mathcal{C}$

heißt *bedingte Wahrscheinlichkeit von A unter (der σ -Algebra) \mathcal{C}* .

Im Sinne der Eingangsbemerkungen ist damit $P(A|\mathcal{C})$ eine natürliche Verallgemeinerung des Begriffs der elementaren bedingten Wahrscheinlichkeit in der Situation von Teil- σ -Algebren \mathcal{C} mit abzählbarem Erzeuger (bzw. abzählbarem Atomsystem). Es ist aber im Allgemeinen nicht grundsätzlich sichergestellt, dass auch die für ein (bedingtes) Wahrscheinlichkeitsmaß notwendige σ -Additivität gilt, d.h. hier:

$$P\left(\bigcup_{n=1}^{\infty} A_n \mid \mathcal{C}\right) = \sum_{n=1}^{\infty} P(A_n \mid \mathcal{C}) \quad P_C\text{-fast sicher}$$

für alle paarweise disjunkten Folgen $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$. Dies liegt daran, dass die Ausnahmemengen, auf denen $P(A|\mathcal{C})$ nicht eindeutig ist, von jeder der Mengen $A \in \mathcal{A}$ abhängen kann, so dass die (gemeinsame) Ausnahmemenge, auf der die σ -Additivität nicht gilt, so gross werden kann, dass sie keine Nullmenge mehr oder ggf. sogar nicht mehr messbar ist. Um dies trotzdem zu gewährleisten, müssen an die Grundmenge Ω bestimmte topologische Anforderungen gestellt werden, z.B. die, *polnisch* zu sein, d.h. es existiert eine die Topologie definierende Metrik ρ über Ω , so dass (Ω, ρ) ein vollständiger separabler metrischer Raum ist (vgl. ELSTRODT (1996), Kapitel VIII und BAUER (2002)). Dies ist nämlich eine hinreichende Bedingung für die Existenz eines abzählbaren Erzeugers der von der Topologie erzeugten σ -Algebra \mathcal{A} über Ω , was für die Konstruktion einer hinreichend großen Ausnahme(-Null-)Menge genügt. Beispielsweise sind die Räume \mathbb{R}^d für $d \in \mathbb{N}$ mit der üblichen Euklidischen Metrik solche polnischen Räume.

II. Grundprinzipien stochastischer Modellierung

Definition 42 (reguläre bedingte Verteilungen): Gilt unter den Bedingungen von Definition 42 zusätzlich für alle paarweise disjunkten Folgen $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$:

- $P\left(\bigcup_{n=1}^{\infty} A_n \mid \mathcal{C}\right) = \sum_{n=1}^{\infty} P(A_n \mid \mathcal{C})$ $P_{\mathcal{C}}$ -fast sicher,

so heißt das durch $P(\cdot \mid \mathcal{C})$ $P_{\mathcal{C}}$ -fast sicher bestimmte (bedingte) Wahrscheinlichkeitsmaß auf \mathcal{A} *regulär*.

Bemerkung: Da es für alle Mengenfolgen $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$, die für diese Bedingung betrachtet werden können, eine gemeinsame Ausnahme(-Null-)Menge gibt, können und werden wir o.B.d.A. bei regulären bedingten Verteilungen stets davon ausgehen, dass $P(\cdot \mid \mathcal{C})(\omega)$ für *alle* $\omega \in \Omega$ ein Wahrscheinlichkeitsmaß über \mathcal{A} ist. Dies lässt sich z.B. einfach dadurch erreichen, dass auf dieser Ausnahmemenge $P(\cdot \mid \mathcal{C})(\omega) = P$ gesetzt wird (vgl. die Bemerkung im Anschluss an Definition 27).

Wir benötigen für das Folgende eine modifizierte Version des Satzes 9.

Satz 45 (Faktorisierungssatz II). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und \mathbf{Y} ein Zufallselement mit Werten in einem Messraum $(\mathcal{X}, \mathcal{B})$. \mathcal{C} bezeichne die von \mathbf{Y} erzeugte Teil- σ -Algebra von \mathcal{A} . Eine reellwertige Zufallsvariable X auf (Ω, \mathcal{A}, P) ist genau dann \mathcal{C} -messbar, wenn eine messbare Abbildung $h: (\mathcal{X}, \mathcal{B}) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$ existiert mit $X = h \circ \mathbf{Y}$.

Beweis: " \Leftarrow ": Sei $X = h \circ \mathbf{Y}$. Dann gilt

$$X^{-1}(B) = \mathbf{Y}^{-1}\left(\underbrace{h^{-1}(B)}_{\in \mathcal{B}}\right) \in \mathcal{C} \text{ für alle } B \in \mathcal{B}^1,$$

d.h. X ist \mathcal{C} -messbar.

" \Rightarrow ": Sei zunächst X eine Elementarfunktion mit der Darstellung (vgl. Lemma 16)

$$X = \sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$$

mit reellen Zahlen $\alpha_1, \dots, \alpha_n$ und Mengen $A_1, \dots, A_n \in \mathcal{C}$, $n \in \mathbb{N}$. Daher existieren Mengen $B_1, \dots, B_n \in \mathcal{B}$ mit $A_i = \mathbf{Y}^{-1}(B_i)$, $i = 1, \dots, n$. Die Abbildung $h := \sum_{i=1}^n \alpha_i \mathbb{1}_{B_i}$ leistet dann das Gewünschte. Ist $X \geq 0$ beliebig messbar, so existiert nach Satz 16 eine monoton wachsende Folge $\{X_n\}_{n \in \mathbb{N}}$ messbarer Elementarfunktionen mit $X = \sup_{n \in \mathbb{N}} \{X_n\}$. Also existiert nach dem gerade Gezeigten auch eine Folge \mathcal{B} -messbarer Abbildungen $\{h_n\}_{n \in \mathbb{N}}$ mit $X_n = h_n \circ \mathbf{Y}$ für alle $n \in \mathbb{N}$. Hier leistet $h := \sup_{n \in \mathbb{N}} \{h_n\}$ Das Gewünschte. Für den beliebigen Fall zerlege man X in Positiv- und Negativteil: $X = X^+ - X^-$. Wiederum existieren nach dem Vorigen \mathcal{B} -messbare Abbildungen h^+ und

II. Grundprinzipien stochastischer Modellierung

h^- mit $X^\pm = h^\pm \circ \mathbf{Y}$. Hier leistet $h := h^+ - h^-$ das Gewünschte. Man beachte, dass dabei h^+ und h^- so gewählt werden können, dass nicht zugleich der Wert $h^+(\mathbf{x}) = h^-(\mathbf{x}) = \infty$ für ein $\mathbf{x} \in \mathfrak{X}$ angenommen wird. Damit ist der Satz bewiesen. ■

Für bedingte Verteilungen $P(\cdot | \sigma(\mathbf{Y}))$ mit Zufallselementen \mathbf{Y} schreiben wir im Folgenden auch kürzer $P(\cdot | \mathbf{Y})$. Die nach dem Faktorisierungssatz II (Satz 45) für jedes Ereignis $A \in \mathcal{A}$ existierende Abbildung h_A mit $P(A | \mathbf{Y}) = h_A(\mathbf{Y})$ wollen wir einprägsamer auch als

$$h_A(\mathbf{y}) = P(A | \mathbf{Y} = \mathbf{y}), \quad \mathbf{y} \in \mathfrak{X}$$

schreiben.

Aufgrund des Transformationssatzes für Maße (Satz 22) ist die grundlegende Gleichung in Definition 41 in diesem Fall äquivalent zu

$$P(A \cap \{\mathbf{Y} \in B\}) = \int_{\{\mathbf{Y} \in B\}} P(A | \mathbf{Y}) dP = \int_B P(A | \mathbf{Y} = \mathbf{y}) P^{\mathbf{Y}}(d\mathbf{y}), \quad A \in \mathcal{A}, B \in \mathcal{B}. (*)$$

Lassen sich die Abbildungen h_A so wählen, daß $P(\cdot | \mathbf{Y} = \mathbf{y})$ für jedes $\mathbf{y} \in \mathfrak{X}$ wieder eine Wahrscheinlichkeitsverteilung ist, so spricht man bei $P(\cdot | \mathbf{Y} = \mathbf{y})$ ebenfalls von der (regulären) bedingten Verteilung unter (der Hypothese) $\mathbf{Y} = \mathbf{y}$.

Der folgende Satz gestattet eine für praktische Zwecke besonders nützliche Konstruktion regulärer bedingter Verteilungen für den Fall, dass die beteiligten Zufallselemente endlichdimensionale Zufallsvektoren mit Dichten bezüglich eines geeigneten Maßes μ (etwa Zählmaß, Lebesgue-Maß) bilden.

Satz 46 (bedingte Verteilungen und Dichten). Es sei \mathbf{X} ein n -dimensionaler und \mathbf{Y} ein m -dimensionaler Zufallsvektor ($n, m \in \mathbb{N}$) auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) derart, dass $P^{(\mathbf{X}, \mathbf{Y})}$ eine $(n + m)$ -dimensionale Dichte $f_{(\mathbf{X}, \mathbf{Y})}$ bezüglich eines geeigneten Produkt-Maßes $\mu = \mu_n \otimes \nu_m$ auf der Borel'schen σ -Algebra $\mathcal{B}^{n+m} = \mathcal{B}^n \otimes \mathcal{B}^m$ besitze. In diesem Fall existiert eine reguläre Version der bedingten Verteilung $P^{\mathbf{X}}(\cdot | \mathbf{Y} = \mathbf{y})$; eine solche ist etwa gegeben durch die zugehörige bedingte μ_n -Dichte

$$f_{\mathbf{X}}(\mathbf{x} | \mathbf{Y} = \mathbf{y}) := \begin{cases} \frac{f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})}, & \text{falls } f_{\mathbf{Y}}(\mathbf{y}) > 0 \\ f_{\mathbf{X}}(\mathbf{x}), & \text{falls } f_{\mathbf{Y}}(\mathbf{y}) = 0, \end{cases} \quad \mathbf{y} \in \mathbb{R}^m.$$

Hierbei bezeichnet $f_{\mathbf{X}}(\mathbf{x})$ die Randdichte von \mathbf{X} (entsprechend für $f_{\mathbf{Y}}(\mathbf{y})$), die gegeben ist durch

$$f_{\mathbf{X}}(\mathbf{x}) = \int_{\mathbb{R}^m} f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) \nu_m(d\mathbf{y}), \quad \mathbf{x} \in \mathbb{R}^n.$$

II. Grundprinzipien stochastischer Modellierung

Beweis: Für $A \in \mathcal{B}^m$, $B \in \mathcal{B}^n$ gilt definitionsgemäß mit dem Satz von Fubini (Satz 23):

$$\begin{aligned} P(\mathbf{X} \in A, \mathbf{Y} \in B) &= \int_B \int_A f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y}) \mu_n(d\mathbf{x}) \nu_m(d\mathbf{y}) = \int_B \int_A \frac{f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \mathbf{y})}{f_{\mathbf{Y}}(\mathbf{y})} \cdot f_{\mathbf{Y}}(\mathbf{y}) \mu_n(d\mathbf{x}) \nu_m(d\mathbf{y}) \\ &= \int_B \left[\int_A f_{\mathbf{X}}(\mathbf{x} | \mathbf{Y} = \mathbf{y}) \mu_n(d\mathbf{x}) \right] \cdot f_{\mathbf{Y}}(\mathbf{y}) \nu_m(d\mathbf{y}) = \int_B \left[\int_A f_{\mathbf{X}}(\mathbf{x} | \mathbf{Y} = \mathbf{y}) \mu_n(d\mathbf{x}) \right] P^{\mathbf{Y}}(d\mathbf{y}); \end{aligned}$$

andererseits gilt nach (*) oben auch

$$P(\mathbf{X} \in A, \mathbf{Y} \in B) = \int_B P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) P^{\mathbf{Y}}(d\mathbf{y}),$$

woraus durch Vergleich

$$P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) = \int_A f_{\mathbf{X}}(\mathbf{x} | \mathbf{Y} = \mathbf{y}) \mu_n(d\mathbf{x})$$

und damit die Behauptung folgt. ■

Der vorangegangene Satz 46 zeigt, dass im diskreten Fall die reguläre bedingte Zähl-dichte $f_{\mathbf{X}}(\mathbf{x} | \mathbf{Y} = \mathbf{y})$ für $P(\mathbf{Y} = \mathbf{y}) > 0$ gerade mit der elementaren bedingten Wahrscheinlichkeit $P(\mathbf{X} = \mathbf{x} | \mathbf{Y} = \mathbf{y})$, $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^m$ (nach Definition 27) übereinstimmt.

Eine anschauliche Erklärung von Satz 46 im Fall stetiger Verteilungen liefert das folgende Resultat.

Satz 47. In der Situation des Satzes 46 sei μ das $(n + m)$ -dimensionale Lebesgue-Maß und zusätzlich $f_{(\mathbf{X}, \mathbf{Y})}(\mathbf{x}, \cdot)$ für alle $\mathbf{x} \in \mathbb{R}^n$ und $f_{\mathbf{Y}}$ in einem Punkt $\mathbf{y} \in \mathbb{R}^m$ stetig mit $f_{\mathbf{Y}}(\mathbf{y}) > 0$.

$U_h(\mathbf{y}) := \prod_{i=1}^m [y_i - h, y_i + h]$ bezeichne für $h > 0$ die Würfel-Umgebung von \mathbf{y} mit Kantenlänge $2h$.

Dann gilt für jede Menge $A \in \mathcal{B}^n$:

$$P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) = \lim_{h \downarrow 0} P(\mathbf{X} \in A | \mathbf{Y} \in U_h(\mathbf{y})).$$

Beweis. Nach Voraussetzung ist

$$P(\mathbf{Y} \in U_h(\mathbf{y})) = \int_{U_h(\mathbf{y})} f_{\mathbf{Y}}(\mathbf{z}) d\mathbf{z} \geq \delta m^m (U_h(\mathbf{y})) = (2h)^m \delta > 0,$$

wobei $\delta > 0$ so existiert, dass $f_{\mathbf{Y}}(\mathbf{z}) \geq \delta > 0$ für alle $\mathbf{z} \in U_h(\mathbf{y})$. Damit existiert $P(\mathbf{X} \in A | \mathbf{Y} \in U_h(\mathbf{y}))$ als elementare bedingte Wahrscheinlichkeit im Sinne von Definition 27 mit

II. Grundprinzipien stochastischer Modellierung

$$P(\mathbf{X} \in A | \mathbf{Y} \in U_h(\mathbf{y})) = \frac{P(\mathbf{X} \in A, \mathbf{Y} \in U_h(\mathbf{y}))}{P(\mathbf{Y} \in U_h(\mathbf{y}))} = \frac{\frac{1}{m^m(U_h(\mathbf{y}))} \int_{U_h(\mathbf{y})} P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) f_Y(\mathbf{z}) d\mathbf{z}}{\frac{1}{m^m(U_h(\mathbf{y}))} \int_{U_h(\mathbf{y})} f_Y(\mathbf{z}) d\mathbf{z}}$$

$$\rightarrow \frac{P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) f_Y(\mathbf{y})}{f_Y(\mathbf{y})} = P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) \quad \text{für } h \downarrow 0;$$

dies folgt z.B. aus dem Mittelwertsatz für mehrfache Riemann-Integrale (vgl. HEUSER (1988), 201.5). ■

Satz 47 zeigt damit, dass unter den genannten Voraussetzungen die bedingte Wahrscheinlichkeit $P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y})$ ähnlich wie im diskreten Fall interpretiert werden kann: die Bedingung $\{\mathbf{Y} \in U_h(\mathbf{y})\}$ besagt dabei, dass \mathbf{Y} Werte "in der Nähe" von \mathbf{y} annimmt.

Man beachte jedoch, dass voraussetzungsgemäß $P(\mathbf{Y} = \mathbf{y}) = 0$ gilt, und zwar für *alle* $\mathbf{y} \in \mathbb{R}^m$ wegen der Stetigkeit der Verteilung $P^{\mathbf{Y}}$! Eine sinnvolle Definition der bedingten Wahrscheinlichkeiten $P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y})$ gemäß Definition 27 ist in solchen Fällen also zunächst gar nicht möglich, da sämtliche bedingenden Ereignisse die Wahrscheinlichkeit Null besitzen. Dies rechtfertigt nachträglich noch einmal den hier gewählten Zugang zu bedingten Verteilungen über σ -Algebren wie in Definition 41.

Beispiel: X und Y seien unabhängige, exponentialverteilte Zufallsvariablen mit $P^X = \mathcal{E}(\lambda)$, $P^Y = \mathcal{E}(\nu)$ mit $\lambda > \nu > 0$. Für die Dichte der Summe $S = X + Y$ erhalten wir dann mit Lemma 38

$$f_S(z) = \int_0^z f_X(x) \cdot f_Y(z-x) dx = \lambda \nu \int_0^z \exp(-\lambda x - \nu(z-x)) dx$$

$$= \lambda \nu e^{-\nu z} \int_0^z e^{-(\lambda-\nu)x} dx = \frac{\lambda \nu}{\lambda - \nu} (e^{-\nu z} - e^{-\lambda z}), \quad z \geq 0.$$

(Man beachte, dass man für den Grenzübergang $\nu \uparrow \lambda$ erhält: $f_S(z) = \frac{\lambda^2}{2} z e^{-\lambda z}$, $z \geq 0$, also die Dichte der $\mathcal{E}(2, \lambda)$ -Verteilung.) Die gemeinsame Dichte von X und S bzw. Y und S kann man etwa direkt aus Satz 26 berechnen über die Abbildung $g(u, v) = \begin{pmatrix} u \\ u + v \end{pmatrix}$, $u, v > 0$ zu

$$f_{(X,S)}(x, z) = f_{(X,Y)}(x, z-x) = \lambda \nu e^{-\lambda x} e^{-\nu(z-x)} \quad 0 < x, y < z.$$

$$f_{(Y,S)}(y, z) = f_{(X,Y)}(z-y, y) = \lambda \nu e^{-\lambda(z-y)} e^{-\nu y},$$

II. Grundprinzipien stochastischer Modellierung

Für die bedingten Dichten $f_X(\cdot | S = z)$ bzw. $f_Y(\cdot | S = z)$ ergibt sich nun nach Satz 46:

$$f_X(\cdot | S = z) = \frac{f_{(X,S)}(x, z)}{f_S(z)} = \frac{\lambda \nu e^{-\lambda x} e^{-\nu(z-x)}}{\frac{\lambda \nu}{\lambda - \nu} (e^{-\nu z} - e^{-\lambda z})} = (\lambda - \nu) \frac{e^{-(\lambda - \nu)x}}{1 - e^{-(\lambda - \nu)z}}, \quad 0 < x < z,$$

$$f_Y(\cdot | S = z) = \frac{f_{(Y,S)}(y, z)}{f_S(z)} = \frac{\lambda \nu e^{-\lambda(z-y)} e^{-\nu y}}{\frac{\lambda \nu}{\lambda - \nu} (e^{-\nu z} - e^{-\lambda z})} = (\lambda - \nu) \frac{e^{(\lambda - \nu)y}}{e^{(\lambda - \nu)z} - 1}, \quad 0 < y < z.$$

Für den Grenzfall $\nu \uparrow \lambda$ erhält man hieraus noch

$$f_X(\cdot | S = z) = \frac{1}{z}, \quad 0 < x < z, \quad f_Y(\cdot | S = z) = \frac{1}{z}, \quad 0 < y < z,$$

d.h. in diesem Fall gilt: $P^X(\cdot | S = z) = P^Y(\cdot | S = z) = \mathcal{U}[0, z]$, $z > 0$.

Für das (iterierte) Rechnen mit allgemeinen bedingten Verteilungen ist das folgende Ergebnis sehr nützlich:

Satz 48. Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum; $\mathbf{X}, \mathbf{Y}, \mathbf{Z}$ seien Zufallselemente mit Werten in Messräumen $(\mathfrak{X}, \mathcal{B})$, $(\mathfrak{Y}, \mathcal{D})$, $(\mathfrak{Z}, \mathcal{F})$. Dann gilt:

a) Ist U eine reellwertige Zufallsvariable auf $(\mathfrak{Y} \times \mathfrak{Z}, \mathcal{D} \otimes \mathcal{F})$ und $P(\cdot | \mathbf{Z} = \mathbf{z})$, $\mathbf{z} \in \mathfrak{Z}$ regulär, so gilt im Falle der Integrierbarkeit:

$$\int U(\mathbf{y}, \mathbf{z}) P^{(\mathbf{Y}, \mathbf{Z})}(d\mathbf{y}, d\mathbf{z}) = \int \int U(\mathbf{y}, \mathbf{z}) P^{\mathbf{Y}}(d\mathbf{y} | \mathbf{Z} = \mathbf{z}) P^{\mathbf{Z}}(d\mathbf{z})$$

b) Ist $P(\cdot | \mathbf{Z} = \mathbf{z})$, $\mathbf{z} \in \mathfrak{Z}$ regulär, so gilt

$$P(\mathbf{X} \in A, \mathbf{Y} \in B | \mathbf{Z} = \mathbf{z}) = \int_B P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z}) P^{\mathbf{Y}}(d\mathbf{y} | \mathbf{Z} = \mathbf{z})$$

$P^{\mathbf{Z}}$ -fast sicher für alle $A \in \mathcal{B}$, $B \in \mathcal{D}$, $\mathbf{z} \in \mathfrak{Z}$.

c) \mathbf{X} und \mathbf{Y} sind stochastisch unabhängig genau dann, wenn es eine reguläre Version von $P^{\mathbf{X}}(\cdot | \mathbf{Y} = \mathbf{y})$, $\mathbf{y} \in \mathfrak{Y}$ bzw. $P^{\mathbf{Y}}(\cdot | \mathbf{X} = \mathbf{x})$, $\mathbf{x} \in \mathfrak{X}$ gibt, die nicht von \mathbf{y} bzw. von \mathbf{x} abhängt.

Beweis:

a) Sei zunächst $U = \mathbb{1}_{D \times F}$ mit $D \in \mathcal{D}$, $F \in \mathcal{F}$. Dann gilt

II. Grundprinzipien stochastischer Modellierung

$$\int U(\mathbf{y}, \mathbf{z}) P^{(\mathbf{Y}, \mathbf{Z})}(d\mathbf{y}, d\mathbf{z}) = P(\mathbf{Y} \in D, \mathbf{Z} \in F) = \int_F P(\mathbf{Y} \in D | \mathbf{Z} = \mathbf{z}) P^{\mathbf{Z}}(d\mathbf{z})$$

$$\int_F \int_D P^{\mathbf{Y}}(d\mathbf{y} | \mathbf{Z} = \mathbf{z}) P^{\mathbf{Z}}(d\mathbf{z}) = \iint U(\mathbf{y}, \mathbf{z}) P^{\mathbf{Y}}(d\mathbf{y} | \mathbf{Z} = \mathbf{z}) P^{\mathbf{Z}}(d\mathbf{z}).$$

Ähnlich wie in Satz 8 lässt sich hieraus schließen, dass durch

$$Q(D \times F) := \int_F \int_D P^{\mathbf{Y}}(d\mathbf{y} | \mathbf{Z} = \mathbf{z}) P^{\mathbf{Z}}(d\mathbf{z}), \quad D \in \mathcal{D}, F \in \mathcal{F}$$

ein Maß Q auf $\mathcal{D} \otimes \mathcal{F}$ gegeben ist, welches auf dem Erzeuger $\{D \times F | D \in \mathcal{D}, F \in \mathcal{F}\}$ und damit auch auf $\mathcal{D} \otimes \mathcal{F}$ mit dem Maß $P^{(\mathbf{Y}, \mathbf{Z})}$ übereinstimmt. Damit gilt die Aussage allgemeiner sogar für alle Indikatorfunktionen $U = \mathbb{1}_C$ mit $C \in \mathcal{D} \otimes \mathcal{F}$. Unter Ausnutzung der Additivität des Integrals folgt hieraus sofort, dass die Aussage auch für Elementarfunktionen U gilt. Durch monotone Approximation ergibt sich die Aussage dann auch für nicht-negative messbare Abbildungen U und schließlich durch Zerlegung in Positiv- und Negativteil für beliebige integrierbare Abbildungen U .

b) Mit dem gerade Gezeigten ergibt sich einerseits für alle $A \in \mathcal{B}$, $B \in \mathcal{D}$, $F \in \mathcal{F}$:

$$\begin{aligned} P(\mathbf{X} \in A, \mathbf{Y} \in B, \mathbf{Z} \in F) &= \iint_{B \times F} P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z}) P^{(\mathbf{Y}, \mathbf{Z})}(d\mathbf{y}, d\mathbf{z}) \\ &= \int_F \int_B P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z}) P^{\mathbf{Y}}(d\mathbf{y} | \mathbf{Z} = \mathbf{z}) P^{\mathbf{Z}}(d\mathbf{z}), \end{aligned}$$

andererseits

$$P(\mathbf{X} \in A, \mathbf{Y} \in B, \mathbf{Z} \in F) = \int_F P(\mathbf{X} \in A, \mathbf{Y} \in B | \mathbf{Z} = \mathbf{z}) P^{\mathbf{Z}}(d\mathbf{z}).$$

Da diese Beziehung für alle $F \in \mathcal{F}$ gilt, folgt durch Vergleich der Terme die Behauptung (vgl. Definition 41, Punkt 3 und den Faktorisierungssatz II oben).

c) Seien zunächst \mathbf{X} und \mathbf{Y} stochastisch unabhängig. Dann ist $P^{\mathbf{X}}(\cdot | \mathbf{Y} = \mathbf{y}) = P^{\mathbf{X}}$, $\mathbf{y} \in \mathfrak{Y}$ eine solche reguläre Version, denn es gilt für alle $B \in \mathcal{D}$:

$$\int_B P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) P^{\mathbf{Y}}(d\mathbf{y}) = P(\mathbf{X} \in A, \mathbf{Y} \in B) = P(\mathbf{X} \in A) P(\mathbf{Y} \in B) = \int_B P(\mathbf{X} \in A) P^{\mathbf{Y}}(d\mathbf{y}),$$

woraus wieder durch Vergleich beider Seiten die Behauptung folgt. Hängt umgekehrt $Q = P^{\mathbf{X}}(\cdot | \mathbf{Y} = \mathbf{y})$ nicht von $\mathbf{y} \in \mathfrak{Y}$ ab, so erhält man analog

$$P(\mathbf{X} \in A, \mathbf{Y} \in B) = \int_B P(\mathbf{X} \in A | \mathbf{Y} = \mathbf{y}) P^{\mathbf{Y}}(d\mathbf{y}) = \int_B Q(A) P^{\mathbf{Y}}(d\mathbf{y}) = Q(A) P(\mathbf{Y} \in B),$$

woraus $Q = P^{\mathbf{X}}$ und damit auch die umgekehrte Behauptung folgt. ■

II. Grundprinzipien stochastischer Modellierung

Die obigen Ausführungen legen die Frage nahe, ob man wie im diskreten Fall auch allgemeiner "Informationen", die in der rechten Bedingungsseite enthalten sind, in die linke Seite einsetzen darf, wenn diese dort ebenfalls vorkommen. Generell ist das aber nicht so einfach möglich, wie das folgende Ergebnis zeigt.

Satz 49 (Ersetzungsformel). Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum; \mathbf{X} und \mathbf{Y} seien Zufallselemente mit Werten in Messräumen $(\mathcal{X}, \mathcal{B})$ bzw. $(\mathcal{Y}, \mathcal{D})$. Die bedingte Verteilung $P^{\mathbf{X}}(\cdot | \mathbf{Y} = \mathbf{y})$, $\mathbf{y} \in \mathcal{Y}$ sei regulär, $G: (\mathcal{X} \times \mathcal{Y}, \mathcal{B} \otimes \mathcal{D}) \rightarrow (\mathcal{Z}, \mathcal{F})$ eine Abbildung mit Werten in dem Messraum $(\mathcal{Z}, \mathcal{F})$. Dann gilt für alle $A \in \mathcal{F}$:

$$\begin{aligned} P(G(\mathbf{X}, \mathbf{Y}) \in A | \mathbf{Y} = \mathbf{y}) &= P(G(\mathbf{X}, \mathbf{y}) \in A | \mathbf{Y} = \mathbf{y}) \\ &:= P(G(\mathbf{X}, \mathbf{z}) \in A | \mathbf{Y} = \mathbf{y}) \Big|_{\mathbf{z}=\mathbf{y}} \quad P^{\mathbf{Y}}\text{-f.s.} \end{aligned}$$

Sind \mathbf{X} und \mathbf{Y} stochastisch unabhängig, so gilt speziell

$$P(G(\mathbf{X}, \mathbf{Y}) \in A | \mathbf{Y} = \mathbf{y}) = P(G(\mathbf{X}, \mathbf{y}) \in A) \quad P^{\mathbf{Y}}\text{-f.s.}$$

Beweis: Wir zeigen die Aussage zunächst für den speziellen Fall $(\mathcal{Z}, \mathcal{F}) = (\mathcal{X} \times \mathcal{Y}, \mathcal{B} \otimes \mathcal{D})$, $G(\mathbf{x}, \mathbf{y}) = (\mathbf{x}, \mathbf{y}) \in (\mathcal{Z}, \mathcal{F})$ mit $A = B \times D$, $B \in \mathcal{B}$, $D \in \mathcal{D}$. Für beliebige $C \in \mathcal{D}$ ist dann

$$\begin{aligned} \int_C P((\mathbf{X}, \mathbf{Y}) \in A | \mathbf{Y} = \mathbf{y}) P^{\mathbf{Y}}(d\mathbf{y}) &= P((\mathbf{X}, \mathbf{Y}) \in B \times D, \mathbf{Y} \in C) \\ &= P(\mathbf{X} \in B, \mathbf{Y} \in D \cap C) = \int_{D \cap C} P(\mathbf{X} \in B | \mathbf{Y} = \mathbf{y}) P^{\mathbf{Y}}(d\mathbf{y}) \\ &= \int_C \mathbb{1}_D(\mathbf{y}) P(\mathbf{X} \in B | \mathbf{Y} = \mathbf{y}) P^{\mathbf{Y}}(d\mathbf{y}). \end{aligned}$$

Ein Vergleich beider Seiten zeigt also

$$\begin{aligned} P(G(\mathbf{X}, \mathbf{Y}) \in A | \mathbf{Y} = \mathbf{y}) &= P((\mathbf{X}, \mathbf{Y}) \in B \times D | \mathbf{Y} = \mathbf{y}) \\ &= \mathbb{1}_D(\mathbf{y}) P(\mathbf{X} \in B | \mathbf{Y} = \mathbf{y}) = \begin{cases} P(\mathbf{X} \in B | \mathbf{Y} = \mathbf{y}) & \text{für } \mathbf{y} \in D \\ 0 & \text{sonst} \end{cases} \\ &= P((\mathbf{X}, \mathbf{z}) \in B \times D | \mathbf{Y} = \mathbf{y}) \Big|_{\mathbf{z}=\mathbf{y}} = P(G(\mathbf{X}, \mathbf{z}) \in A | \mathbf{Y} = \mathbf{y}) \Big|_{\mathbf{z}=\mathbf{y}} \quad P^{\mathbf{Y}}\text{-f.s.} \end{aligned}$$

Mit einer ähnlichen Begründung wie im Beweis zu Satz 48 folgt dann auch die Gültigkeit der Aussage für beliebige Mengen $A \in \mathcal{B} \otimes \mathcal{D}$. Der generelle Fall ergibt sich hieraus unmittelbar wegen $\{G(\mathbf{X}, \mathbf{Y}) \in A\} = \{(\mathbf{X}, \mathbf{Y}) \in G^{-1}(A)\}$, $A \in \mathcal{F}$.

Im Falle der stochastischen Unabhängigkeit von \mathbf{X} und \mathbf{Y} hat man nur die Gültigkeit von

$$P(G(\mathbf{X}, \mathbf{z}) \in A | \mathbf{Y} = \mathbf{y}) = P(G(\mathbf{X}, \mathbf{z}) \in A) \quad P^{\mathbf{Y}}\text{-f.s.}$$

zu beachten, die aus Satz 48 c) folgt. ■

II. Grundprinzipien stochastischer Modellierung

Man beachte, dass der oben auftretende Ausdruck $P(G(\mathbf{X}, \mathbf{y}) \in A | \mathbf{Y} = \mathbf{y})$ nur im Fall diskreter Verteilungen unmittelbar Sinn macht, während i.Allg. die angegebene Umschreibung mit der "Dummy"-Variablen \mathbf{z} unverzichtbar ist, da der Ausdruck im Sinne von Definition 41 im Kontext des Faktorisierungssatzes II sonst nicht erklärt ist.

Sind beispielsweise X und Y reellwertige Zufallsvariablen und betrachtet man $U := \max(X, Y)$, so ergibt sich mit der Ersetzungsformel

$$\begin{aligned} P(U \leq u | Y = y) &= P(X \leq u, Y \leq u | Y = y) = P(X \leq u, z \leq u | Y = y) \Big|_{z=y} \\ &= \mathbb{1}_{(-\infty, u]}(z) \Big|_{z=y} P(X \leq u | Y = y) = \mathbb{1}_{(-\infty, u]}(y) P(X \leq u | Y = y), \quad u, y \in \mathbb{R}. \end{aligned}$$

Sind X und Y stochastisch unabhängig, so vereinfacht sich dies noch zu

$$P(U \leq u | Y = y) = P(X \leq u, Y \leq u | Y = y) = \mathbb{1}_{(-\infty, u]}(y) P(X \leq u), \quad u, y \in \mathbb{R}.$$

Ist \mathcal{C} eine Teil- σ -Algebra von \mathcal{A} und die bedingte Verteilung $P(\cdot | \mathcal{C})$ regulär, so kann man natürlich alle Begriffsbildungen wie Verteilungsfunktion, Dichte, Erwartungswert usw., die für gewöhnliche Verteilungen existieren, entsprechend auf die bedingte Verteilung übertragen. Im letzten Fall gelangt man dann etwa zum Begriff des bedingten Erwartungswerts, den wir wegen seiner zentralen Bedeutung hier noch einmal gesondert behandeln wollen.

Definition 43 (regulärer bedingter Erwartungswert): Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\mathcal{C} \sqsubseteq \mathcal{A}$ eine Teil- σ -Algebra von \mathcal{A} ; $P(\cdot | \mathcal{C})$ sei regulär. Ist X eine reellwertige Zufallsvariable auf (Ω, \mathcal{A}, P) , so heißt im Falle der Existenz

$$E(X | \mathcal{C})(\omega) := \int X(\eta) P(d\eta | \mathcal{C})(\omega) = \int X dP(\cdot | \mathcal{C})(\omega), \quad \omega \in \Omega,$$

oder kürzer

$$E(X | \mathcal{C}) = \int X dP(\cdot | \mathcal{C})$$

bedingter Erwartungswert von X unter \mathcal{C} . Ist $\mathcal{C} = \sigma(\mathbf{Y})$ für ein Zufallselement \mathbf{Y} mit Werten in einem mit Werten in einem Messraum $(\mathfrak{X}, \mathcal{B})$, so heißt entsprechend auch

$$E(X | \sigma(\mathbf{Y})) = E(X | \mathbf{Y}) \quad \text{bzw.} \quad E(X | \mathbf{Y} = \mathbf{y}), \quad \mathbf{y} \in \mathfrak{X}$$

bedingter Erwartungswert von \mathbf{X} unter \mathbf{Y} bzw. unter $\mathbf{Y} = \mathbf{y}$.

Bemerkung: Insbesondere ist damit $E(X | \mathcal{C})$ eine \mathcal{C} -messbare Abbildung.

II. Grundprinzipien stochastischer Modellierung

Die Ersetzungsformel aus Satz 49 überträgt sich völlig analog auch auf reguläre bedingte Erwartungswerte; man erhält hier, mit den Bezeichnungen von Satz 49:

$$E(G(\mathbf{X}, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}) = E(G(\mathbf{X}, \mathbf{y}) | \mathbf{Y} = \mathbf{y}) := E(G(\mathbf{X}, \mathbf{z}) | \mathbf{Y} = \mathbf{y}) \Big|_{\mathbf{z}=\mathbf{y}} \quad P^{\mathbf{Y}}\text{-f.s.}$$

Im Falle der Unabhängigkeit von \mathbf{X} und \mathbf{Y} ergibt sich entsprechend

$$E(G(\mathbf{X}, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}) = E(G(\mathbf{X}, \mathbf{y})) \quad P^{\mathbf{Y}}\text{-f.s.}$$

Im Falle der bedingten quadratischen Integrierbarkeit kann man auch die *bedingte Varianz* einer Zufallsvariablen definieren als Varianz bezüglich der entsprechenden bedingten Verteilung, in Zeichen: $Var(X | \mathcal{C})$.

Lemma 49 (Zerlegungslemma): Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\mathcal{C} \sqsubseteq \mathcal{A}$ eine Teil- σ -Algebra von \mathcal{A} ; $P(\cdot | \mathcal{C})$ sei regulär. Ist die reellwertige Zufallsvariable X auf (Ω, \mathcal{A}, P) integrierbar, so ist sie auch bezüglich $P(\cdot | \mathcal{C})$ integrierbar, und es gilt

$$E(X) = E(E(X | \mathcal{C})) = \int E(X | \mathcal{C}) dP.$$

Ist ferner X sogar quadratisch integrierbar, so auch bezüglich $P(\cdot | \mathcal{C})$, und es gilt

$$Var(X) = E(Var(X | \mathcal{C})) + Var(E(X | \mathcal{C})).$$

Beweis: Sei zunächst $X = \mathbb{1}_A$ mit $A \in \mathcal{A}$. Dann ist X bezüglich $P(\cdot | \mathcal{C})$ integrierbar mit

$$E(X | \mathcal{C}) = \int X dP(\cdot | \mathcal{C}) = \int \mathbb{1}_A dP(\cdot | \mathcal{C}) = P(A | \mathcal{C}) \text{ sowie}$$

$$E(E(X | \mathcal{C})) = \int E(X | \mathcal{C}) dP = \int P(A | \mathcal{C}) dP = P(A) = E(X)$$

nach Definition 41. Unter Ausnutzung der Linearität des Integrals folgt hieraus, dass die Aussage auch für Elementarfunktionen richtig bleibt. Für nicht-negative X bleibt die Aussage damit ebenfalls richtig (nach Anwendung des Satzes von der monotonen Konvergenz); durch Zerlegung von X in Positiv- und Negativteil ergibt sich damit die Aussage im Allgemeinen.

Der zweite Teil des Lemmas ergibt sich so:

$$\begin{aligned} E(Var(X | \mathcal{C})) &= E\left[E(X^2 | \mathcal{C}) - \{E(X | \mathcal{C})\}^2\right] \\ Var(E(X | \mathcal{C})) &= E\left[\{E(X | \mathcal{C})\}^2\right] - [E(E(X | \mathcal{C}))]^2 \end{aligned}$$

Die Summation beider Terme ergibt

II. Grundprinzipien stochastischer Modellierung

$$\begin{aligned}
 E(\text{Var}(X|C)) + \text{Var}(E(X|C)) &= E\left[E(X^2|C) - \{E(X|C)\}^2\right] + E\left[\{E(X|C)\}^2\right] - [E(E(X|C))]^2 \\
 &= E\left[E(X^2|C)\right] - E\left[\{E(X|C)\}^2\right] + E\left[\{E(X|C)\}^2\right] - [E(E(X|C))]^2 \\
 &= E\left[E(X^2|C)\right] - [E(E(X|C))]^2 = E(X^2) - [E(X)]^2 = \text{Var}(X)
 \end{aligned}$$

und damit die Behauptung. ■

Das folgende Beispiel zeigt, dass Zufallsvariablen bezüglich der bedingten Verteilung $P(\cdot|C)$ (sogar quadratisch) integrierbar sein können, ohne dass dies für die unbedingte Verteilung zutrifft:

Seien dazu X und Y stochastisch unabhängige, jeweils $\mathcal{U}(0,1)$ -verteilte Zufallsvariablen. Setzt man $Z := -\frac{1}{Y} \ln X$, so ist die bedingte Verteilung $P^Z(\cdot|Y=y) = \mathcal{E}(y)$ für jedes $y \in (0,1)$ eine Exponentialverteilung; man erhält also

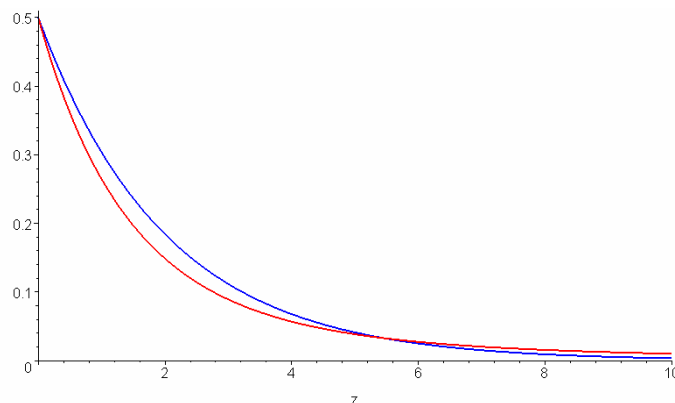
$$E(Z|Y=y) = \frac{1}{y} < \infty \quad \text{und} \quad \text{Var}(Z|Y=y) = \frac{1}{y^2} < \infty \quad \text{für jedes } y \in (0,1);$$

jedoch ist $E(Z) = \int E(Z|Y=y)P^Y(y) = \int_0^1 \frac{1}{y} f_Y(y) dy = \int_0^1 \frac{1}{y} dy = \infty$, d.h. $E(Z)$ existiert nicht als endliche Größe, somit existiert auch keine (unbedingte) Varianz. Für die Verteilungsfunktion von Z erhält man übrigens noch den Ausdruck

$$F_Z(z) = \int_0^1 P(Z \leq z | Y=y) P^Y(dy) = \int_0^1 (1 - e^{-yz}) dy = 1 - \frac{1 - e^{-z}}{z}, \quad z > 0.$$

Für die Dichte ergibt sich hieraus durch Differenzieren

$$f_Z(z) = \frac{e^{-z}(1+z) - 1}{z^2}, \quad z > 0.$$



Plot der Dichte von Z (rot) im Vergleich zur Dichte der $\mathcal{E}\left(\frac{1}{2}\right)$ -Verteilung (blau)

II. Grundprinzipien stochastischer Modellierung

Es gibt prinzipiell auch die Möglichkeit, abstrakte bedingte Erwartungen direkt mit Hilfe des Satzes 21 (von Radon-Nikodym) einzuführen. Sei dazu (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\mathcal{C} \subseteq \mathcal{A}$ eine Teil- σ -Algebra von \mathcal{A} . Ist nun X eine integrierbare Zufallsvariable auf (Ω, \mathcal{A}, P) , so wird durch

$$Q^+(C) := \int_C X^+ dP \quad \text{und} \quad Q^-(C) := \int_C X^- dP \quad \text{für alle } C \in \mathcal{C}$$

jeweils ein (endliches) Maß auf \mathcal{C} definiert. Bezeichnet nun P^C die Einschränkung von P auf \mathcal{C} und betrachten wir den (kleineren) Wahrscheinlichkeitsraum $(\Omega, \mathcal{C}, P^C)$, so dominiert offenbar P^C sowohl Q^+ als auch Q^- . Folglich existieren \mathcal{C} -messbare Dichten f^+ und f^- mit

$$Q^+(C) = \int_C X^+ dP = \int_C f^+ dP^C = \int_C f^+ dP \quad \text{und} \quad Q^-(C) = \int_C X^- dP = \int_C f^- dP^C = \int_C f^- dP$$

für alle $C \in \mathcal{C}$. Mit der Setzung

$$E(X | \mathcal{C}) := f^+ - f^-$$

erhält man dann die gewünschte abstrakte Definition der bedingten Erwartung, die im Falle der Existenz einer regulären bedingten Verteilung $P(\cdot | \mathcal{C})$ mit dem vorher eingeführten Begriff der regulären bedingten Erwartung zusammenfällt. Man beachte aber, dass die zuletzt eingeführte abstrakte bedingte Erwartung nur P^C -fast sicher eindeutig bestimmt ist. Man spricht deshalb auch von den verschiedenen (fast sicher gleichen) *Versionen* der bedingten Erwartung. Jede solche Version $E(X | \mathcal{C})$ ist also eine \mathcal{C} -messbare Lösung der Integralgleichung

$$\int_C X dP = \int_C E(X | \mathcal{C}) dP \quad \text{für alle } C \in \mathcal{C};$$

vgl. Definition 41. Vermöge der Setzung

$$P(A | \mathcal{C}) := E(\mathbb{1}_A | \mathcal{C}) \quad \text{für alle } A \in \mathcal{A}$$

erhält man dann auch noch allgemeine bedingte Wahrscheinlichkeiten, die ebenfalls nur fast sicher eindeutig bestimmt sind. Im Allgemeinen ist die bedingte Verteilung $P(\cdot | \mathcal{C})$ aber nicht regulär.

Abschließend stellen wir noch einige wichtige Eigenschaften (regulärer) bedingter Erwartungen zusammen, die vor allem im Zusammenhang mit Stochastischen Prozessen (Markoff-Ketten, Markoff-Prozesse, Martingale) wichtig sind.

II. Grundprinzipien stochastischer Modellierung

Satz 50 (Eigenschaften [regulärer] bedingter Erwartungen): Es sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\mathcal{C} \subseteq \mathcal{A}$ eine Teil- σ -Algebra von \mathcal{A} ; die reellwertige Zufallsvariable X auf (Ω, \mathcal{A}, P) sei integrierbar. Dann gilt:

- a) Ist Y eine weitere reellwertige Zufallsvariable auf (Ω, \mathcal{A}, P) , die \mathcal{C} -messbar ist und für die das Produkt $X \cdot Y$ integrierbar ist, dann gilt:

$$E(X \cdot Y | \mathcal{C}) = Y \cdot E(X | \mathcal{C}) \quad (\text{Faktor-Regel})$$

Insbesondere gilt im Fall $X \equiv 1$:

$$E(Y | \mathcal{C}) = Y$$

- b) Ist $\mathcal{D} \subseteq \mathcal{C}$ eine weitere Teil- σ -Algebra von \mathcal{A} , so gilt:

$$E[E(X | \mathcal{C}) | \mathcal{D}] = E[E(X | \mathcal{D}) | \mathcal{C}] = E(X | \mathcal{D}) \quad (\text{Turm-Eigenschaft})$$

- c) Sind die σ -Algebren $\sigma(X)$ und \mathcal{C} stochastisch unabhängig (im Sinne von Definition 26), so gilt

$$E(X | \mathcal{C}) = E(X) \quad (\text{Unabhängigkeits-Eigenschaft})$$

Insbesondere gilt im Fall $\mathcal{C} = \{\emptyset, \Omega\}$:

$$E(X | \mathcal{C}) = E(X).$$

Beweis:

- a) Für beliebige $A \in \mathcal{A}$ und $C, D \in \mathcal{C}$ gilt (vgl. Definition 41):

$$\int_D \mathbb{1}_C \cdot P(A | \mathcal{C}) dP = \int_{C \cap D} P(A | \mathcal{C}) dP = P(A \cap C \cap D) = \int_D P(A \cap C | \mathcal{C}) dP,$$

d.h. es gilt

$$E(\mathbb{1}_A \mathbb{1}_C | \mathcal{C}) = E(\mathbb{1}_{A \cap C} | \mathcal{C}) = P(A \cap C | \mathcal{C}) = \mathbb{1}_C \cdot P(A | \mathcal{C}) = \mathbb{1}_C \cdot E(\mathbb{1}_A | \mathcal{C}),$$

womit die Aussage zunächst für den Fall $X = \mathbb{1}_A$ und $Y = \mathbb{1}_C$ gezeigt ist. Unter Ausnutzung der Linearität des Integrals bleibt die Aussage dann aber auch für entsprechend messbare Elementarfunktionen X und Y richtig. Mit Hilfe des Satzes von der monotonen Konvergenz schließt man dann weiter auf die Gültigkeit der Aussage für nicht-negative X, Y und schließlich durch Zerlegung in Positiv- und Negativteil auf den allgemeinen Fall.

Für $X \equiv 1$ ergibt sich speziell

$$E(Y | \mathcal{C}) = E(X \cdot Y | \mathcal{C}) = Y \cdot E(X | \mathcal{C}) = Y \cdot E(\mathbb{1}_\Omega | \mathcal{C}) = Y \cdot P(\Omega | \mathcal{C}) = Y.$$

- b) Nach Voraussetzung ist $E(X | \mathcal{D})$ als \mathcal{D} -messbare Abbildung auch \mathcal{C} -messbar, so dass mit Teil a) folgt

$$E[E(X | \mathcal{D}) | \mathcal{C}] = E(X | \mathcal{D}) \cdot E(1 | \mathcal{C}) = E(X | \mathcal{D}).$$

Andererseits folgt für jedes $D \in \mathcal{D} \subseteq \mathcal{C}$

II. Grundprinzipien stochastischer Modellierung

$$\int_D E(X|C) dP = \int_D X dP = \int_D E(X|D) dP, \text{ also auch } E[E(X|C)|D] = E(X|D),$$

woraus die Aussage folgt.

- c) Aus der Unabhängigkeitsannahme folgt, dass auch X und $\mathbb{1}_C$ für alle $C \in \mathcal{C}$ stochastisch unabhängig sind. Damit gilt:

$$\int_C E(X) dP = E(X) \cdot P(C) = E(X) \cdot E(\mathbb{1}_C) = E(X \cdot \mathbb{1}_C) = \int_C X dP = \int_C E(X|C) dP$$

für alle $C \in \mathcal{C}$, also $E(X|C) = E(X)$, wie behauptet.

Die σ -Algebren $\sigma(X)$ und $\mathcal{C} = \{\emptyset, \Omega\}$ sind trivialerweise stochastisch unabhängig.

Damit ist Satz 50 bewiesen. ■

Für das diskrete Beispiel auf S. 91 ergibt sich mit der Faktor-Regel wegen $E(X|Y) = \frac{Y+1}{2}$ entsprechend dem Ergebnis auf S. 102:

$$E(Z) = E\left[E\left(\frac{X}{Y} \middle| Y\right)\right] = E\left[\frac{1}{Y} E(X|Y)\right] = E\left(\frac{Y+1}{2Y}\right) = \frac{1}{2} + \frac{1}{2} E\left(\frac{1}{Y}\right) = \frac{1}{2} + \frac{1}{2} \sum_{n=1}^{\infty} \frac{1}{n 2^n} = \frac{1 + \ln(2)}{2}.$$

(Man beachte, hierbei, dass die Zufallsvariable $\frac{1}{Y}$ \mathcal{C} -messbar ist für $\mathcal{C} = \sigma(Y)$.)

II.9. Die mehrdimensionale Normalverteilung

In diesem Abschnitt wollen wir kurz die multivariate Normalverteilung vorstellen, die nicht nur für die stochastische Modellierung, sondern auch für die Statistik eine fundamentale Rolle spielt.

Definition 44 (mehrdimensionale Normalverteilung): Ein d -dimensionaler Zufallsvektor $X = (X_1, \dots, X_d)^{tr}$ mit Werten in $(\mathbb{R}^d, \mathcal{B}^d)$ besitzt eine d -dimensionale (nicht-entartete) Normalverteilung $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$, wenn er bezüglich des Lebesgue-Maßes m^d die Dichte

$$f(x_1, \dots, x_d) = \frac{1}{\sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^{tr} \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\right)$$

mit $\mathbf{x} = (x_1, \dots, x_d)^{tr} \in \mathbb{R}^d$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)^{tr} \in \mathbb{R}^d$ besitzt, wobei Σ eine symmetrische positiv-definite $d \times d$ -Matrix ist.

II. Grundprinzipien stochastischer Modellierung

Bemerkung: Die Matrix Σ enthält die (paarweisen) Kovarianzen der Komponenten von \mathbf{X} , d.h. es gilt

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \cdots & \sigma_d^2 \end{bmatrix}$$

mit $\sigma_k^2 = \text{Var}(X_k)$, $k = 1, \dots, d$ und $\sigma_{ij} = \text{Kov}(X_i, X_j)$ für $1 \leq i < j \leq d$. Sie heißt deshalb auch die *Varianz-Kovarianz-Matrix* der Verteilung. Der Vektor $\boldsymbol{\mu}$ heißt entsprechend *Mittelwert- oder Erwartungswertvektor* der Verteilung.

Wegen der positiven Definitheit von Σ existiert (mindestens) eine quadratische invertierbare Matrix A mit $\Sigma = A \cdot A^T$. Diese Eigenschaft kann man z.B. zur Simulation $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ -verteilter Zufallsvektoren verwenden. Es gibt mehrere Möglichkeiten, eine solche Matrix A zu bestimmen; im Allgemeinen existieren sogar mehrere Lösungen. Ein Weg besteht in der *Spektralzerlegung* der Matrix Σ : voraussetzungsgemäß besitzt Σ d (ggf. mit Vielfachheiten vorkommende) positive *Eigenwerte* $\lambda_1, \dots, \lambda_d$; ferner existiert eine $d \times d$ -Matrix T aus orthonormalen *Eigenvektoren* mit der Eigenschaft

$$\Sigma = T \Delta T^T = T \Delta T^{-1} \text{ mit } \Delta = \begin{bmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & \vdots \\ \vdots & \vdots & & \lambda_{d-1} & 0 \\ 0 & 0 & \cdots & 0 & \lambda_d \end{bmatrix}.$$

Hier ist $A = T \Delta^{1/2}$ eine mögliche Wahl mit $\Delta^{1/2} = \begin{bmatrix} \sqrt{\lambda_1} & 0 & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & 0 & \cdots & 0 \\ 0 & 0 & \ddots & \cdots & \vdots \\ \vdots & \vdots & & \sqrt{\lambda_{d-1}} & 0 \\ 0 & 0 & \cdots & 0 & \sqrt{\lambda_d} \end{bmatrix}.$

Eine andere Möglichkeit besteht in der *Cholesky-Zerlegung* von Σ ; Die gesuchte Matrix A wird dabei als *untere Dreiecksmatrix* angenommen:

$$A = \begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{bmatrix}.$$

II. Grundprinzipien stochastischer Modellierung

Hiermit ergibt sich

$$\Sigma = [\sigma_{ij}] = AA^r = \begin{bmatrix} a_{11}^2 & a_{11}a_{21} & \cdots & a_{11}a_{d1} \\ a_{21}a_{11} & a_{21}^2 + a_{22}^2 & \cdots & a_{21}a_{d1} + a_{22}a_{d2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1}a_{11} & a_{d1}a_{21} + a_{d2}a_{22} & \cdots & \sum_{k=1}^d a_{dk}^2 \end{bmatrix}.$$

Diese Gleichung kann rekursiv aufgelöst werden zu

$$a_{11} = \sqrt{\sigma_{11}}, \quad a_{kk} = \sqrt{\sigma_{kk} - \sum_{i=1}^{k-1} a_{ki}^2}, \quad a_{k1} = \frac{\sigma_{k1}}{a_{11}}, \quad a_{kj} = \frac{\sigma_{kj} - \sum_{i=1}^{j-1} a_{ki}a_{ji}}{a_{jj}}, \quad 1 \leq k, j \leq d.$$

Beispiel: Für $\Sigma = \begin{bmatrix} 5 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 5 \end{bmatrix}$ erhält man mit der Einheitsmatrix \mathbf{I} das charakteristische Polynom

$$\varphi(\lambda) = \det(\Sigma - \lambda\mathbf{I}) = -\lambda^3 + 11\lambda^2 - 11\lambda + 1$$

mit den drei Nullstellen

$$\lambda_1 = 1, \quad \lambda_{2,3} = 5 \pm 2\sqrt{6}$$

und der zugehörigen Orthonormalmatrix der Eigenvektoren

$$T = \begin{bmatrix} -0,7071 & 0,6739 & -0,2142 \\ 0 & 0,3029 & 0,9530 \\ 0,7071 & 0,6739 & -0,2142 \end{bmatrix}.$$

Hiermit ergibt sich

$$A = T\Delta^{1/2} = \begin{bmatrix} -0,7071 & 2,1202 & -0,0681 \\ 0 & 0,9530 & 0,3029 \\ 0,7072 & 2,1202 & -0,0681 \end{bmatrix}.$$

Mit der Cholesky-Zerlegung erhält man stattdessen

$$A = \begin{bmatrix} \sqrt{5} & 0 & 0 \\ \frac{2}{5}\sqrt{5} & \frac{1}{5}\sqrt{5} & 0 \\ \frac{4}{5}\sqrt{5} & \frac{2}{5}\sqrt{5} & 1 \end{bmatrix} = \begin{bmatrix} 2,2361 & 0 & 0 \\ 0,8944 & 0,4472 & 0 \\ 1,7889 & 0,8944 & 1 \end{bmatrix}.$$

II. Grundprinzipien stochastischer Modellierung

Die Gleichung $\Sigma = AA^T$ besitzt im Allgemeinen noch weitere Lösungen, hier z.B.

$$A = \begin{bmatrix} 1 & 2 & 0 \\ 0 & 1 & 0 \\ 0 & 2 & 1 \end{bmatrix}.$$

Satz 51 (Eigenschaften der multivariaten Normalverteilung I): Es sei \mathbf{X} ein $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ -verteilter Zufallsvektor. Dann ist der Zufallsvektor $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}$ $\mathcal{N}(\mathbf{0}, \Sigma)$ -verteilt. Ist umgekehrt $X \mathcal{N}(\mathbf{0}, \Sigma)$ -verteilt, dann ist $\mathbf{Y} = \mathbf{X} + \boldsymbol{\mu}$ $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ -verteilt.

Beweis: Folgt sofort aus dem Transformationssatz 26 mit der Abbildung $g(\mathbf{x}) = \mathbf{x} - \boldsymbol{\mu}$ bzw. $g^{-1}(\mathbf{y}) = \mathbf{y} + \boldsymbol{\mu}$. ■

Satz 52 (Eigenschaften der multivariaten Normalverteilung II): Es sei \mathbf{X} ein d -dimensionaler $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ -verteilter Zufallsvektor und B eine beliebige $m \times d$ -Matrix mit $m < d$ und vollem Rang m . Dann ist der m -dimensionale Zufallsvektor $\mathbf{Y} = B\mathbf{X}$ ebenfalls multivariat normalverteilt mit $P^{\mathbf{Y}} = \mathcal{N}(B\boldsymbol{\mu}, B\Sigma B^T)$. Im speziellen Fall $m = 1$ bedeutet dies, dass jede Linearkombination multivariat normalverteilter Zufallsvariablen wieder (univariat) normalverteilt ist. Insbesondere ist jede Komponente X_k von \mathbf{X} selbst univariat normalverteilt mit Erwartungswert μ_k und Varianz σ_k^2 für $k = 1, \dots, d$.

Beweis: Im Fall $m = d$ kann wieder der Transformationssatz 26 angewendet werden, und zwar zunächst mit $\boldsymbol{\mu} = \mathbf{0}$, mit der Abbildung $g(\mathbf{x}) = B\mathbf{x}$ und $g^{-1}(\mathbf{y}) = B^{-1}\mathbf{y}$. Hier ist dann $\Delta g(\mathbf{x}) = B$. Der Zufallsvektor $\mathbf{Y} = B\mathbf{X}$ hat damit die Dichte

$$\begin{aligned} h(x_1, \dots, x_d) &= \frac{1}{\det B \sqrt{(2\pi)^d \det \Sigma}} \exp\left(-\frac{1}{2} (B^{-1}\mathbf{x})^T \Sigma^{-1} (B^{-1}\mathbf{x})\right) \\ &= \frac{1}{\sqrt{(2\pi)^d \det(B\Sigma B^T)}} \exp\left(-\frac{1}{2} \mathbf{x}^T (B\Sigma B^T)^{-1} \mathbf{x}\right) \end{aligned}$$

für $\mathbf{x} \in \mathbb{R}^d$. Im allgemeinen Fall kann die Matrix B zu einer invertierbaren $d \times d$ -Matrix C ergänzt werden, die Aussage ergibt sich dann durch Teilauswahl der Komponenten des transformierten Zufallsvektors $\mathbf{Y} = C\mathbf{X}$. Für die Situation mit beliebigem $\boldsymbol{\mu}$ argumentiert man mit Satz 51.

Die zweite Aussage ist trivial, die dritte ergibt sich daraus, indem man in der $1 \times d$ -Matrix B (also eigentlich ein Zeilenvektor) in der k -ten Komponente eine 1 wählt und sonst Nullen. ■

Bemerkung: Eine wesentliche Folgerung aus Satz 52 ist, dass jeder $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ -verteilte Zufallsvektor durch eine affin-lineare Transformation von d stochastisch unabhängigen Standardnormalverteilten Zufallsvariablen dargestellt werden kann: ist nämlich $\mathbf{Y} \mathcal{N}(\mathbf{0}, \mathbf{I})$ -verteilt, mit der Einheitmatrix \mathbf{I} (d.h. alle Komponenten von \mathbf{Y} sind paarweise unkorreliert und damit hier auch sto-

II. Grundprinzipien stochastischer Modellierung

chastisch unabhängig), und ist A eine invertierbare Matrix mit $\Sigma = AA^T$, so ist der Zufallsvektor $\mathbf{X} = A\mathbf{Y} + \boldsymbol{\mu}$ nach den Sätzen 51 und 52 ebenfalls (multivariat) normalverteilt nach $\mathcal{N}(\boldsymbol{\mu}, A\mathbf{I}A^T) = \mathcal{N}(\boldsymbol{\mu}, \Sigma)$.

Achtung: die Eigenschaft, dass Unkorreliertheit die stochastische Unabhängigkeit nach sich zieht, ist im Allgemeinen falsch und gilt bis auf Trivialfälle nur bei multivariater Normalverteilung!

Lemma 50: Für stochastisch unabhängige multivariat normalverteilte Zufallsvektoren \mathbf{X} und \mathbf{Y} derselben Dimension d gilt folgende Rechenregel:

$$P^{\mathbf{X}} = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}}, \Sigma_{\mathbf{X}}) \text{ und } P^{\mathbf{Y}} = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{Y}}) \Rightarrow P^{\mathbf{X}+\mathbf{Y}} = \mathcal{N}(\boldsymbol{\mu}_{\mathbf{X}} + \boldsymbol{\mu}_{\mathbf{Y}}, \Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Y}}).$$

Beweis: Wir setzen $\mathbf{Z} = \begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix}$; dann ist \mathbf{Z} ein $2d$ -dimensionaler Zufallsvektor mit einer $\mathcal{N}(\boldsymbol{\mu}, \Sigma)$ -

Verteilung mit $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_{\mathbf{X}} \\ \boldsymbol{\mu}_{\mathbf{Y}} \end{pmatrix}$ und $\Sigma = \begin{bmatrix} \Sigma_{\mathbf{X}} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{Y}} \end{bmatrix}$. Mit der $d \times 2d$ -Matrix $B = [\mathbf{I} \ \mathbf{I}]$ folgt dann

$B\mathbf{Z} = \mathbf{X} + \mathbf{Y}$; die Aussage ergibt sich dann nach Satz 52. Beachte:

$$B\Sigma B^T = [\mathbf{I} \ \mathbf{I}] \begin{bmatrix} \Sigma_{\mathbf{X}} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\mathbf{Y}} \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} = [\Sigma_{\mathbf{X}} \ \Sigma_{\mathbf{Y}}] \begin{bmatrix} \mathbf{I} \\ \mathbf{I} \end{bmatrix} = \Sigma_{\mathbf{X}} + \Sigma_{\mathbf{Y}}. \quad \blacksquare$$

Man spricht allgemeiner auch dann noch von einer multivariaten (degenerierten) Normalverteilung, wenn Σ die Form $\Sigma = AA^T$ hat, die Matrix A aber nicht invertierbar ist. In diesem Fall existiert keine Dichte der Verteilung im üblichen Sinne. Die Verteilung ist hier konzentriert auf einen niedriger dimensional affinen Unterraum von \mathbb{R}^d vom Lebesgue-Maß Null. Für Einzelheiten sei auf die Monographie von FAHRMEIR ET AL. (1996) verwiesen, dort sind auch noch weitere Eigenschaften der multivariaten Normalverteilung (z.B. für bedingte Verteilungen) angegeben.

III. Grundprinzipien der Statistik

Historisch geht das Wort Statistik wohl auf das lateinische „*statisticum*“ („den Staat betreffend“) zurück; es wird manchmal auch dem italienischen „*statista*“ („Staatsmann“ oder „Politiker“) zugeschrieben. Die deutsche Statistik, eingeführt von *Gottfried Achenwall* (1719 – 1772), bezeichnete ursprünglich die Lehre von den Daten über den Staat, also eine Art Staatstheorie. Mit seinem Werk *Abriss der neuesten Staatswissenschaft der vornehmsten Europäischen Reiche und Republiken* von 1749 gab er als einer der Ersten der Statistik einen wissenschaftlichen Charakter. Das englische Pendant *statistics* wurde erstmals im 19. Jahrhundert von dem Engländer *Sir John Sinclair* (1754 – 1835) in seiner heutigen Bedeutung des allgemeinen Sammelns und Auswertens von Daten benutzt. In diese Zeit (1834) fällt auch die Gründung der berühmten, später *Royal Statistical Society* genannten Gesellschaft in London. Die Entwicklung der eigentlichen *mathematischen Statistik* wurde sehr stark von *Karl Pearson* (1857 – 1936), dem Begründer der *Biometrie*, und *William Sealy Gosset* (1876 – 1937) beeinflusst; den bekannten χ^2 -Anpassungstest publizierte Pearson schon im Jahr 1900. Gosset arbeitete eine Zeit lang bei der berühmten Guinness-Brauerei, die seinerzeit ein angesehenes agro-chemisches Unternehmen war. Den Angestellten von Guinness war es allerdings wegen einer vorgekommenen Indiskretion strikt untersagt, wissenschaftliche Arbeiten zu publizieren, auch solche nicht, die keinen Bezug zu Guinness hatten. Um dieser Auflage zu entgehen, publizierte Gosset seine Arbeiten unter dem Pseudonym *Student*, das auch heute noch eine Reihe statistischer Methoden und Begriffe kennzeichnet (so etwa die *Student'sche t-Verteilung*). Einen großen Schub erhielt die Weiterentwicklung der mathematischen Statistik durch *Ronald Aylmer Fisher* (1890 – 1962) mit dem 1922 erschienenen Werk *On the mathematical foundations of theoretical statistics*, in dem neben der *Maximum-Likelihood-Methode* auch schon die später maßtheoretisch fundierten Begriffe *Konsistenz* und *Suffizienz* diskutiert wurden. *Jerzy Neyman* (1894 – 1981) und *Egon Pearson* (1895 – 1980), Sohn von Karl Pearson, waren weitere wichtige Protagonisten der Fortentwicklung der Statistik. Nach ihnen ist das fundamentale *Neyman-Pearson-Lemma* der Testtheorie benannt, das die Autoren 1928 in ihrer gemeinsamen Arbeit *On the Use and Interpretation of Certain Test Criteria for Purposes of Statistical Inference* vorstellten. 1933 behandelten sie erstmalig das Problem der *Effizienz* in der Testtheorie.

Eine zentrale Aufgabe der Statistik besteht in der Entwicklung wahrscheinlichkeitstheoretisch fundierter Methoden, mit denen man aus Beobachtungen ("Daten"), die zufallsbehafteten Vorgängen [der realen Welt] entstammen, auf die ihnen zu Grunde liegenden Gesetzmäßigkeiten schließen kann. Ein erster Ansatz in diese Richtung ist das schon auf Jakob Bernoulli im 17. Jahrhundert zurückgehende *Gesetz der großen Zahlen* in der Form des Lemmas 45, welches es erlaubt, den Parameter p der den Experimenten zu Grunde liegenden Binomialverteilung durch relative Häufigkeiten zu "schätzen". Diese Idee werden wir in dem nächsten Abschnitt aufgreifen und verallgemeinern.

Definition 45 (statistisches Modell): Es sei $(\mathcal{X}, \mathcal{B})$ ein Messraum. Ein *statistisches Modell* ist das Tripel $(\mathcal{X}, \mathcal{B}, \{Q_\vartheta\}_{\vartheta \in \Theta})$, bei dem Θ eine nicht-leere *Parametermenge* und jedes Q_ϑ , $\vartheta \in \Theta$ eine Wahrscheinlichkeitsverteilung über \mathcal{B} ist.

Dahinter verbirgt sich die Vorstellung, dass $(\mathcal{X}, \mathcal{B})$ der Bildraum einer Zufallsvariablen oder eines Zufallsvektors X auf einem geeigneten (nicht notwendig explizit bekannten) Wahrscheinlichkeitsraum mit Grundmenge Ω und σ -Algebra \mathcal{A} ist, der als Realisationen $X(\omega)$ die "Daten" liefert (vgl. hierzu auch die ausführliche Bemerkung im Anschluss an Satz 12). Im ersten Fall ist

III. Grundprinzipien der Statistik

$(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^1, \mathcal{B}^1)$, im zweiten $(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}^d)$ mit $d \in \mathbb{N}$. Dem Grundraum ist dann ebenfalls ein statistisches Modell zugeordnet, nämlich $(\Omega, \mathcal{A}, \{P_\vartheta\}_{\vartheta \in \Theta})$ mit

$$Q_\vartheta = P_\vartheta^X, \vartheta \in \Theta.$$

Wir sprechen dann von einem statistischen Modell für die Zufallsvariable bzw. den Zufallsvektor X .

Typischerweise sind die Parametermengen Θ Borel-messbare Teilmengen von \mathbb{R}^1 oder \mathbb{R}^m mit $m \in \mathbb{N}$ [etwa (ggf. auch unendliche) Intervalle]. Im ersten Fall spricht man dann auch von einem eindimensionalen, im zweiten Fall von einem m -dimensionalen Parameter (bzw. von einem statistischen Modell mit einem bzw. m Parametern).

Motiviert durch das Gesetz der großen Zahlen werden häufig statistische Modelle betrachtet, die durch unabhängige Wiederholung von Einzelexperimenten entstehen, denen selbst jeweils ein statistisches Modell zu Grunde liegt.

Definition 46 (statistisches Produkt-Modell): Es sei $(\mathfrak{X}, \mathcal{B}, \{Q_\vartheta\}_{\vartheta \in \Theta})$ ein statistisches Modell und

$n \in \mathbb{N}$. Dann heißt $\left(\prod_{i=1}^n \mathfrak{X}, \prod_{i=1}^n \mathcal{B}, \left\{ \prod_{i=1}^n Q_\vartheta \right\}_{\vartheta \in \Theta} \right)$ das zugehörige n -fache *statistische Produkt-Modell*.

$\left(\prod_{i=1}^{\infty} \mathfrak{X}, \prod_{i=1}^{\infty} \mathcal{B}, \left\{ \prod_{i=1}^{\infty} Q_\vartheta \right\}_{\vartheta \in \Theta} \right)$ heißt entsprechend das *unendliche statistische Produkt-Modell*.

Um die statistische Methodik und Berechenbarkeit zu vereinfachen, wird in der Regel angenommen, dass die Q_ϑ , $\vartheta \in \Theta$ je eine Dichte bezüglich eines gemeinsamen Maßes μ auf \mathcal{B} bzw. auf

$\prod_{i=1}^n \mathcal{B}$ besitzen. Ein solches Maß heißt dann *dominierendes Maß*. Die wichtigsten Anwendungsfälle betreffen die folgenden beiden Situationen:

- Die Menge \mathfrak{X} ist abzählbar mit $\mathcal{B} = \mathfrak{P}(\mathfrak{X})$; das dominierende Maß ist das abzählende Maß $\mu = \#$ (vgl. Definition 2) [diskreter Fall].
- Es ist $(\mathfrak{X}, \mathcal{B}) = (\mathbb{R}^1, \mathcal{B}^1)$; das dominierende Maß ist das Lebesgue-Maß $\mu = m^1$ [stetiger Fall].

Für das Produkt-Modell gilt entsprechendes; das dort dominierende Maß ist dann das Produktmaß

$\prod_{i=1}^n \mu$ (also wieder das abzählende Maß [auf der (Produkt-) σ -Algebra $\mathfrak{P}\left(\prod_{i=1}^n \mathfrak{X}\right)$] bzw. das Lebesgue-Maß m^n auf der Borel'schen σ -Algebra \mathcal{B}^n).

III. Grundprinzipien der Statistik

Die folgende Tabelle listet einige bekannte (eindimensionale) statistische Modelle auf:

diskreter Fall:

P^X	Name	(Zähl-)Dichte $f(k) = P(X = k)$ bzgl. #	Θ
$B(n, \vartheta)$	Binomialverteilung	$\binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}, k = 0, \dots, n$	$[0, 1]$
$NB(\beta, \vartheta)$	negative Binomialverteilung	$\binom{\beta+k-1}{k} \vartheta^\beta (1-\vartheta)^k, k \in \mathbb{Z}^+; \beta > 0$	$(0, 1]$
$\mathcal{P}(\vartheta)$	Poisson-Verteilung	$e^{-\vartheta} \frac{\vartheta^k}{k!}, k \in \mathbb{Z}^+$	$(0, \infty)$

stetiger Fall:

P^X	Name	Dichte $f(x)$ bzgl. m^1	Θ
$\mathcal{U}[0, \vartheta]$	Gleichverteilung	$\frac{1}{\vartheta}, 0 \leq x \leq \vartheta$	$(0, \infty)$
$\mathcal{E}(\vartheta)$	Exponentialverteilung	$\vartheta e^{-\vartheta x}, x \geq 0$	$(0, \infty)$
$\Gamma(\alpha, \vartheta)$	Gamma-Verteilung	$\vartheta^\alpha \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-\vartheta x}, x > 0; \alpha > 0$	$(0, \infty)$
$\mathcal{N}(\vartheta, \sigma^2)$	Normalverteilung	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\vartheta)^2}{2\sigma^2}}, x \in \mathbb{R}; \sigma > 0$	\mathbb{R}

Einige der aufgeführten Verteilungsklassen besitzen in natürlicher Weise sogar *zweidimensionale* Parametermengen Θ , so die

- Binomialverteilung $B(n, p)$ mit $\vartheta = (n, p) \in \Theta = \mathbb{N} \times (0, 1]$
- negative Binomialverteilung $NB(\beta, p)$ mit $\vartheta = (\beta, p) \in \Theta = (0, \infty) \times (0, 1]$
- Gamma-Verteilung $\Gamma(\alpha, \lambda)$ mit $\vartheta = (\alpha, \lambda) \in \Theta = (0, \infty) \times (0, \infty)$
- Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mit $\vartheta = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$.

Jede beliebige [feste] Verteilung über der Borel'schen σ -Algebra \mathcal{B}^1 bzw. eine Zufallsvariable Z mit dieser Verteilung induziert übrigens in natürlicher Weise ein statistisches Modell mit einer zweidimensionalen Parametermenge Θ , nämlich die zugehörige *Lage-Skalen-Familie* $\mathcal{V}(\mu, \sigma)$ mit $\mu \in \mathbb{R}, \sigma > 0$, die ausführlicher in Abschnitt III.4 besprochen wird. Sie entsteht anschaulich durch Betrachtung der Verteilungen, die zu den Zufallsvariablen

$$X = \mu + \sigma Z \text{ mit } \vartheta = (\mu, \sigma) \in \Theta = \mathbb{R} \times (0, \infty)$$

gehören. μ heißt hier *Lageparameter*, σ *Skalenparameter*. Die Familie der Normalverteilungen $\mathcal{N}(\mu, \sigma^2)$ ist offenbar ein Spezialfall hiervon, wenn $P^Z = N(0, 1)$ gewählt wird.

III.1. Elementare Schätzverfahren

Ausgehend vom Gesetz der Großen Zahlen besprechen wir zuerst ein Schätzverfahren, das als *Momentenmethode* bekannt ist. Es beruht auf folgender Idee:

Ist $(\mathcal{X}, \mathcal{B}, \{Q_\vartheta\}_{\vartheta \in \Theta}) = (\mathbb{R}^1, \mathcal{B}^1, \{Q_\vartheta\}_{\vartheta \in \Theta})$ ein statistisches Modell für eine Zufallsvariable X auf einem geeigneten (nicht notwendig bekannten) Wahrscheinlichkeitsraum mit $P_\vartheta^X = Q_\vartheta$, $\vartheta \in \Theta \subseteq \mathbb{R}^1$, und existiert eine (von $\vartheta \in \Theta$ unabhängige) messbare Abbildung $g: (\mathbb{R}^1, \mathcal{B}^1) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$ mit

$$E_\vartheta(g(X)) = \int g(X) dP_\vartheta = \int g dQ_\vartheta = \vartheta \text{ für alle } \vartheta \in \Theta,$$

so ist

$$\hat{\vartheta}_n := \frac{1}{n} \sum_{k=1}^n g(X_k)$$

ein "plausibler" Schätzer für $\vartheta \in \Theta$, wenn X_1, \dots, X_n stochastisch unabhängige, wie X verteilte Zufallsvariablen mit $n \in \mathbb{N}$ sind (also dem Produkt-Modell $(\mathbb{R}^n, \mathcal{B}^n, \{\bigotimes_{i=1}^n Q_\vartheta\}_{\vartheta \in \Theta})$ zuzuordnen sind).

Dieser Schätzer heißt *Momentenschätzer*.

Die "Plausibilität" kommt dabei daher, dass nach dem (schwachen wie starken) Gesetz der Großen Zahlen gilt:

$$\hat{\vartheta}_n = \frac{1}{n} \sum_{k=1}^n g(X_k) \xrightarrow{P \text{ f.s.}} E_\vartheta(g(X)) = \vartheta \text{ für } n \rightarrow \infty,$$

wenn sogleich eine unabhängige Folge $\{X_n\}_{n \in \mathbb{N}}$ stochastisch unabhängiger, wie X verteilter Zufallsvariablen betrachtet wird (dies korrespondiert mit dem *unendlichen* statistischen Produkt-Modell $(\prod_{i=1}^{\infty} \mathbb{R}^1, \prod_{i=1}^{\infty} \mathcal{B}^1, \{\bigotimes_{i=1}^{\infty} Q_\vartheta\}_{\vartheta \in \Theta})$).

Eine andere Variante des Momentenschätzers entsteht, wenn eine stetige, streng monotone (von $\vartheta \in \Theta$ unabhängige) Abbildung $h: (\mathbb{R}^1, \mathcal{B}^1) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$ existiert mit

$$E_\vartheta(X) = \int X dP_\vartheta = \int id dQ_\vartheta = h(\vartheta) \text{ für alle } \vartheta \in \Theta;$$

in diesem Falle ist

$$\hat{\vartheta}_n := h^{-1} \left(\frac{1}{n} \sum_{k=1}^n X_k \right)$$

ebenfalls ein "plausibler" Schätzer für $\vartheta \in \Theta$ wegen

$$\hat{\vartheta}_n = h^{-1} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) \xrightarrow{P \text{ f.s.}} h^{-1}(E_\vartheta(X)) = h^{-1}(h(\vartheta)) = \vartheta \text{ für } n \rightarrow \infty.$$

III. Grundprinzipien der Statistik

Die Momentenmethode lässt sich analog auch für den Fall eines mehrdimensionalen Parameters $\vartheta = (\vartheta_1, \dots, \vartheta_m) \in \Theta \subseteq \mathbb{R}^m$ mit $m \in \mathbb{N}$ anwenden, indem man das beschriebene Verfahren auf jede einzelne Komponente ϑ_k mit $k \in \{1, \dots, m\}$ anwendet.

Die nachfolgende Tabelle listet einige Möglichkeiten für Momentenschätzer für $\vartheta \in \Theta \subseteq \mathbb{R}^1$ auf.

P^x	Dichte f bezgl. μ (dominierendes Maß)	Θ	$g(x)$	$h(x)$	$h^{-1}(y)$
$B(m, \vartheta)$	$\binom{m}{k} \vartheta^k (1-\vartheta)^{m-k}$, $k = 0, \dots, m$ bekannt	$[0, 1]$	$\frac{x}{m}$	$m \cdot x$	$\frac{y}{m}$
$NB(\beta, \vartheta)$	$\binom{\beta+k-1}{k} \vartheta^k (1-\vartheta)^{\beta-k}$, $k \in \mathbb{Z}^+$; $\beta > 0$ bekannt	$(0, 1]$	s.u.	$\beta \frac{1-x}{x}$	$\frac{\beta}{\beta+y}$
$\mathcal{P}(\vartheta)$	$e^{-\vartheta} \frac{\vartheta^k}{k!}$, $k \in \mathbb{Z}^+$	$(0, \infty)$	x	x	y
$\mathcal{U}[0, \vartheta]$	$\frac{1}{\vartheta}$, $0 \leq x \leq \vartheta$	$(0, \infty)$	$2x$	$\frac{x}{2}$	$2y$
$\mathcal{E}(\vartheta)$	$\vartheta e^{-\vartheta x}$, $x \geq 0$	$(0, \infty)$	s.u.	$\frac{1}{x}$	$\frac{1}{y}$
$\Gamma(\alpha, \vartheta)$	$\vartheta^\alpha \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-\vartheta x}$, $x > 0$; $\alpha > 0$ bekannt	$(0, \infty)$	s.u.	$\frac{\alpha}{x}$	$\frac{\alpha}{y}$
$\mathcal{N}(\vartheta, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\vartheta)^2}{2\sigma^2}}$, $x \in \mathbb{R}$; $\sigma > 0$ bekannt	\mathbb{R}	x	x	y

Für die negativen Binomialverteilungen $NB(\beta, \vartheta)$ gibt es für beliebige, insbesondere nicht-ganzzahlige $\beta > 0$ keine einfachen Momentenschätzer der ersten Art für $\vartheta \in \Theta = (0, 1]$. Für $\beta = 1$ (geometrische Verteilung) gibt es aber im Prinzip genau einen derartigen Schätzer, den wir jetzt herleiten wollen. Wenn wir annehmen, dass eine Abbildung g der geforderten Art mit

$$E_{\vartheta}(g(X)) = \int g(X) dP_{\vartheta} = \int g dQ_{\vartheta} = \sum_{k=0}^{\infty} g(k) \cdot \vartheta (1-\vartheta)^k = \vartheta \text{ für alle } \vartheta \in (0, 1]$$

existiert, ergibt sich die äquivalente Gleichung

$$\sum_{k=0}^{\infty} g(k) \cdot (1-\vartheta)^k = 1 \text{ für alle } \vartheta \in (0, 1]. \quad (*)$$

Insbesondere ist die Reihe – sogar absolut (Lebesgue-Integral!) – für $\vartheta = \frac{1}{2}$ konvergent. Sei nun $0 < \varepsilon < \frac{1}{2}$. Dann gilt nach (*) mit $0 < \vartheta = 1 - \varepsilon < 1$

$$|1 - g(0)| = \left| \sum_{k=1}^{\infty} g(k) \cdot (1-\vartheta)^k \right| \leq \sum_{k=1}^{\infty} |g(k)| \cdot \varepsilon^k = \sum_{k=1}^{\infty} \frac{|g(k)|}{2^k} \cdot (2\varepsilon)^k \leq 2\varepsilon \sum_{k=1}^{\infty} \frac{|g(k)|}{2^k}$$

III. Grundprinzipien der Statistik

und damit, da $\varepsilon > 0$ beliebig (klein) war, $g(0) = 1$. Bedingung (*) reduziert sich damit, nach Kürzen durch $(1-\vartheta)$ für $\vartheta \in (0,1)$, auf

$$g(1) + \sum_{k=2}^{\infty} g(k) \cdot (1-\vartheta)^{k-1} = 0 \quad \text{für alle } \vartheta \in (0,1). \quad (**)$$

Ein ähnliches Argument wie zuvor zeigt dann, dass $g(1) = 0$ sein muss. Diesen Beweisgang kann man induktiv fortsetzen und erhält dann insgesamt die (auf \mathbb{Z}^+ eindeutige) Lösung

$$g(k) = \begin{cases} 1, & k = 0 \\ 0, & k \in \mathbb{N}. \end{cases}$$

Der Momentenschätzer ist damit aber wegen der Diskretheit der Verteilungen in diesem Modell fast sicher eindeutig bestimmt, mit der Darstellung

$$\hat{\vartheta}_n = \frac{1}{n} \sum_{k=1}^n g(X_k) = \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{\{0\}}(X_k), \quad n \in \mathbb{N},$$

d.h. dieser Schätzer zählt die relative Häufigkeit der Nullen in der Beobachtungsserie. Zum Vergleich: der alternative Momentenschätzer der zweiten Art ist hier gegeben durch

$$\hat{\vartheta}_n := h^{-1} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{1 + \frac{1}{n} \sum_{k=1}^n X_k} = \frac{n}{n + \sum_{k=1}^n X_k}, \quad n \in \mathbb{N}.$$

Wir zeigen hier noch exemplarisch die Existenz eines Momentenschätzers für die negative Binomialverteilung mit $\beta = 2$; für größere ganzzahlige Werte von β kann man analog vorgehen. Aus dem Ansatz

$$E_{\vartheta}(g(X)) = \int g(X) dP_{\vartheta} = \int g dQ_{\vartheta} = \sum_{k=0}^{\infty} g(k) \cdot (k+1) \vartheta^2 (1-\vartheta)^k = \vartheta \quad \text{für alle } \vartheta \in (0,1]$$

und einem Vergleich mit der geometrischen Reihe $\sum_{k=0}^{\infty} \vartheta (1-\vartheta)^k = 1$ erkennt man schnell, dass eine

Lösung durch $g(k) = \frac{1}{k+1}$ für alle $k \in \mathbb{Z}^+$ gegeben ist. Der zugehörige Momentenschätzer lautet also

$$\hat{\vartheta}_n = \frac{1}{n} \sum_{k=1}^n g(X_k) = \frac{1}{n} \sum_{k=1}^n \frac{1}{X_k + 1}, \quad n \in \mathbb{N}.$$

Der Alternativschätzer ist hier gegeben durch

$$\hat{\vartheta}_n := h^{-1} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{2}{2 + \frac{1}{n} \sum_{k=1}^n X_k} = \frac{2n}{2n + \sum_{k=1}^n X_k}, \quad n \in \mathbb{N}.$$

III. Grundprinzipien der Statistik

Für die Exponentialverteilungen $\mathcal{E}(\vartheta) = \Gamma(1, \vartheta)$ mit $\vartheta > 0$ ist hier – im Gegensatz zur geometrischen Verteilung – interessant, dass es einen Momentenschätzer der ersten Art nicht geben kann. Wenn wir nämlich annehmen, dass eine Abbildung g der geforderten Art mit

$$E_{\vartheta}(g(X)) = \int g(X) dP_{\vartheta} = \int g dQ_{\vartheta} = \vartheta \int_0^{\infty} g(x) e^{-\vartheta x} dx = \vartheta \text{ für alle } \vartheta \in \Theta = (0, \infty)$$

existiert, so wäre also

$$\int_0^{\infty} g(x) e^{-\vartheta x} dx = 1 \text{ für alle } \vartheta > 0. \quad (***)$$

Da das Integral $\int g dQ_{\vartheta}$ für alle $\vartheta > 0$ im Sinne von Lebesgue existieren muss, ist also insbesondere die Abbildung $|g(x)|e^{-x}$ (Fall $\vartheta = 1$) über \mathbb{R}^+ integrierbar bezüglich des Lebesgue-Maßes m^1 . Sei nun μ das Maß auf \mathcal{B}^1 mit der m^1 -Dichte $|g(x)|e^{-x} \cdot \mathbb{1}_{[0, \infty)}(x)$, $x \in \mathbb{R}$. Dann impliziert (***) die Ungleichung (wähle $\vartheta = n + 1$)

$$\int_0^{\infty} e^{-nx} \mu(dx) = \int_0^{\infty} |g(x)| e^{-(n+1)x} dx \geq \left| \int_0^{\infty} g(x) e^{-(n+1)x} dx \right| = 1 \text{ für alle } n \in \mathbb{N}. \quad (***)$$

Die Funktionenfolge $\{f_n\}_{n \in \mathbb{N}}$ mit $f_n(x) = e^{-nx} \cdot \mathbb{1}_{[0, \infty)}(x)$ für $n \in \mathbb{N}$ und $x \in \mathbb{R}$ konvergiert nun aber für $n \rightarrow \infty$ punktweise gegen $\mathbb{1}_{\{0\}}$ und ist betragsmäßig durch die μ -integrierbare Majorante 1 beschränkt. Nach dem Satz 18 von der majorisierten Konvergenz folgt also

$$\lim_{n \rightarrow \infty} \int_0^{\infty} e^{-nx} \mu(dx) = \lim_{n \rightarrow \infty} \int_0^{\infty} f_n d\mu = \int_0^{\infty} \mathbb{1}_{\{0\}} d\mu = \mu(\{0\}) = 0$$

im Widerspruch zu (***) .

Für die Gammaverteilungen $\Gamma(\alpha, \vartheta)$ mit $\vartheta > 0$ ist dies jedoch möglich mit $g(x) = \frac{\alpha-1}{x}$, $x > 0$, wenn $\alpha > 1$ ist, denn es gilt dann

$$E_{\vartheta}(g(X)) = \int g dQ_{\vartheta} = (\alpha-1) \int_0^{\infty} \vartheta^{\alpha} \frac{x^{\alpha-2}}{\Gamma(\alpha)} e^{-\vartheta x} dx = \vartheta \int_0^{\infty} \underbrace{\vartheta^{\alpha-1} \frac{x^{\alpha-2}}{\Gamma(\alpha-1)} e^{-\vartheta x}}_{\text{Dichte der } \Gamma(\alpha-1, \vartheta)\text{-Verteilung}} dx = \vartheta \text{ für alle } \vartheta > 0,$$

wie gefordert. (Man beachte, dass g zwar von α , nicht aber von dem hier betrachteten Parameter ϑ abhängt.) Der Momentenschätzer für $\vartheta > 0$ ist hier also gegeben durch

$$\hat{\vartheta}_n = \frac{1}{n} \sum_{k=1}^n g(X_k) = \frac{\alpha-1}{n} \sum_{k=1}^n \frac{1}{X_k}, \quad n \in \mathbb{N}.$$

Für die zweite Variante ergibt sich übrigens der um den Faktor $\frac{\alpha}{\alpha-1}$ größere Schätzer

$$\hat{\vartheta}_n := h^{-1} \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{\alpha}{n} \sum_{k=1}^n \frac{1}{X_k}, \quad n \in \mathbb{N}.$$

III. Grundprinzipien der Statistik

Die beiden Ansätze für Momentenschätzer lassen sich natürlich auch geeignet kombinieren. Wir zeigen dies an folgendem

Beispiel: Wir betrachten die Familie $\{Q_\vartheta\}_{\vartheta \in \Theta} = \{\mathcal{P}(\vartheta)\}_{\vartheta \in \Theta}$ der Poisson-Verteilungen mit $\Theta = (0, \infty)$. Dann gilt für $g(x) = x^2$, $x \geq 0$:

$$E_\vartheta(g(X)) = E_\vartheta(X^2) = \text{Var}(X) + \{E(X)\}^2 = \vartheta^2 + \vartheta = h(\vartheta)$$

mit $h^{-1}(y) = -\frac{1}{2} + \frac{1}{2}\sqrt{4y+1}$, $y \geq 0$. Ein sinnvoller kombinierter Momentenschätzer ist also

$$\hat{\vartheta}_n = h^{-1}\left(\frac{1}{n} \sum_{k=1}^n g(X_k)\right) = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\bar{X}_n^2} \xrightarrow{\text{f.s.}} h^{-1}(E_\vartheta(g(X))) = \vartheta.$$

Wenn verschiedene Schätzer für einen Parameter zur Verfügung stehen, benötigt man sinnvollerweise Kriterien, nach denen man einen guten oder sogar „den besten“ Schätzer auswählen kann.

Definition 47 (Schätzer und ihre Gütekriterien). Gegeben sei ein statistisches (Produkt-)Modell

$\left(\prod_{i=1}^n \mathcal{X}, \prod_{i=1}^n \mathcal{B}, \left\{ \prod_{i=1}^n Q_\vartheta \right\}_{\vartheta \in \Theta} \right)$ mit $n \in \mathbb{N}$ und einer Borel-messbaren Parametermenge $\Theta \subseteq \mathbb{R}$. Ein [hier

eindimensionaler] Schätzer $\hat{\vartheta}_n$ (auch Schätzfunktion genannt) ist zunächst eine reellwertige, $\prod_{i=1}^n \mathcal{B}$ -messbare Abbildung auf $\prod_{i=1}^n \mathcal{X}$.

1. Der Schätzer heißt *zulässig*, wenn seine Werte in der Parametermenge Θ liegen.
2. Der Schätzer heißt *erwartungstreu*, wenn er integrierbar ist bezüglich des Maßes Q_ϑ mit

$$E_\vartheta(\hat{\vartheta}_n) = \vartheta \text{ für alle } \vartheta \in \Theta.$$

3. Bezüglich des unendlichen statistischen Produktmodells gilt:

Die Folge $\{\hat{\vartheta}_n\}_{n \in \mathbb{N}}$ der Schätzer [bzw. der Schätzer] heißt (*schwach* bzw. *stark*) *konsistent*, wenn gilt:

$$\lim_{n \rightarrow \infty} \hat{\vartheta}_n = \vartheta \text{ (stochastisch bzw. fast sicher).}$$

4. Der *mittlere quadratische Fehler* des Schätzers (engl.: **Mean Square Error**, MSE) ist definiert durch

$$MSE(\hat{\vartheta}_n) = E_\vartheta \left[(\hat{\vartheta}_n - \vartheta)^2 \right].$$

In der Regel sind nach der entsprechenden Eingangsbemerkung Schätzfunktionen Transformationen von Zufallsvariablen auf $(\Omega, \mathcal{A}, \{P_\vartheta\}_{\vartheta \in \Theta})$ mit Werten in $(\mathcal{X}, \mathcal{B}, \{Q_\vartheta\}_{\vartheta \in \Theta}) = (\mathbb{R}^1, \mathcal{B}^1, \{Q_\vartheta\}_{\vartheta \in \Theta})$, so dass sich diese Begriffe hierauf sinngemäß übertragen.

III. Grundprinzipien der Statistik

Nach dem gewählten Zugang sind Momentenschätzer des ersten Typs, also $\hat{\vartheta}_n = \frac{1}{n} \sum_{k=1}^n g(X_k)$ für $n \in \mathbb{N}$, grundsätzlich erwartungstreu und stark (und damit auch schwach) konsistent, während Momentenschätzer des zweiten Typs, also $\hat{\vartheta}_n := h^{-1} \left(\frac{1}{n} \sum_{k=1}^n X_k \right)$ für $n \in \mathbb{N}$, in der Regel nicht erwartungstreu, aber (wegen der Stetigkeit von h und damit auch h^{-1}) stark (und damit auch schwach) konsistent sind. Der MSE ist daher für Momentenschätzer des zweiten Typs tendenziell höher als der für Momentenschätzer des ersten Typs, denn es gilt nach Lemma 37 a):

$$MSE(\hat{\vartheta}_n) = E_{\vartheta} \left[\left(\hat{\vartheta}_n - \vartheta \right)^2 \right] = \text{Var}(\hat{\vartheta}_n) + \left(E(\hat{\vartheta}_n) - \vartheta \right)^2.$$

Der MSE wird also genau dann minimal, wenn $E_{\vartheta}(\hat{\vartheta}_n) = \vartheta$ für $\vartheta \in \Theta$ gilt, der Schätzer also erwartungstreu ist.

Für das obige Poisson-Beispiel folgt etwa mit der Jensen'schen Ungleichung (Lemma 36), angewendet auf die (strikt konkave) Wurzelfunktion:

$$\begin{aligned} E(\hat{\vartheta}_n - \vartheta) &= \frac{1}{2} E \left(\sqrt{1 + 4X_n^2} - (1 + 2\vartheta) \right) < \frac{1}{2} \left(\sqrt{1 + 4E(X_n^2)} - (1 + 2\vartheta) \right) \\ &= \frac{1}{2} \left(\sqrt{1 + 4\vartheta + 4\vartheta^2} - (1 + 2\vartheta) \right) = \frac{1}{2} \left(\sqrt{(1 + 2\vartheta)^2} - (1 + 2\vartheta) \right) = 0, \end{aligned}$$

also $E(\hat{\vartheta}_n) < \vartheta$. Dieser Momentenschätzer unterschätzt den wahren Parameter also systematisch. Die hieraus resultierende Abweichung wird auch als *Bias* bezeichnet. Leider kann der Bias für dieses Beispiel nicht explizit berechnet werden, ebenso wenig wie der MSE.

Das folgende Beispiel verdeutlicht den Effekt von Schätzern unterschiedlicher Güte etwas besser.

Beispiel: Wir betrachten die Familie $\{\mathcal{Q}_{\vartheta}\}_{\vartheta \in \Theta} = \{\mathcal{U}[0, \vartheta]\}_{\vartheta \in \Theta}$ der stetigen Gleichverteilungen über dem Intervall $[0, \vartheta]$ mit $\vartheta \in \Theta = (0, \infty)$. Als Momentenschätzer verwenden wir

$$\hat{\vartheta}_n = \frac{2}{n} \sum_{k=1}^n X_k \quad \text{mit} \quad E(\hat{\vartheta}_n) = \frac{2}{n} \sum_{k=1}^n E(X_k) = 2 \cdot \frac{\vartheta}{2} = \vartheta \quad \text{für alle } n \in \mathbb{N} \text{ und } \vartheta \in \Theta.$$

Dieser Schätzer ist also erwartungstreu, stark konsistent und besitzt den MSE

$$MSE(\hat{\vartheta}_n) = \text{Var}(\hat{\vartheta}_n) = 4 \cdot \frac{\vartheta^2}{12n} = \frac{\vartheta^2}{3n} \quad \text{für alle } n \in \mathbb{N} \text{ und } \vartheta \in \Theta.$$

Als Alternative, die *kein* Momentenschätzer ist, betrachten wir den nahe liegenden Schätzer

$$\hat{\vartheta}_n := \frac{n+1}{n} \max\{X_1, \dots, X_n\} \quad \text{für } n \in \mathbb{N}.$$

III. Grundprinzipien der Statistik

Der Vorfaktor wurde dabei so gewählt, dass $\hat{\vartheta}_n$ für alle $n \in \mathbb{N}$ erwartungstreu ist. Es gilt nämlich:

$$F_n(x) := P(\max\{X_1, \dots, X_n\} \leq x) = P\left(\bigcap_{k=1}^n \{X_k \leq x\}\right) = \prod_{k=1}^n P(X_k \leq x) = \left(\frac{x}{\vartheta}\right)^n$$

für alle $x \in [0, \vartheta]$ und $n \in \mathbb{N}$, so dass nach Satz 32 d) folgt:

$$E(\max\{X_1, \dots, X_n\}) = \int_0^{\vartheta} F_n^{-1}(u) du = \vartheta \int_0^1 \sqrt[n]{u} du = \vartheta \frac{u^{\frac{1}{n}+1}}{\frac{1}{n}+1} \Big|_0^1 = \frac{n}{n+1} \vartheta.$$

Ferner ist die Folge $\{\max\{X_1, \dots, X_n\}\}_{n \in \mathbb{N}}$ nicht-negativ, schwach monoton wachsend und nach oben fast sicher durch ϑ beschränkt, also auch fast sicher konvergent gegen eine nicht-negative Zufallsvariable Y . Die Folge $\{\max\{X_1, \dots, X_n\}\}_{n \in \mathbb{N}}$ ist aber auch schwach konvergent gegen ϑ wegen

$$\lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 1, & x \geq \vartheta \\ 0, & x < \vartheta. \end{cases}$$

Damit ist aber $Y = \vartheta$ fast sicher und somit der Schätzer $\hat{\vartheta}_n$ ebenso wie $\hat{\vartheta}_n$ erwartungstreu und stark konsistent. Der Unterschied besteht jedoch in den unterschiedlichen Größenordnungen des MSE (also hier: der Varianz), denn es gilt nach obigem

$$G_n(x) := P(\max\{X_1, \dots, X_n\}^2 \leq x) = P(\max\{X_1, \dots, X_n\} \leq \sqrt{x}) = \left(\frac{\sqrt{x}}{\vartheta}\right)^n$$

für alle $x \in [0, \vartheta]$ und $n \in \mathbb{N}$, so dass wiederum nach Satz 32 d) folgt:

$$E(\max\{X_1, \dots, X_n\}^2) = \int_0^{\vartheta^2} G_n^{-1}(u) du = \vartheta^2 \int_0^1 \sqrt[n]{u^2} du = \vartheta^2 \frac{u^{\frac{2}{n}+1}}{\frac{2}{n}+1} \Big|_0^1 = \frac{n}{n+2} \vartheta^2.$$

Damit ergibt sich

$$MSE(\hat{\vartheta}_n) = \text{Var}(\hat{\vartheta}_n) = E(\hat{\vartheta}_n^2) - \left[E(\hat{\vartheta}_n)\right]^2 = \vartheta^2 \left[\left(\frac{n+1}{n}\right)^2 \frac{n}{n+2} - 1 \right] = \frac{1}{n(n+2)} \vartheta^2 \leq \frac{\vartheta^2}{3n} = MSE(\hat{\vartheta}_n)$$

für alle $n \in \mathbb{N}$ und $\vartheta \in \Theta$. Für $n > 1$ besitzt $\hat{\vartheta}_n$ sogar einen strikt kleineren MSE (Varianz) als $\hat{\vartheta}_n$ und wäre daher für eine konkrete Anwendung die bessere Wahl. [Für $n = 1$ sind beide Schätzer gleich].

Dieses Beispiel zeigt, dass es durchaus „gute“ Schätzer unterschiedlicher Qualität geben kann. Der zuletzt verwendete alternative Schätzer $\hat{\vartheta}_n$ ist dabei eine Variante eines so genannten *Maximum-Likelihood-Schätzers*, dessen Prinzip jetzt vorgestellt wird.

III. Grundprinzipien der Statistik

Wir gehen wieder von einem statistisches Modell $(\mathfrak{X}, \mathcal{B}, \{Q_\vartheta\}_{\vartheta \in \Theta})$ bzw. dem daraus abgeleiteten Produkt-Modell $\left(\prod_{i=1}^n \mathfrak{X}, \prod_{i=1}^n \mathcal{B}, \left\{ \prod_{i=1}^n Q_\vartheta \right\}_{\vartheta \in \Theta} \right)$ aus, bei dem die Verteilungen Q_ϑ , $\vartheta \in \Theta$ eine Dichte f_ϑ bezüglich eines dominierenden Maes μ besitzen. Im englischen Sprachgebrauch wird eine solche Dichte auch als *Likelihood* bezeichnet. Im Fall des abzählenden Maes μ ist die Likelihood mit dem Begriff der Wahrscheinlichkeit identisch; im Fall des Lebesgue-Maes μ ist die Dichte aber *keine* Wahrscheinlichkeit.

Definition 48. Im statistischen Produkt-Modell heißt die durch

$$L_n(x_1, \dots, x_n; \vartheta) = \prod_{k=1}^n f_\vartheta(x_k), \quad (x_1, \dots, x_n) \in \prod_{k=1}^n \mathfrak{X}, \quad \vartheta \in \Theta$$

gegebene Produktdichte *Likelihood-Funktion* zum Parameter $\vartheta \in \Theta$. Ein von den zugehörigen Zufallsvariablen X_1, \dots, X_n (messbar) abhängender Schätzer $\hat{\vartheta}_n = \hat{\vartheta}_n(X_1, \dots, X_n)$ mit der Eigenschaft

$$L_n(x_1, \dots, x_n; \hat{\vartheta}_n(x_1, \dots, x_n)) = \sup_{\vartheta \in \Theta} L_n(x_1, \dots, x_n; \vartheta)$$

heißt *Maximum-Likelihood-Schätzer* (ML-Schätzer) für den Parameter $\vartheta \in \Theta$.

Im Falle des abzählenden Maes bedeutet dies, dass der dem Experiment zu Grunde liegende Parameter $\vartheta \in \Theta$ so geschätzt wird, dass unter diesem die gezogenen Beobachtungen „am wahrscheinlichsten“ sind, d.h. es gilt

$$P_{\hat{\vartheta}_n}(X_1 = x_1, \dots, X_n = x_n) = \sup_{\vartheta \in \Theta} P_\vartheta(X_1 = x_1, \dots, X_n = x_n) \text{ für alle } (x_1, \dots, x_n) \in \prod_{k=1}^n \mathfrak{X}.$$

Bemerkung: Häufig lässt sich das Supremum der Likelihood-Funktion mit Methoden der Analysis, etwa durch Differenzieren, ermitteln. Die Rechnung wird dabei meist einfacher, wenn man statt der Funktion L_n die so genannte *Log-Likelihood-Funktion* $\ell_n = \ln L_n$ betrachtet; dabei ist nur zu beachten, dass die Menge

$$\left\{ (x_1, \dots, x_n) \in \prod_{k=1}^n \mathfrak{X} \mid L_n(x_1, \dots, x_n; \vartheta) = 0 \right\}$$

für jedes $\vartheta \in \Theta$ eine $\prod_{i=1}^n Q_\vartheta$ -Nullmenge und damit $\ell_n = \ln L_n$ $\prod_{i=1}^n \mu$ -fast überall reellwertig ist. Ein typisches Vorgehen für die Berechnung eines ML-Schätzers (für einen eindimensionalen Parameter) ist also – unter gewissen Regularitätsbedingungen – die Lösung der Gleichung (notwendige Bedingung)

$$\frac{\partial}{\partial \vartheta} \ell_n(x_1, \dots, x_n; \vartheta) = 0$$

mit anschließender Überprüfung der hinreichenden Bedingung für die Lösung $\hat{\vartheta}_n = \hat{\vartheta}_n(x_1, \dots, x_n)$:

$$\frac{\partial^2}{\partial \vartheta^2} \ell_n(x_1, \dots, x_n; \hat{\vartheta}_n) < 0.$$

III. Grundprinzipien der Statistik

Die nachfolgende Tabelle listet die (eindeutigen) ML-Schätzer für $\vartheta \in \Theta \subseteq \mathbb{R}^1$ für die schon am Anfang dieses Abschnitts betrachteten Verteilungsfamilien auf. Man vergleiche die Ergebnisse!

P^X	Dichte f bezgl. μ (dominierendes Maß)	Θ	$\hat{\vartheta}_n$
$B(m, \vartheta)$	$\binom{m}{k} \vartheta^k (1-\vartheta)^{m-k}$, $k = 0, \dots, m$ bekannt	$[0, 1]$	$\frac{1}{nm} \sum_{i=1}^n X_i$
$NB(\beta, \vartheta)$	$\binom{\beta+k-1}{k} \vartheta^k (1-\vartheta)^{\beta-k}$, $k \in \mathbb{Z}^+$; $\beta > 0$ bekannt	$(0, 1]$	$\frac{n\beta}{n\beta + \sum_{i=1}^n X_i}$
$\mathcal{P}(\vartheta)$	$e^{-\vartheta} \frac{\vartheta^k}{k!}$, $k \in \mathbb{Z}^+$	$(0, \infty)$	$\frac{1}{n} \sum_{i=1}^n X_i$
$\mathcal{U}[0, \vartheta]$	$\frac{1}{\vartheta}$, $0 \leq x \leq \vartheta$	$(0, \infty)$	$\max(X_1, \dots, X_n)$
$\mathcal{E}(\vartheta)$	$\vartheta e^{-\vartheta x}$, $x \geq 0$	$(0, \infty)$	$\frac{n}{\sum_{i=1}^n X_i}$
$\Gamma(\alpha, \vartheta)$	$\vartheta^\alpha \frac{x^{\alpha-1}}{\Gamma(\alpha)} e^{-\vartheta x}$, $x > 0$; $\alpha > 0$ bekannt	$(0, \infty)$	$\frac{n\alpha}{\sum_{i=1}^n X_i}$
$\mathcal{N}(\vartheta, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\vartheta)^2}{2\sigma^2}}$, $x \in \mathbb{R}$; $\sigma > 0$ bekannt	\mathbb{R}	$\frac{1}{n} \sum_{i=1}^n X_i$

Wir zeigen die zugehörige Rechnung beispielhaft für folgende Verteilungsfamilien:

$NB(\beta, \vartheta)$: hier ist $\ell_n(x_1, \dots, x_n; \vartheta) = \ln \left\{ \prod_{k=1}^n \binom{\beta + x_k - 1}{x_k} \right\} + n\beta \ln \vartheta + \ln(1-\vartheta) \sum_{k=1}^n x_k$ mit

$$\frac{\partial}{\partial \vartheta} \ell_n(x_1, \dots, x_n; \vartheta) = \frac{n\beta}{\vartheta} - \frac{\sum_{k=1}^n x_k}{1-\vartheta} = 0 \Leftrightarrow \vartheta = \frac{n\beta}{n\beta + \sum_{i=1}^n x_i} = \hat{\vartheta}_n \quad \text{und}$$

$$\frac{\partial^2}{\partial \vartheta^2} \ell_n(x_1, \dots, x_n; \hat{\vartheta}_n) = -\frac{n\beta}{\hat{\vartheta}_n^2} - \frac{\sum_{k=1}^n x_k}{(1-\hat{\vartheta}_n)^2} < 0.$$

III. Grundprinzipien der Statistik

$\Gamma(\alpha, \vartheta)$: hier ist $\ell_n(x_1, \dots, x_n; \vartheta) = \ln \left\{ \frac{\prod_{k=1}^n x_k^{\alpha-1}}{\Gamma(\alpha)^n} \right\} + n\alpha \ln \vartheta - \vartheta \sum_{k=1}^n x_k$ mit

$$\frac{\partial}{\partial \vartheta} \ell_n(x_1, \dots, x_n; \vartheta) = \frac{n\alpha}{\vartheta} - \sum_{k=1}^n x_k = 0 \quad \Leftrightarrow \quad \vartheta = \frac{n\alpha}{\sum_{i=1}^n x_i} = \hat{\vartheta}_n \quad \text{und}$$

$$\frac{\partial^2}{\partial \vartheta^2} \ell_n(x_1, \dots, x_n; \hat{\vartheta}_n) = -\frac{n\alpha}{\hat{\vartheta}_n^2} < 0.$$

$\mathcal{U}[0, \vartheta]$: hier ist $L_n(x_1, \dots, x_n; \vartheta) = \begin{cases} \frac{1}{\vartheta^n} & \text{für } 0 \leq x_1, \dots, x_n \leq \vartheta \Leftrightarrow \vartheta \geq \max\{x_1, \dots, x_n\} \\ 0 & \text{sonst.} \end{cases}$

Das Supremum der Likelihood-Funktion lässt sich hier also *nicht* durch Differenzieren bestimmen, aber aus Monotoniegründen ist dieses offenbar gegeben durch $\hat{\vartheta}_n = \max\{X_1, \dots, X_n\}$.

ML-Schätzer besitzen in der Regel viele wichtige statistische Güteeigenschaften (z.B. Konsistenz), die wir hier aber nicht genauer besprechen wollen; siehe etwa PRUSCHA (2000), Kapitel V. Die Erwartungstreue geht dabei im Allgemeinen aber verloren, wie man am Beispiel der stetigen Gleichverteilung $\mathcal{U}[0, \vartheta]$ sieht: hier gilt ja nach der obigen Rechnung

$$E(\max\{X_1, \dots, X_n\}) = \frac{n}{n+1} \vartheta < \vartheta.$$

Ähnlich wie bei den Momentenschätzern lassen sich ML-Schätzer auch für mehrdimensionale Parameter ϑ bestimmen. Wir zeigen dies exemplarisch am Beispiel der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ mit $\vartheta = (\vartheta_1, \vartheta_2) = (\mu, \sigma^2)$: hier ist

$$\ell_n(x_1, \dots, x_n; \vartheta_1, \vartheta_2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\vartheta_2) - \frac{1}{2\vartheta_2} \sum_{k=1}^n (x_k - \vartheta_1)^2$$

mit

$$\frac{\partial}{\partial \vartheta_1} \ell_n(x_1, \dots, x_n; \vartheta_1, \vartheta_2) = \frac{1}{\vartheta_2} \sum_{k=1}^n (x_k - \vartheta_1)$$

$$\frac{\partial}{\partial \vartheta_2} \ell_n(x_1, \dots, x_n; \vartheta_1, \vartheta_2) = -\frac{n}{2\vartheta_2} + \frac{1}{2\vartheta_2^2} \sum_{k=1}^n (x_k - \vartheta_1)^2.$$

Nullsetzen der partiellen Ableitungen (notwendige Bedingung) liefert die eindeutige und messbare Lösung

$$\hat{\vartheta}_{1n} = \frac{1}{n} \sum_{k=1}^n x_k =: \bar{x}_n, \quad \hat{\vartheta}_{2n} = \frac{1}{n} \sum_{k=1}^n (x_k - \hat{\vartheta}_{1n})^2 = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x}_n)^2 =: s_n^2.$$

III. Grundprinzipien der Statistik

Weiter gilt:

$$\frac{\partial^2}{\partial \vartheta_1^2} \ell_n(x_1, \dots, x_n; \vartheta_1, \vartheta_2) = -\frac{n}{\vartheta_2}$$

$$\frac{\partial^2}{\partial \vartheta_1 \partial \vartheta_2} \ell_n(x_1, \dots, x_n; \vartheta_1, \vartheta_2) = -\frac{1}{\vartheta_2^2} \sum_{k=1}^n (x_k - \vartheta_1)$$

$$\frac{\partial^2}{\partial \vartheta_2^2} \ell_n(x_1, \dots, x_n; \vartheta_1, \vartheta_2) = \frac{n}{2\vartheta_2^2} - \frac{1}{\vartheta_2^3} \sum_{k=1}^n (x_k - \vartheta_1)^2.$$

Für die Hesse-Matrix $H(\hat{\vartheta}_n)$ an der Stelle $\hat{\vartheta}_n$ erhält man also (hinreichende Bedingung)

$$H(\hat{\vartheta}_n) = \begin{bmatrix} -\frac{n}{s_n^2} & 0 \\ 0 & -\frac{n}{2s_n^4} \end{bmatrix} \quad \text{mit } \det H(\hat{\vartheta}_n) = \frac{n^2}{2s_n^6} > 0 \quad \text{und} \quad -\frac{n}{s_n^2} < 0,$$

woraus folgt, dass $\hat{\vartheta}_n = (\bar{x}_n, s_n^2)$ ein (relatives) Maximum der (Log-)Likelihood-Funktion ist. Wir erhalten also zusammenfassend als ML-Schätzer für die beiden (eindimensionalen) Parameter der Normalverteilung:

$$\boxed{\hat{\mu}_n = \bar{X}_n, \quad \hat{\sigma}_n^2 = s_n^2 = \frac{1}{n} \sum_{k=1}^n (X_k - \bar{X}_n)^2.}$$

Der erste Schätzer ist offenkundig erwartungstreu, nicht aber der zweite, denn es gilt:

$$\begin{aligned} E \left[\sum_{k=1}^n (X_k - \bar{X}_n)^2 \right] &= E \left[\sum_{k=1}^n (X_k - \mu + \mu - \bar{X}_n)^2 \right] \\ &= E \left[\sum_{k=1}^n (X_k - \mu)^2 \right] + 2E \left[(\mu - \bar{X}_n) \sum_{k=1}^n (X_k - \mu) \right] + nE \left[(\mu - \bar{X}_n)^2 \right] \\ &= n\sigma^2 - n\text{Var}(\bar{X}_n - \mu) = n\sigma^2 - \sigma^2 = (n-1)\sigma^2. \end{aligned}$$

Die erwartungstreue Modifikation des zweiten Schätzers wäre demnach

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2.$$

Dieser sowie der nicht-modifizierte Schätzer ist offenbar für beliebige Verteilungen mit existierender Varianz ein erwartungstreuere und nach dem Gesetz der Großen Zahlen auch stark konsistenter Schätzer für die Varianz.

III. Grundprinzipien der Statistik

An dieser Stelle sind einige grundsätzliche, kritische Anmerkungen zur praktischen Anwendung der Statistik angebracht. Wie aus den obigen Ausführungen ersichtlich ist, bezieht sich die statistische Methodik zunächst auf *statistische Modelle*, d.h. sie wird im Rahmen der maßtheoretischen Stochastik auf der Basis von Zufallsvariablen (oder Zufallsvektoren oder Folgen von Zufallsvariablen) entwickelt. In der realen Anwendung wird diese Methodik aber auf „Daten“ oder „Stichproben“ bezogen, deren Herkunft nicht unmittelbar aus mathematischen Modellen abgeleitet werden kann. Selbst der Allem zu Grunde liegende Wahrscheinlichkeitsbegriff ist nicht ohne weiteres in die Realität übertragbar. Die Formalisierung des Wahrscheinlichkeitsbegriffs mittels σ -additiver Mengenfunktionen erscheint zwar plausibel und führt auch zu einer genügend reichhaltigen Theorie, ist aber nicht die einzige Weise, wie man diesen Begriff formalisieren kann. Es gibt z.B. eine durchaus ernst zu nehmende Wahrscheinlichkeitstheorie, die lediglich mit endlich-additiven Mengenfunktionen operiert; so gehört etwa der bekannte Italiener Bruno de Finetti (1906 - 1985) zu einem Vertreter dieser Richtung. (De Finetti hat darüber hinaus - neben anderen - eine subjektive Wahrscheinlichkeitstheorie entwickelt; von ihm stammt u.a. der legendäre Ausspruch „Es existiert keine objektive Wahrscheinlichkeit“, siehe das Vorwort zur deutschen Ausgabe seines Buchs „Wahrscheinlichkeitstheorie“ (DE FINETTI (1981); vgl. auch DE FINETTI (1990) und JAYNES (2006)). Unter Verwendung lediglich endlich-additiver Mengenfunktionen läßt sich z.B. widerspruchsfrei eine diskrete „Gleichverteilung“ über \mathbb{Z} angeben, bei der alle arithmetischen Mengen der Form $A(n, k) = n + k\mathbb{Z}$ mit $n \in \mathbb{Z}, k \in \mathbb{N}$ die „Wahrscheinlichkeit“ $1/k$ besitzen, was unter der Annahme der σ -Additivität des Maßes natürlich nicht möglich ist. Eine Konsequenz dieses Ansatzes ist allerdings, dass alle *endlichen* Teilmengen von \mathbb{Z} die Wahrscheinlichkeit Null besitzen, und dass eine Fortsetzung der so eingeführten „Wahrscheinlichkeit“ auf größere Mengensysteme im Allgemeinen nur mit dem Auswahlaxiom, also nicht-konstruktivistisch möglich ist.

Seit etwa der Mitte des letzten Jahrhunderts gibt es darüber hinaus verstärkt Versuche, den Begriff des „Zufalls“ ohne Maßtheorie - im Rahmen der Komplexitätstheorie - zu formalisieren; vgl. etwa NIES (2009) oder DOWNEY AND HIRSCHFELDT (2010).

Besonders kritisch ist die Anwendung statistischer Methoden auf Daten zu sehen, die nicht aus physikalisch erklärbaren Zusammenhängen (wie z.B. das Würfeln) entstehen. Hierzu gehören etwa Daten aus den Human- und Sozialwissenschaften, aber auch der Ökonomie. Wirtschaftliche Entscheidungen, die beispielsweise den Verlauf von Aktienkursen beeinflussen, kommen oft durch Dinge wie den menschlichen Willen zustande, der schlechterdings nicht mathematisierbar ist. Aber selbst im Fall des Würfeln macht das stochastische Gesetz der großen Zahlen keine Aussagen über eine Abfolge konkreter Versuche mit einem realen Würfel - auch wenn sich dieses „Gesetz“ meist empirisch erfolgreich belegen läßt. Vgl. hierzu auch die Ausführungen in DE FINETTI (1981), Kapitel VII, Abschnitt 5.5.

Schließlich sei noch angemerkt, dass „Stichproben“ aus Verteilungen mit einer stetigen Verteilungsfunktion eigentlich gar nicht „gezogen“ werden können, weil jede „Realisierung“, d.h. jede einpunktige Größe, eine Wahrscheinlichkeit von Null besitzt. Strenggenommen ist daher die Maximum-Likelihood-Methode in solchen Fällen gar nicht anwendbar. (Man kann sie nur dadurch rechtfertigen, dass man unterstellt, dass die gezogenen Daten selbst ungenau und eigentlich Repräsentanten für ein Intervall sind, was im Prinzip zu einer gewissen Diskretisierung der Verteilung führt, für die das Maximum-Likelihood-Verfahren dann gerechtfertigt ist).

Trotz aller dieser Einwendungen werden statistische Methoden aber wie selbstverständlich fast überall in der Praxis verwendet, bis hin zu legislativen, aufsichtsrechtlichen Vorgaben etwa im Versicherungs- und Finanzwesen, aber auch in der Medizin. Dies ist zwar mangels geeigneter praktikabler Alternativen akzeptabel, darf aber nicht zu einem grundsätzlichen Automatismus („Modellgläubigkeit“) führen, sondern sollte immer auch - auf der Basis eines ausreichenden Fachwissens - kritisch hinterfragt werden.

III.2. Elementare Testverfahren

Wir besprechen hier zunächst nur *einfache* Hypothesentests. Dazu betrachten wir ein statistisches Modell $(\mathcal{X}, \mathcal{B}, \{Q_\vartheta\}_{\vartheta \in \Theta})$ für einen Zufallsvektor \mathbf{X} mit $P_\vartheta^{\mathbf{X}} = Q_\vartheta$, $\vartheta \in \Theta := \{\vartheta_0, \vartheta_1\}$ (der Zufallsvektor darf auch eindimensional sein; Q_ϑ kann dabei auch ein Produktmaß sein). Auf Grund einer Stichprobe \mathbf{x} ("Realisation von \mathbf{X} ") soll entschieden werden, ob der Parameter ϑ_0 (Nullhypothese) oder der Parameter ϑ_1 (Alternative) vorliegt:

$$\text{Nullhypothese } H_0: \vartheta = \vartheta_0 \qquad \text{Alternative } H_1: \vartheta = \vartheta_1$$

In dieser Situation gibt es genau vier verschiedene Entscheidungssituationen:

		Es liegt vor	
		Nullhypothese H_0	Alternative H_1
Die Entscheidung fällt für	Nullhypothese H_0	korrekt	Fehler 2. Art
	Alternative H_1	Fehler 1. Art	korrekt

Ziel ist es, bei vorgegebener maximaler Fehlerwahrscheinlichkeit 1. Art von $\alpha \in [0,1]$ die Fehlerwahrscheinlichkeit 2. Art zu minimieren. Die eigentliche Entscheidung geschieht dabei auf der Basis einer so genannten *Testgröße* oder auch *Teststatistik*, das ist eine geeignete messbare Abbildung $T : (\mathcal{X}, \mathcal{B}) \rightarrow (\mathbb{R}^1, \mathcal{B}^1)$, die auf den Zufallsvektor \mathbf{X} bzw. die Realisation \mathbf{x} angewandt wird.

Man kann zeigen, dass dieses Problem durch einen so genannten *Likelihood-Quotienten-Test* gelöst wird, bei dem die Testgröße der Likelihood-Quotient

$$T = \frac{f_1}{f_0}$$

ist, wobei die f_i Dichten der Verteilungen Q_{ϑ_i} , $i = 1,2$ bezüglich eines geeigneten dominierenden Maßes μ sind (vgl. etwa PRUSCHA (2000)). Die Entscheidungsregel bei diesem Test lautet erwartungsgemäß:

$$\begin{aligned} T(\mathbf{X}) \leq c_\alpha &: \text{Entscheidung für } H_0 \\ T(\mathbf{X}) > c_\alpha &: \text{Entscheidung für } H_1. \end{aligned}$$

Dabei ist die Ablehnschranke c_α (in der Regel minimal) so zu bestimmen, dass

$$P_{\vartheta_0}(T(\mathbf{X}) > c_\alpha) \leq \alpha \text{ (Testniveau)}$$

bleibt. Die Fehlerwahrscheinlichkeit 2. Art ist hier gegeben durch

$$\beta = P_{\vartheta_1}(T(\mathbf{X}) \leq c_\alpha).$$

Heuristische Begründung: Ist $T(\mathbf{X})$ "groß", so ist typischerweise $f_1(\mathbf{X}) > f_0(\mathbf{X})$, d.h. die Likelihood an der Stelle \mathbf{X} ist unter der Alternative größer als unter der Nullhypothese. Man beachte, dass

III. Grundprinzipien der Statistik

im diskreten Fall die Likelihood gerade einer Wahrscheinlichkeit entspricht, was die Erläuterung noch plausibler macht.

Im Fall zweier einfacher Hypothesen ist die Bestimmung eines dominierenden Maßes μ leicht: man wähle etwa das Summenmaß $\mu = Q_{\vartheta_0} + Q_{\vartheta_1}$. Nach dem Satz von Radon-Nikodym sind dann beide Maße Q_{ϑ_0} und Q_{ϑ_1} stetig bezüglich μ , so dass Dichten f_0 und f_1 existieren mit

$$Q_{\vartheta_i}(B) = \int_B f_i d\mu, \quad i = 1, 2, \quad B \in \mathcal{B}.$$

Beispiel: Wir wählen Binomialverteilungen $Q_{\vartheta_i} = B(1, \vartheta_i)$ mit $\vartheta_i \in (0, 1)$. Dann sind die Dichten

$$f_0 = \frac{1-\vartheta_0}{2-\vartheta_0-\vartheta_1} \mathbb{1}_{\{0\}} + \frac{\vartheta_0}{\vartheta_0+\vartheta_1} \mathbb{1}_{\{1\}}, \quad f_1 = \frac{1-\vartheta_1}{2-\vartheta_0-\vartheta_1} \mathbb{1}_{\{0\}} + \frac{\vartheta_1}{\vartheta_0+\vartheta_1} \mathbb{1}_{\{1\}}$$
 eine mögliche Wahl wegen

$$Q_{\vartheta_i}(\{1\}) = \vartheta_i = \frac{\vartheta_i}{\vartheta_0+\vartheta_1} \cdot (\vartheta_0+\vartheta_1) = \frac{\vartheta_i}{\vartheta_0+\vartheta_1} \mu(\{1\}) = \int \frac{\vartheta_i}{\vartheta_0+\vartheta_1} \mathbb{1}_{\{1\}} d\mu = \int_{\{1\}} f_i d\mu \quad \text{und}$$

$$Q_{\vartheta_i}(\{0\}) = 1 - \vartheta_i = \frac{1-\vartheta_i}{2-\vartheta_0-\vartheta_1} \cdot (2-\vartheta_0-\vartheta_1) = \frac{1-\vartheta_i}{2-\vartheta_0-\vartheta_1} \mu(\{0\}) = \int \frac{1-\vartheta_i}{2-\vartheta_0-\vartheta_1} \mathbb{1}_{\{0\}} d\mu = \int_{\{0\}} f_i d\mu.$$

Als Teststatistik ergibt sich hier

$$T = \frac{f_1}{f_0} = \frac{1-\vartheta_1}{1-\vartheta_0} \mathbb{1}_{\{0\}} + \frac{\vartheta_1}{\vartheta_0} \mathbb{1}_{\{1\}} \quad \text{oder anders ausgedrückt} \quad T(x) = \begin{cases} \frac{1-\vartheta_1}{1-\vartheta_0}, & \text{falls } x = 0 \\ \frac{\vartheta_1}{\vartheta_0}, & \text{falls } x = 1. \end{cases}$$

Alternativ kann natürlich auch $\mu = \#$ (abzählendes Maß) gewählt werden; dann wären

$$f_0 = (1-\vartheta_0) \mathbb{1}_{\{0\}} + \vartheta_0 \mathbb{1}_{\{1\}}, \quad f_1 = (1-\vartheta_1) \mathbb{1}_{\{0\}} + \vartheta_1 \mathbb{1}_{\{1\}},$$

mit derselben Teststatistik T .

Wir nehmen jetzt beispielhaft an, dass $\vartheta_0 \leq \alpha \leq \vartheta_1$ gilt. In diesem Fall ist auch $\frac{1-\vartheta_1}{1-\vartheta_0} < \frac{\vartheta_1}{\vartheta_0}$ und damit jede Ablehnschranke c_α geeignet, die die Bedingung

$$\frac{1-\vartheta_1}{1-\vartheta_0} < c_\alpha < \frac{\vartheta_1}{\vartheta_0}$$

erfüllt wegen

$$P_{\vartheta_0}(T(X) > c_\alpha) = P_{\vartheta_0}(X = 1) = \vartheta_0 \leq \alpha.$$

Für die Fehlerwahrscheinlichkeit 2. Art ergibt sich noch

III. Grundprinzipien der Statistik

$$\beta = P_{\vartheta_1}(T(\mathbf{X}) \leq c_\alpha) = P_{\vartheta_1}(X = 0) = 1 - \vartheta_1 \geq 1 - \alpha !$$

Eine solche Situation ist typisch in der Testtheorie: in der Regel lassen sich nicht beide Fehlerwahrscheinlichkeiten gleichzeitig minimieren; einer "kleinen" Fehlerwahrscheinlichkeit 1. Art entspricht oft eine "große" Fehlerwahrscheinlichkeit 2. Art.

Es ist deshalb extrem wichtig, bei der Benennung der Hypothesen auf die Konsequenzen einer Fehlentscheidung zu achten. Aus diesem Grund muss man diejenige Hypothese, bei der eine Fehlentscheidung fatal ist, entgegen der landläufigen Meinung grundsätzlich als die Alternative formulieren!

Oder anders ausgedrückt: möchte man eine bestimmte Hypothese mit großer Sicherheit verifizieren, muss man sie als Alternative formulieren! Denn eine irrtümliche Entscheidung für diese Hypothese wird ja gerade durch das vorgegebene (kleine) α kontrolliert.

Allerdings handelt man sich damit auf der anderen Seite ein, dass eine Entscheidung für die Nullhypothese praktisch wertlos ist, weil die Fehlerwahrscheinlichkeit 2. Art in der Regel *nicht* klein gehalten werden kann.

Bei dem gerade behandelten Beispiel sieht man, dass anstelle der Teststatistik $T(X)$ ebenso gut die Größe X selbst hätte verwendet werden können, denn es gilt hier (positiv lineare Transformation):

$$X = \frac{T(X) - \frac{1 - \vartheta_1}{1 - \vartheta_0}}{\frac{\vartheta_1 - 1 - \vartheta_1}{\vartheta_0} \cdot \frac{1 - \vartheta_1}{1 - \vartheta_0}} = \frac{1 - \vartheta_0}{\vartheta_1 - \vartheta_0} T(X) - \frac{1 - \vartheta_1}{\vartheta_1 - \vartheta_0}.$$

Wir wollen das obige Beispiel nun auf den Fall mehrerer Experimentwiederholungen erweitern.

Beispiel (Binomialtest): Wir wählen für $n \in \mathbb{N}$ Binomialverteilungen $Q_{\vartheta_i} = \bigotimes_{k=1}^n B(1, \vartheta_i)$ mit $\vartheta_i \in \Theta = (0, 1)$, $i \in \{1, 2\}$ und (Produkt-)Dichten f_i bezüglich des Produktmaßes $\bigotimes_{k=1}^n \#$, d.h.

$$f_i(\mathbf{x}) = P_{\vartheta_i}(\mathbf{X} = \mathbf{x}) = \vartheta_i^{\sum_{k=1}^n x_k} (1 - \vartheta_i)^{n - \sum_{k=1}^n x_k} \quad \text{für } \mathbf{x} \in \{0, 1\}^n.$$

Als Testgröße erhalten wir hieraus

$$T(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = \left(\frac{\vartheta_1}{\vartheta_0} \cdot \frac{1 - \vartheta_0}{1 - \vartheta_1} \right)^{\sum_{k=1}^n x_k} \left(\frac{1 - \vartheta_1}{1 - \vartheta_0} \right)^n \quad \text{für } \mathbf{x} \in \{0, 1\}^n.$$

III. Grundprinzipien der Statistik

$T(\mathbf{x})$ ist offensichtlich eine monotone Transformation von $\sum_{k=1}^n x_k$, so dass wir gleich auch

$S(\mathbf{X}) = \sum_{k=1}^n X_k$ als Testgröße verwenden können. Dies hat insbesondere den Vorteil, dass die Verteilung von $S(\mathbf{X})$ unter beiden Hypothesen bekannt ist, nämlich $P_{\vartheta_i}^{S(\mathbf{X})} = B(n, \vartheta_i)$, $i \in \{1, 2\}$.

Fall I: $\vartheta_0 < \vartheta_1$:

$S(\mathbf{X}) \leq c_\alpha$: Entscheidung für H_0

$S(\mathbf{X}) > c_\alpha$: Entscheidung für H_1

Hierbei ist die Ablehnschranke $c_\alpha \in \mathbb{Z}^+$ minimal so zu bestimmen, dass

$$P_{\vartheta_0}(S(\mathbf{X}) > c_\alpha) = \sum_{k=c_\alpha+1}^n \binom{n}{k} \vartheta_0^k (1-\vartheta_0)^{n-k} \leq \alpha$$

gilt. Für die Fehlerwahrscheinlichkeit 2. Art ergibt sich hier:

$$\beta = P_{\vartheta_1}(S(\mathbf{X}) \leq c_\alpha) = \sum_{k=0}^{c_\alpha} \binom{n}{k} \vartheta_1^k (1-\vartheta_1)^{n-k}.$$

Fall II: $\vartheta_0 > \vartheta_1$:

$S(\mathbf{X}) > c_\alpha$: Entscheidung für H_0

$S(\mathbf{X}) \leq c_\alpha$: Entscheidung für H_1

Hierbei ist die Ablehnschranke $c_\alpha \in \mathbb{Z}^+$ maximal so zu bestimmen, dass

$$P_{\vartheta_0}(S(\mathbf{X}) \leq c_\alpha) = \sum_{k=0}^{c_\alpha} \binom{n}{k} \vartheta_0^k (1-\vartheta_0)^{n-k} \leq \alpha$$

gilt. Für die Fehlerwahrscheinlichkeit 2. Art ergibt sich hier:

$$\beta = P_{\vartheta_1}(S(\mathbf{X}) > c_\alpha) = \sum_{k=c_\alpha+1}^n \binom{n}{k} \vartheta_1^k (1-\vartheta_1)^{n-k}.$$

Konkretes Anwendungsbeispiel: In zwei äußerlich nicht unterscheidbare Beutel werden rote und schwarze Kugeln gegeben; die Gesamtzahl der Kugeln sei in beiden Fällen 10. Im ersten Beutel befinden sich 3 rote, im zweiten Beutel 6 rote Kugeln. Es wird ein Beutel auf gut Glück ausgewählt; mit Hilfe des Binomialtests ist zu entscheiden, um welchen Beutel es sich handelt. Dazu werden aus dem Beutel $n = 20$ Kugeln mit Zurücklegen gezogen und die jeweilige Farbe notiert. Als Testniveau sei $\alpha = 0,05$ vereinbart.

III. Grundprinzipien der Statistik

Situation I:

Nullhypothese $H_0: \vartheta = 0,3$

Alternative $H_1: \vartheta = 0,6$

Hier ist die Ablehnschranke $c_\alpha \in \mathbb{Z}^+$ minimal so zu bestimmen, dass

$$P_{\vartheta_0}(S(\mathbf{X}) > c_\alpha) = \sum_{k=c_\alpha+1}^n \binom{n}{k} \vartheta_0^k (1-\vartheta_0)^{n-k} \leq \alpha$$

gilt. Tabelle:

n	ϑ	k	$P_\vartheta(S(\mathbf{X}) \leq k)$	$P_\vartheta(S(\mathbf{X}) > k)$	ϑ	k	$P_\vartheta(S(\mathbf{X}) \leq k)$	$P_\vartheta(S(\mathbf{X}) > k)$
20	0,3	0	0,00079792	0,99920208	0,6	0	0,00000001	0,99999999
		1	0,00763726	0,99236274		1	0,00000034	0,99999966
		2	0,03548313	0,96451687		2	0,00000504	0,99999496
		3	0,10708680	0,89291320		3	0,00004734	0,99995266
		4	0,23750778	0,76249222		4	0,00031703	0,99968297
		5	0,41637083	0,58362917		5	0,00161152	0,99838848
		6	0,60800981	0,39199019		6	0,00646588	0,99353412
		7	0,77227180	0,22772820		7	0,02102893	0,97897107
		8	0,88666854	0,11333146		8	0,05652637	0,94347363
		9	0,95203810	0,04796190		9	0,12752125	0,87247875
		10	0,98285518	0,01714482		10	0,24466280	0,75533720

Es ergibt sich also $c_\alpha = 9$. Bei 10 und mehr roten Kugeln in der Stichprobe (das sind 50% Rotanteil und höher) wird auf H_1 entschieden, ansonsten auf H_0 . Fehlerwahrscheinlichkeit 2. Art:

$$\beta = P_{\vartheta_1}(S(\mathbf{X}) \leq 9) = 0,1275.$$

Situation II:

Nullhypothese $H_0: \vartheta = 0,6$

Alternative $H_1: \vartheta = 0,3$

Hier ist die Ablehnschranke $c_\alpha \in \mathbb{Z}^+$ maximal so zu bestimmen, dass

$$P_{\vartheta_0}(S(\mathbf{X}) \leq c_\alpha) = \sum_{k=0}^{c_\alpha} \binom{n}{k} \vartheta_0^k (1-\vartheta_0)^{n-k} \leq \alpha$$

gilt. Aus obiger Tabelle ergibt sich hier $c_\alpha = 7$. Bei 7 und weniger roten Kugeln in der Stichprobe (das sind 35% Rotanteil und weniger) wird auf H_1 entschieden, ansonsten auf H_0 . Fehlerwahrscheinlichkeit 2. Art:

$$\beta = P_{\vartheta_1}(S(\mathbf{X}) > 7) = 0,2277.$$

III. Grundprinzipien der Statistik

Abschließend wollen wir exemplarisch noch ein Beispiel für eine stetige Verteilungssituation besprechen.

Beispiel Exponentialverteilung $\mathcal{E}(\vartheta)$:

Nullhypothese H_0 : $\vartheta = 2$

Alternative H_1 : $\vartheta = 5$

$\mathcal{Q}_{\vartheta_i} = \bigotimes_{k=1}^n \mathcal{E}(\vartheta_i)$ mit Lebesgue-Dichten $f_i(\mathbf{x}) = \prod_{k=1}^n \vartheta_i^n \exp(-\vartheta_i x_k)$, $\mathbf{x} \geq \mathbf{0}$ (komponentenweise), mit $\alpha = 0,05$ und $n = 10$:

Teststatistik:

$$T(\mathbf{x}) = \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} = \prod_{k=1}^n \left\{ \left(\frac{\vartheta_1}{\vartheta_0} \right)^n \exp((\vartheta_0 - \vartheta_1)x_k) \right\} = \exp\left((\vartheta_0 - \vartheta_1) \sum_{k=1}^n x_k \right) \cdot \prod_{k=1}^n \left(\frac{\vartheta_1}{\vartheta_0} \right)^n$$

oder alternativ (wegen Monotonie) $S(\mathbf{X}) = \sum_{k=1}^n X_k$.

$S(\mathbf{X}) > c_\alpha$: Entscheidung für H_0

$S(\mathbf{X}) \leq c_\alpha$: Entscheidung für H_1

Hier ist die Ablehnschranke $c_\alpha \in \mathbb{R}^+$ maximal so zu bestimmen, dass (Erlang-Verteilung, siehe Abschnitt II.5)

$$P_{\vartheta_0}(S(\mathbf{X}) \leq c_\alpha) = \int_0^{c_\alpha} \vartheta_0^n \frac{x^{n-1}}{(n-1)!} e^{-\vartheta_0 x} dx \leq \alpha \text{ bzw. } \int_0^{c_\alpha} \vartheta_0^n \frac{x^{n-1}}{(n-1)!} e^{-\vartheta_0 x} dx = \alpha$$

Hier gilt (mit Software berechnet): $c_\alpha = 2,7127$

Fehlerwahrscheinlichkeit 2. Art: $\beta = P_{\vartheta_1}(S(\mathbf{X}) > c_\alpha) = 1 - \int_0^{c_\alpha} \vartheta_1^n \frac{x^{n-1}}{(n-1)!} e^{-\vartheta_1 x} dx = 0,1317$.

Bemerkungen: Es ist wegen des Dichtequotienten-Tests nicht verwunderlich, dass die hier auftretenden Testgrößen (bzw. monotone Transformationen hiervon) früher schon als Maximum-Likelihood-Schätzer in Erscheinung getreten sind.

Man kann die oben beschriebene Vorgehensweise in vielen Fällen auch verwenden, wenn zusammengesetzte Hypothesen betrachtet werden, also Hypothesen der Form

Nullhypothese H_0 : $\vartheta \in \Theta_0$ gegen Alternative H_1 : $\vartheta \in \Theta_1$

mit $\Theta_0 \cap \Theta_1 = \emptyset$, $\Theta_0, \Theta_1 \subseteq \Theta$, insbesondere dann, wenn die – ggf. monoton transformierte – Teststatistik $T(\mathbf{X})$ für alle $\vartheta \in \Theta_0$ in gleicher Weise und mit derselben Entscheidungsregel gewählt werden kann.

III. Grundprinzipien der Statistik

Lautet die Entscheidungsregel wie am Anfang des Abschnitts etwa

$$T(\mathbf{X}) \leq c_\alpha : \text{Entscheidung für } H_0$$

$$T(\mathbf{X}) > c_\alpha : \text{Entscheidung für } H_1,$$

so ist die Ablehnschranke c_α hier so zu bestimmen, dass die (maximale) Fehlerwahrscheinlichkeit 1. Art durch $\alpha \in (0,1)$ beschränkt wird, also

$$\sup_{\vartheta \in \Theta_0} P_\vartheta(T(\mathbf{X}) > c_\alpha) \leq \alpha$$

gilt. Für die (maximale) Fehlerwahrscheinlichkeit 2. Art ergibt sich entsprechend

$$\beta = \sup_{\vartheta \in \Theta_1} P_\vartheta(T(\mathbf{X}) \leq c_\alpha).$$

Wie die obigen Herleitungen zeigen, kann in dieser Weise z.B. beim Binomialtest verfahren werden. Betrachtet man dazu das erste obige Anwendungsbeispiel in der Situation I mit der Modifikation

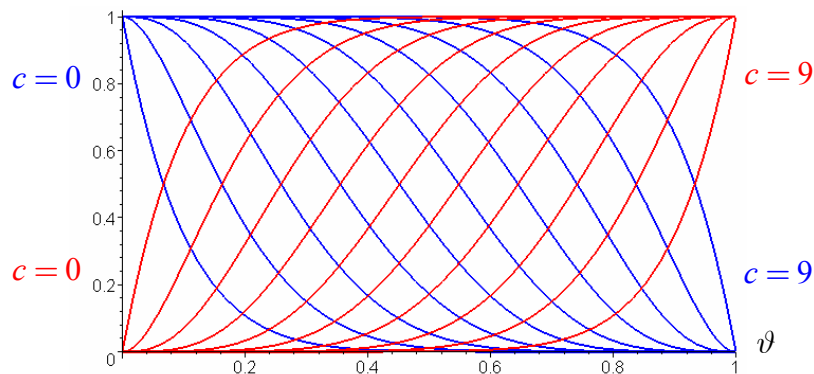
$$\text{Nullhypothese } H_0 : \vartheta \leq \vartheta_0 := 0,3 \quad \text{gegen} \quad \text{Alternative } H_1 : \vartheta > 0,3,$$

So ist die Ablehnschranke $c_\alpha \in \mathbb{Z}^+$ jetzt minimal so zu bestimmen, dass

$$\sup_{\vartheta \in \Theta_0} P_\vartheta(S(\mathbf{X}) > c_\alpha) = \sup_{\vartheta \leq \vartheta_0} \sum_{k=c_\alpha+1}^n \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k} = \sum_{k=c_\alpha+1}^n \binom{n}{k} \vartheta_0^k (1-\vartheta_0)^{n-k} \leq \alpha$$

gilt. Dies folgt aus der wachsenden Monotonie der Abbildungen $G : (\vartheta, c) \mapsto \sum_{k=c+1}^n \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k}$ für $\vartheta \in (0,1)$ und $c \in \{0,1,\dots,n\}$. Die Ablehnschranke bleibt also unverändert $c_\alpha = 9$, lediglich die (maximale) Fehlerwahrscheinlichkeit 2. Art wird größer; sie beträgt jetzt wegen der fallenden Monotonie der Abbildungen $F : (\vartheta, c) \mapsto \sum_{k=0}^c \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k} = 1 - G(\vartheta, c)$ für $\vartheta \in (0,1)$ und $c \in \{0,1,\dots,n\}$

$$\beta = \sup_{\vartheta \in \Theta_1} P_\vartheta(T(\mathbf{X}) \leq c_\alpha) = \sup_{\vartheta > \vartheta_0} \sum_{k=0}^{c_\alpha} \binom{n}{k} \vartheta^k (1-\vartheta)^{n-k} = \sum_{k=0}^{c_\alpha} \binom{n}{k} \vartheta_0^k (1-\vartheta_0)^{n-k} \geq 1 - \alpha !$$



Graph der Abbildungen $G(\vartheta, c)$ [rot] und $F(\vartheta, c)$ [blau] für $n = 10$, $\vartheta \in (0,1)$ und $c \in \{0,1,\dots,9\}$

III.3. Lineare und nicht-lineare Regression

Das hier besprochene Verfahren ist schon recht alt und geht in der Form der "Methode der kleinsten Quadrate" bereits auf Carl Friedrich Gauß zurück (1777 – 1855). Wir beschränken uns dabei auf einen rein datenanalytischen Zugang, die allgemeinere Theorie wird in der Regel im Rahmen der so genannten "Linearen Statistischen Modelle" behandelt (siehe etwa FAHRMEIR ET AL. (1996)).

Gegeben seien n Datenpaare (x_i, y_i) , $i = 1, \dots, n$ mit $n \in \mathbb{N}$, $n \geq 2$. Ferner wird hier zunächst vorausgesetzt, dass alle x_i paarweise verschieden sind. Gesucht wird eine *lineare Ausgleichsfunktion* $y = ax + b$ mit reellen Koeffizienten a, b , die die Datenpaare "möglichst gut" approximiert (so genannte *Regressionsgerade*). Als Gütemaß (Fehlermaß) legen wir die quadratische Abweichung zwischen den "Ist-Werten" y_i und den "Soll-Werten" $ax_i + b$ zu Grunde und erhalten damit folgendes Optimierungsproblem:

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

Dieses Problem kann mit den klassischen Methoden der Analysis gelöst werden; dazu betrachten wir die partiellen Ableitungen und setzen diese Null (notwendige Bedingung für – relative – Extrema):

$$\frac{\partial F}{\partial a} = 2 \sum_{i=1}^n (ax_i + b - y_i) \cdot x_i = 0$$

$$\frac{\partial F}{\partial b} = 2 \sum_{i=1}^n (ax_i + b - y_i) = 0$$

mit den Lösungen

$$\hat{a} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i - \left(\frac{1}{n} \sum_{i=1}^n x_i \right) \left(\frac{1}{n} \sum_{i=1}^n y_i \right)}{\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2} = \frac{s_{xy}}{s_{xx}}, \quad \hat{b} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{a} \cdot \frac{1}{n} \sum_{i=1}^n x_i = \bar{y} - \hat{a} \cdot \bar{x}.$$

Die Größen s_{xy} und s_{xx} können dabei als empirische Kovarianz bzw. empirische Varianz der Datenreihen aufgefasst werden, die Größen \bar{x} und \bar{y} entsprechen den üblichen arithmetischen Mittelwerten. Die Abbildung F ist nun aber global konvex wegen

$$\nabla^2 F = \begin{bmatrix} \frac{\partial^2 F}{\partial a^2} & \frac{\partial^2 F}{\partial a \partial b} \\ \frac{\partial^2 F}{\partial a \partial b} & \frac{\partial^2 F}{\partial b^2} \end{bmatrix} = 2 \begin{bmatrix} \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & n \end{bmatrix} \quad (\text{Hesse-Matrix})$$

mit

$$\det(HF) = \frac{\partial^2 F}{\partial a^2} \cdot \frac{\partial^2 F}{\partial b^2} - \left(\frac{\partial^2 F}{\partial a \partial b} \right)^2 = 2n^2 \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 \right) = 2n^2 \text{Var}(X) > 0$$

und

III. Grundprinzipien der Statistik

$$\frac{\partial^2 F}{\partial b^2} = n > 0,$$

wobei formal X eine diskret gleichverteilte Zufallsvariable über der Menge $\{x_1, \dots, x_n\}$ bezeichne. (Man beachte, dass wegen $n \geq 2$ der Ausdruck $\text{Var}(X) > 0$ und somit die Hesse-Matrix positiv definit ist!) Die Lösungen (Schätzwerte) \hat{a} und \hat{b} sind somit die eindeutig bestimmten globalen Minimumstellen der Funktion F .

Die obigen Rechnungen zeigen, dass das Verfahren auch dann noch funktioniert, wenn die x_i mehrfach auftreten (d.h. multiple Zuordnungen vorkommen), solange die Größe

$$s_{xx} = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2 > 0$$

bleibt, was genau dann eintritt, wenn wenigstens zwei der x_i voneinander verschieden sind. Eine solche Situation liegt beispielsweise vor, wenn es sich bei den y_i um Wiederholte Messwerte aus mehreren Experimenten (z.B. physikalische Körperausdehnung) bei teilweise gleichen x_i (z.B. eingestellte Temperaturen) handelt.

Manchmal wird auch nur eine lineare Ausgleichsfunktion der Form $y = ax$ gesucht (man spricht dann von einer *Regression durch den Nullpunkt*). In diesem Fall vereinfachen sich die obigen Rechnungen erheblich, mit der Lösung

$$\hat{a} = \frac{\frac{1}{n} \sum_{i=1}^n x_i y_i}{\frac{1}{n} \sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}.$$

Die Methode der kleinsten Quadrate ist auch anwendbar, wenn nicht-lineare Zusammenhänge zwischen den x - und y -Werten unterstellt werden. Dazu bedient man sich zweckmäßigerweise geeigneter *linear unabhängiger Funktionensysteme* $\{f_k\}_{k \in \mathbb{N}}$, z.B. Polynome unterschiedlichen Grades. Das abstrakte Modell ist hier

$$y = \sum_{k=1}^K a_k \cdot f_k(x) \quad \text{mit } K \in \mathbb{N} \text{ und reellen Koeffizienten } a_1, \dots, a_K.$$

Das zugehörige quadratische Optimierungsproblem lautet dann sinngemäß

$$\min_{a_1, \dots, a_K}! F(a_1, \dots, a_K) = \sum_{i=1}^n \left(y_i - \sum_{k=1}^K a_k f_k(x_i) \right)^2 = \min_{\mathbf{a}}! F(\mathbf{a}) = \|\mathbf{y} - D\mathbf{a}\|^2$$

III. Grundprinzipien der Statistik

mit dem Koeffizientenvektor $\mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_K \end{pmatrix}$, dem Wertevektor $\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ und der $n \times K$ -(Design-)Matrix

$D = [f_k(x_i)]_{i=1 \dots n, k=1 \dots K}$. Wegen

$$F(\mathbf{a}) = \|\mathbf{y} - D\mathbf{a}\|^2 = (\mathbf{y} - D\mathbf{a})^T (\mathbf{y} - D\mathbf{a}) = \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T D\mathbf{a} + \mathbf{a}^T D^T D\mathbf{a}$$

ergibt sich für den Gradienten von F :

$$\nabla F = -2D^T \mathbf{y} + 2D^T D\mathbf{a}.$$

Durch Nullsetzen der partiellen Ableitungen ergibt sich also das Gleichungssystem

$$D^T D\mathbf{a} = D^T \mathbf{y}$$

mit der Lösung

$$\hat{\mathbf{a}} = (D^T D)^{-1} D^T \mathbf{y}.$$

Man beachte, dass die Matrix $D^T D$ stets positiv semidefinit ist (vgl. Abschnitt II.9). Die Inverse von $D^T D$ existiert also genau dann, wenn $D^T D$ positiv definit ist (d.h. alle Eigenwerte von $D^T D$ sind positiv). Dieser Fall liegt vor, wenn die Matrix D vollen Rang hat. Ferner gilt für die Matrix der zweiten partiellen Ableitungen (Hesse-Matrix)

$$\nabla^2 F = 2D^T D,$$

was im Fall der positiven Definitheit bedeutet, dass bei der gefundenen Lösung ein Minimum vorliegt.

Im Fall der klassischen linearen Regression ist etwa $\mathbf{a} = \begin{pmatrix} b \\ a \end{pmatrix}$ mit $D = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, also

$$D^T D = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \cdot \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^n x_i \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \end{bmatrix}$$

und

$$D^T \mathbf{y} = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \cdot \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \end{pmatrix},$$

woraus sich wieder die oben angegebene Lösung ergibt.

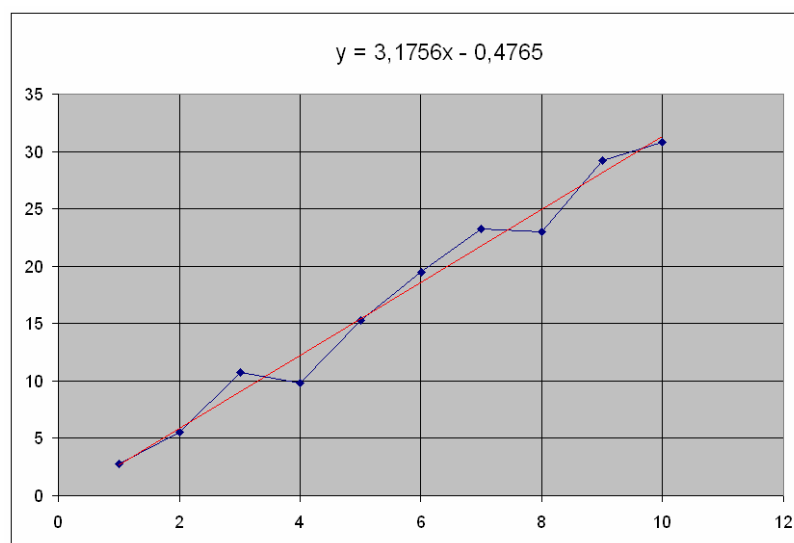
III. Grundprinzipien der Statistik

Beispiel: An den folgenden Datensatz soll einmal eine lineare, zum anderen eine quadratische Funktion angepasst werden:

x_i	1	2	3	4	5	6	7	8	9	10
y_i	2,771	5,522	10,753	9,828	15,262	19,480	23,218	23,030	29,231	30,797

1. *Lineare Regression:* Es ist

$$s_{xy} = 26,199 \quad s_{xx} = 8,25 \quad \hat{a} = \frac{s_{xy}}{s_{xx}} = 3,17558 \quad \hat{b} = \bar{y} - \hat{a} \cdot \bar{x} = -0,47647$$



lineare Regression

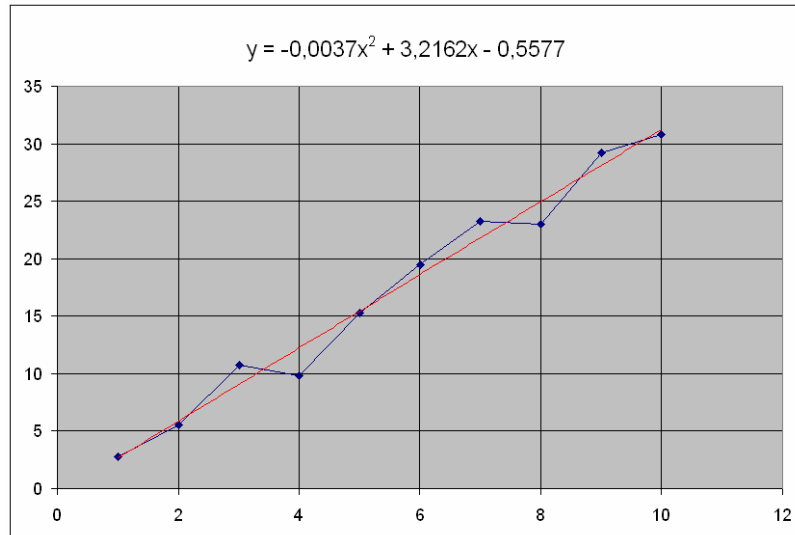
2. *Quadratische Regression:* Hier ist

$$D = \begin{bmatrix} 1 & x_1 & x_1^2 \\ \vdots & \vdots & \vdots \\ 1 & x_{10} & x_{10}^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \\ 1 & 7 & 49 \\ 1 & 8 & 64 \\ 1 & 9 & 81 \\ 1 & 10 & 100 \end{bmatrix}, \quad D^T D = \begin{bmatrix} 10 & 55 & 385 \\ 55 & 385 & 3025 \\ 385 & 3025 & 25333 \end{bmatrix},$$

III. Grundprinzipien der Statistik

$$(D^T D)^{-1} = \frac{1}{2640} \begin{bmatrix} 3651 & -1386 & 110 \\ -1386 & 637 & -55 \\ 110 & -55 & 5 \end{bmatrix}, \quad D^T \mathbf{y} = \begin{pmatrix} 169,892 \\ 1196,391 \\ 9420,727 \end{pmatrix}$$

$$\hat{\mathbf{a}} = (D^T D)^{-1} D^T \mathbf{y} = \begin{pmatrix} -0,5577 \\ 3,2162 \\ -0,0037 \end{pmatrix}$$



quadratische Regression

Bisher haben wir in dem obigen Ansatz keine *zufälligen* Einflüsse explizit modelliert. Man kann dies aber leicht erreichen, wenn wir etwa annehmen, dass die Messergebnisse \mathbf{y} Realisationen eines Zufallsvektors der Form

$$\mathbf{Y} = D\mathbf{a} + \boldsymbol{\varepsilon}$$

mit einem Fehlervektor (Zufallsvektor) $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T$ sind, der paarweise unkorrelierte Koordinaten besitzt mit Erwartungswerten $E(\varepsilon_i) = 0$ und (gleichen) Varianzen $Var(\varepsilon_i) = \sigma^2 > 0$ für $i = 1, \dots, n$. Ist der Fehlervektor beispielsweise multivariat normalverteilt, so erhält man das Verteilungsmodell

$$P^{\mathbf{Y}} = \mathcal{N}(D\mathbf{a}, \sigma^2 \mathbf{I})$$

mit der Einheitsmatrix \mathbf{I} . In diesem Fall stimmt die angegebene Lösung des quadratischen Minimierungsproblems mit dem Maximum-Likelihood-Schätzer für den Parametervektor \mathbf{a} überein; dies ist sofort aus der Form der zugehörigen Dichte ersichtlich (vgl. Abschnitt II.9.):

$$f(\mathbf{y}) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - D\mathbf{a})^T(\mathbf{y} - D\mathbf{a})\right) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - D\mathbf{a}\|^2\right), \quad \mathbf{y} \in \mathbb{R}^n.$$

Modelle dieser Art werden in dem *Lineare statistische Modelle* genannten Teilbereich der Statistik behandelt; siehe etwa PRUSCHA (2000), Kapitel III.

III.4. Lage-Skalen-Familien, Q-Q-Plots

Wir kommen hier auf die schon in Abschnitt III angesprochene Situation einer durch eine Zufallsvariable Z induzierte *Lage-Skalen-Familie* $\mathcal{V}(\mu, \sigma)$ mit $\mu \in \mathbb{R}$, $\sigma > 0$ zurück, die durch den Übergang zu der positiv-lineartransformierten Zufallsvariablen

$$X := \mu + \sigma Z$$

entsteht. Besitzt Z die Verteilungsfunktion F_Z , so besitzt X die Verteilungsfunktion

$$F_X(x) = P(\mu + \sigma Z \leq x) = F_Z\left(\frac{x - \mu}{\sigma}\right), \quad x \in \mathbb{R}.$$

Gilt insbesondere $E(Z) = 0$, $Var(Z) = 1$, so folgt

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

Zu solchen Familien gehören, wie schon erwähnt, natürlicherweise die Normalverteilungen $N(\mu, \sigma^2)$. Ist $\mu = 0$, so spricht man auch nur von einer *Skalen-Familie* (in $\sigma > 0$). Eine solche ist die Familie der $\Gamma(\alpha, \lambda)$ -Verteilungen bei festem $\alpha > 0$ in $\sigma = \frac{1}{\lambda} > 0$.

Eine einfache graphische Methode zur Schätzung der Parameter μ und σ (und gleichzeitig zum Prüfen der Hypothese, ob der den Daten zu Grunde liegende Verteilungstyp P^Z vorliegt) bilden die so genannten *Q-Q-Plots*. Wir betrachten dazu grundsätzlich den Fall, dass die Verteilungsfunktion F_Z im Bereich $\{x \in \mathbb{R} \mid 0 < F_Z(x) < 1\}$ *streng monoton wachsend* und *stetig* ist.

Nach obigem gilt umgekehrt auch:

$$F_Z(x) = P(Z \leq x) = P\left(\frac{X - \mu}{\sigma} \leq x\right) = P(X \leq \mu + \sigma x) = F_X(\mu + \sigma x), \quad x \in \mathbb{R}.$$

Die *Quantilfunktion* Q_Z ist definiert durch die (Pseudo-)Inverse der Verteilungsfunktion:

$$Q_Z(u) = F_Z^{-1}(u) \text{ bzw. } F_Z(Q_Z(u)) = u, \quad 0 < u < 1.$$

Für die Zufallsvariable X ergibt sich daraus folgender Zusammenhang:

$$Q_X(u) = \mu + \sigma Q_Z(u), \quad 0 < u < 1,$$

wie man durch Vergleich der Argumente feststellen kann:

$$u = F_X(Q_X(u)) = F_Z(Q_Z(u)) = F_X(\mu + \sigma Q_Z(u)), \quad 0 < u < 1.$$

Trägt man also in einem Koordinatensystem die Paare $(Q_Z(u), Q_X(u))$ für $0 < u < 1$ gegeneinander auf, so erhält man eine Gerade mit (positiver) Steigung σ und Achsenabschnitt μ .

III. Grundprinzipien der Statistik

Diese Beobachtung kann man sich insbesondere zu Nutze machen, um zu prüfen, ob einem Datensatz (x_1, \dots, x_n) eine durch Z induzierte Lage- Skalen-Familie zu Grunde liegt

Dazu trägt man in ein Koordinatensystem für geeignete Werte u_1, \dots, u_n die aus den Beobachtungen abgeleiteten Paare $(Q_Z(u_k), \hat{Q}_X(u_k))$, $k=1, \dots, n$ gegeneinander auf, wobei \hat{Q}_X die *empirische Quantilfunktion* bezeichnet; das ist die Pseudo-Inverse der (rechtsseitig stetigen) empirischen Verteilungsfunktion

$$\hat{F}(x) := \frac{1}{n} \sum_{k=1}^n \mathbb{1}_{(-\infty, x]}(x_k), \quad x \in \mathbb{R}.$$

Es handelt sich also formal um eine (von den Beobachtungen abhängige) Verteilungsfunktion mit Sprüngen der Höhe $1/n$ an den durch die Beobachtungen vorgegebenen Stellen. Um numerische Komplikationen zu vermeiden, wählt man hier gelegentlich auch einen anderen Vorfaktor als $1/n$. Dies werden wir für das hier besprochene Verfahren ebenfalls tun.

Ordnet man die Daten (x_1, \dots, x_n) der Größe nach mit den Werten (sog. *Ordnungsstatistiken*)

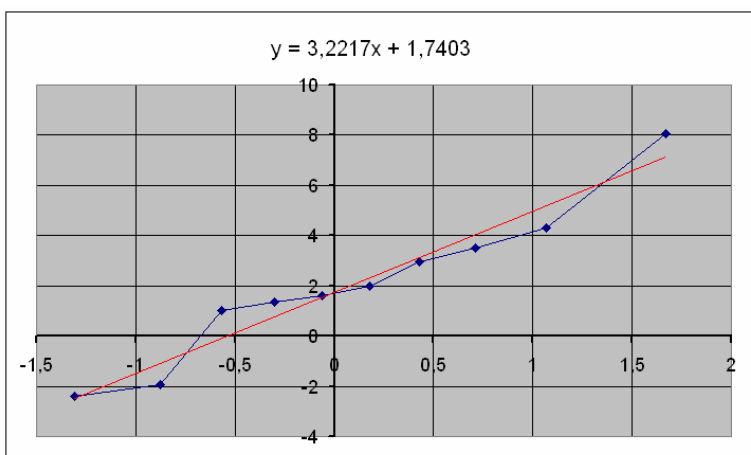
$x_{(1)} < \dots < x_{(n)}$ und wählt man z.B. $u_k = \frac{k}{n + \tau}$ für $k=1, \dots, n$ mit $0 < \tau \leq 1$, so ist $\hat{Q}_X(u_k) \approx x_{(k)}$,

d.h. man trägt die geordneten Beobachtungswerte ("*beobachtete Werte*") gegen die Werte $Q_Z(u_k)$ ("*theoretische Quantile*") auf.

Mit Hilfe einer Ausgleichsgeraden durch diese n Punktepaare (z.B. durch lineare Regression, siehe Abschnitt III.3.) lässt sich dann zunächst durch visuelle Überprüfung abschätzen, ob die getroffene Verteilungsannahme haltbar ist. Dazu sollten die Punktepaare nicht „zu weit“ von der Ausgleichsgeraden entfernt sein. Findet man mit dieser Methode die Verteilungsannahme gerechtfertigt, lassen sich anschließend durch den Achsenabschnitt und die Steigung der Geraden die unbekanntenen Parameter μ und σ schätzen.

Beispiel: Der angegebene Datensatz mit $n=10$ entstammt einer simulierten Stichprobe aus einer $\mathcal{N}(2,9)$ -Normalverteilung. Für den Q-Q-Plot wählen wir $\tau = \frac{1}{2}$.

i	1	2	3	4	5	6	7	8	9	10
x_i	3,5139	2,9697	8,0516	4,3234	-2,3952	1,3471	-1,9499	1,6026	1,0095	1,9759



$$\hat{\mu} = 1,7403$$

$$\hat{\sigma} = 3,2217$$

zum Vergleich:

$$\mu = 2$$

$$\sigma = 3$$

III. Grundprinzipien der Statistik

Nach Abschnitt III.3 lassen sich die aus der linearen Regression gewonnenen Schätzwerte $\hat{\mu}$ und $\hat{\sigma}$ auch so schreiben:

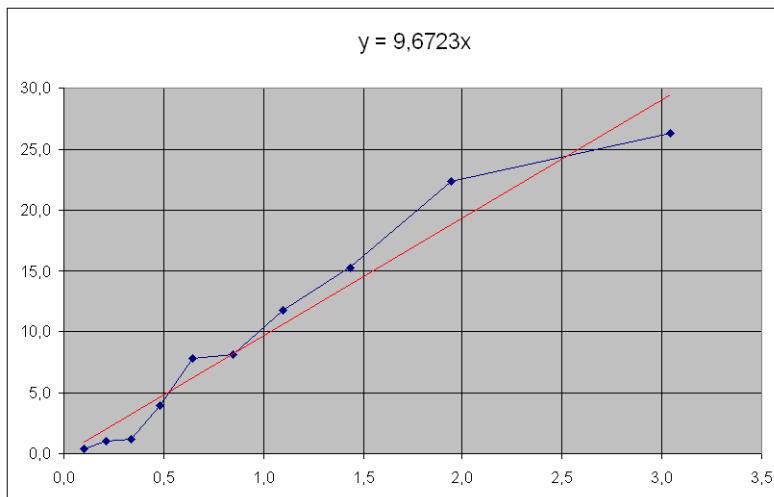
$$\hat{\sigma} = \frac{\frac{1}{n} \sum_{k=1}^n x_{(k)} Q_Z(u_k) - \left(\frac{1}{n} \sum_{k=1}^n x_{(k)} \right) \left(\frac{1}{n} \sum_{k=1}^n Q_Z(u_k) \right)}{\frac{1}{n} \sum_{k=1}^n Q_Z^2(u_k) - \left(\frac{1}{n} \sum_{k=1}^n Q_Z(u_k) \right)^2}, \quad \hat{\mu} = \frac{1}{n} \sum_{k=1}^n x_{(k)} - \frac{\hat{\sigma}}{n} \sum_{k=1}^n Q_Z(u_k).$$

Für reine Skalenfamilien kann man ein entsprechendes Verfahren mit einer linearen Regression *durch den Nullpunkt* anwenden, weil hier ja gerade $\mu = 0$ gilt. Es ergibt sich analog

$$\hat{\sigma} = \frac{\frac{1}{n} \sum_{k=1}^n x_{(k)} Q_Z(u_k)}{\frac{1}{n} \sum_{k=1}^n Q_Z^2(u_k)} = \frac{\sum_{k=1}^n x_{(k)} Q_Z(u_k)}{\sum_{k=1}^n Q_Z^2(u_k)}.$$

Beispiel: Der angegebene Datensatz mit $n = 10$ entstammt einer simulierten Stichprobe aus einer Exponentialverteilung $\mathcal{E}(\lambda)$ mit Parameter $\lambda = \frac{1}{10}$:

i	1	2	3	4	5	6	7	8	9	10
x_i	8,135	22,339	1,218	0,372	7,780	26,271	11,749	15,254	1,010	3,906



$$\hat{\sigma} = 9,6723$$

zum Vergleich:

$$\sigma = \frac{1}{\lambda} = 10$$

IV. Stochastische Prozesse

Stochastische Prozesse werden häufig zur Modellierung stochastischer Phänomene herangezogen, die eine strukturelle zeitliche Komponente besitzen, z.B. in der Warteschlangentheorie (Bediensysteme) oder in der stochastischen Finanzmathematik (Aktienkursverläufe, Zinsstrukturkurven), aber auch in der Physik (radioaktiver Zerfall, Wärmeleitung, Brown'sche Bewegung).

Wir werden in diesem Kapitel die Theorie überwiegend nur plakativ (ohne ausführliche Beweise) behandeln, um damit einen Einstieg in nachfolgende speziellere Lehrveranstaltungen zu erleichtern. Der Text orientiert sich überwiegend an der Monographie von MATHAR UND PFEIFER (1990); vgl. auch die einschlägigen Kapitel in KLENKE (2008).

IV.1 Markoff-Ketten

Markoff-Ketten beschreiben das zeitliche Verhalten von stochastischen Modellen mit einer sehr schwachen Art stochastischer Abhängigkeit, die mit einer gewissen Art von „Gedächtnislosigkeit“ zu vergleichen ist.

Markoff-Modelle finden z.B. in der Versicherungsmathematik vor allem Anwendung bei Bonus-Malus-Systemen in der Kraftfahrzeug-Haftpflichtversicherung und in der modernen Lebensversicherung, sie sind aber auch Grundlage für die allgemeineren Markoff-Prozesse, die speziell in Form von (geometrischen) Wiener-Prozessen oder Lösungen stochastischer Differenzialgleichungen in der Finanzmathematik (Asset-Liability-Management (ALM); Kapitalanlage-Risiken) Verwendung finden.

Definition 49. Unter einer *Markoff-Kette vom Typ I* versteht man eine Folge $\{X_n\}_{n \in \mathbb{Z}^+}$ reellwertiger Zufallsvariablen mit Werten in einer höchstens abzählbaren (meist endlichen) Menge \mathcal{S} (Zustandsraum genannt), die folgende fundamentale Eigenschaft („Gedächtnislosigkeit“) besitzt:

$$P(X_n = x_n | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

für alle Zustände $x_0, \dots, x_n \in \mathcal{S}$ und alle $n \in \mathbb{N}$. Die Markoff-Kette heißt (zeitlich) *homogen*, wenn die Übergangswahrscheinlichkeiten $P(X_n = x_n | X_{n-1} = x_{n-1})$ nicht von $n \in \mathbb{N}$ abhängen. (Um Probleme zu vermeiden, wird dabei $P(A|B) = P(A)$ gesetzt, wenn $P(B) = 0$ gilt.)

Bemerkung: Wegen der Abzählbarkeit des Zustandsraums können wir ohne Beschränkung der Allgemeinheit gleich $\mathcal{S} \subseteq \mathbb{N}$ annehmen. Die Übergangswahrscheinlichkeiten

$$p_{ij}(n) := P(X_n = j | X_{n-1} = i), \quad i, j \in \mathcal{S}$$

lassen sich dann zu so genannten *Übergangsmatrizen* (stochastischen Matrizen)

$$\Pi_n := [p_{ij}(n)]_{i,j \in \mathcal{S}}$$

zusammenfassen, mit denen die Verteilung der Markoff-Kette eindeutig beschrieben werden kann. Wir bezeichnen noch mit

$$\mathbf{p}(n) := [P(X_n = i)]_{i \in \mathcal{S}}$$

den Zeilenvektor der Wahrscheinlichkeitsverteilung von X_n für $n \in \mathbb{Z}^+$.

IV. Stochastische Prozesse

Lemma 51. Markoff-Ketten vom Typ I besitzen die folgenden strukturellen Eigenschaften:

- Für die Verteilung von X_n mit $n \in \mathbb{N}$ gilt:

$$\mathbf{p}(n) = \mathbf{p}(n-1) \cdot \Pi_n = \dots = \mathbf{p}(0) \cdot \Pi_1 \cdot \Pi_2 \cdot \dots \cdot \Pi_n$$

- Für die höheren Übergangswahrscheinlichkeiten gilt:

$$P(X_n = j | X_k = i) = (\Pi_{k+1} \cdot \Pi_{k+2} \cdot \dots \cdot \Pi_n)_{i,j} \text{ für alle } i, j \in \mathcal{S}, 0 \leq k < n$$

- Für die gemeinsame Verteilung gilt:

$$P(X_k = i_k, \dots, X_n = i_n) = p_{i_k}(k) \cdot \prod_{j=k+1}^n p_{i_{j-1} i_j}(j)$$

für alle $0 \leq k < n$ und $i_k, \dots, i_n \in \mathcal{S}$.

Definition 50. Unter *Markoff-Ketten vom Typ II* verstehen wir allgemeiner Markoff-Ketten mit Werten in beliebigen (in der Regel überabzählbaren) Zustandsräumen (Messräumen) $(\mathcal{S}, \mathcal{A})$, in denen reguläre Versionen bedingter Verteilungen existieren (vgl. Kapitel II.8).

Die fundamentale Eigenschaft der „Gedächtnislosigkeit“ lautet hier analog:

$$P(X_n \in A | X_0 = x_0, \dots, X_{n-1} = x_{n-1}) = P(X_n \in A | X_{n-1} = x_{n-1}) \text{ [fast sicher]}$$

für alle Zustände $x_0, \dots, x_{n-1} \in \mathcal{S}$, $A \in \mathcal{A}$ und alle $n \in \mathbb{N}$.

Die Ausdrücke

$$Q_x^{[n]}(A) := P(X_n \in A | X_{n-1} = x), \quad x \in \mathcal{S}, A \in \mathcal{A}, n \in \mathbb{N}$$

heissen hier *Übergangskerne*. Auf Grund der vorausgesetzten Regularität der bedingten Wahrscheinlichkeiten ist dabei jedes $Q_x^{[n]}$ ein Maß auf \mathcal{A} für alle $x \in \mathcal{S}$.

Bemerkung: Die gemeinsame Verteilung der Markoff-Kette erhält man komplizierter durch Integration:

$$P\left(\bigcap_{k=0}^n \{X_k \in A_k\}\right) = \int_{A_0} \dots \int_{A_n} \prod_{k=1}^n Q_{x_{n-k}}^{[n-k+1]}(dx_{n-k+1}) P^{X_0}(dx_0)$$

für $A_0, \dots, A_n \in \mathcal{A}$, $n \in \mathbb{N}$.

Die Situation wird etwas einfacher, wenn die Übergangskerne und die Anfangsverteilung P^{X_0} (Übergangs-)Dichten bezüglich eines dominierenden Maßes besitzen, beispielsweise bezüglich des Lebesgue-Maßes im Fall $(\mathcal{S}, \mathcal{A}) = (\mathbb{R}^1, \mathcal{B}^1)$, etwa

$$Q_x^{[n]}(A) = P(X_n \in A | X_{n-1} = x) = \int_A f^{[n]}(x, y) dy, \quad P(X_0 \in A) = \int_A f^{[0]}(y) dy$$

für $x \in \mathbb{R}^1$, $A \in \mathcal{B}^1$, $n \in \mathbb{N}$. In diesem Fall ist dann

IV. Stochastische Prozesse

$$g^{[n]}(x_0, x_1, \dots, x_n) = f^{[0]}(x_0) \cdot \prod_{k=1}^n f^{[k]}(x_{k-1}, x_k), \quad x_0, x_1, \dots, x_n \in \mathbb{R}^1$$

eine (gemeinsame) Dichte von (X_0, X_1, \dots, X_n) für $n \in \mathbb{N}$.

Markoff-Ketten vom Typ II heißen analog (zeitlich) *homogen*, wenn die Übergangskerne bzw. Übergangsdichten nicht von n abhängen, also die hochgestellten Indices $[n]$ entfallen.

Markoff-Ketten beider Typen können in gewisser Weise kanonisch rekursiv konstruiert werden, wodurch ihr stochastisches Verhalten besonders gut verdeutlicht wird.

Satz 53 (Konstruktionssatz): Es sei X_0 eine Zufallsvariable mit Werten in \mathcal{S} und $\{U_n\}_{n \in \mathbb{N}}$ eine Folge (auch von X_0) unabhängiger reellwertiger Zufallsvariablen. Ferner seien für $n \in \mathbb{N}$ die Abbildungen $f_n : \mathcal{S} \times \mathbb{R} \rightarrow \mathcal{S}$ messbar. Dann bildet die rekursiv definierte Folge $\{X_n\}_{n \in \mathbb{N}}$ mit

$$X_n := f_n(X_{n-1}, U_n) \quad \text{für } n \in \mathbb{N}$$

eine Markoff-Kette mit den Übergangswahrscheinlichkeiten

$$p_{ij}(n) = P(f_n(i, U_n) = j) \quad \text{für } i, j \in \mathcal{S}, n \in \mathbb{N} \text{ bei Ketten vom Typ I}$$

bzw. den Übergangskernen

$$Q_x^{[n]}(A) = P(f_n(x, U_n) \in A) \quad \text{für } x \in \mathcal{S}, A \in \mathcal{A}, n \in \mathbb{N} \text{ bei Ketten vom Typ II.}$$

Since die Abbildungen f_n unabhängig von $n \in \mathbb{N}$ und die U_n identisch verteilt, so ist die Markoff-Kette homogen.

Beweis: Dies folgt direkt aus der Ersetzungsformel, Satz 49. ■

Bemerkung: Der in Satz 53 dargestellte Sachverhalt ist sogar charakteristisch für Markoff-Ketten, denn es lässt sich für *jede* Markoff-Kette $\{\tilde{X}_n\}_{n \in \mathbb{Z}^+}$ ein wahrscheinlichkeitstheoretisches Modell $\{X_n\}_{n \in \mathbb{Z}^+}$ der obigen rekursiven Art angeben mit der Eigenschaft, dass alle endlich-dimensionalen Randverteilungen von $\{\tilde{X}_n\}_{n \in \mathbb{Z}^+}$ und $\{X_n\}_{n \in \mathbb{Z}^+}$ übereinstimmen, d.h. beide Modelle sind statistisch nicht unterscheidbar.

Eine wichtige Klasse von Markoff-Ketten (beider Typen) sind Markoff-Ketten mit *unabhängigen Zuwächsen*, die sich aus dem Konstruktionssatz für geeignete Zustandsräume $\mathcal{S} \subseteq \mathbb{R}$ (Gruppen bezüglich der Addition) ergeben vermöge der Abbildungen

$$f_n(x, y) = x + y, \quad x, y \in \mathcal{S}.$$

In diesem Fall gilt nämlich $U_n = X_n - X_{n-1}$ für alle $n \in \mathbb{N}$, d.h. die Folge $\{X_0, X_n - X_{n-1}\}_{n \in \mathbb{N}}$ ist stochastisch unabhängig.

IV. Stochastische Prozesse

Für den Rest dieses Abschnittes werden wir uns nur noch mit *homogenen* Markoff-Ketten vom Typ I beschäftigen. In diesem Fall reicht die Ein-Schritt-Übergangsmatrix $\Pi = \Pi_1$ allein zur Beschreibung des stochastischen Verhaltens der Markoff-Kette aus. Speziell gilt hier für die n -Schritt-Übergangswahrscheinlichkeiten:

$$\begin{aligned} p_{ij}^{(n)} &:= P(X_n = j | X_0 = i) = (\Pi^n)_{ij} \\ p_{ij}^{(n)} &= P(X_{n+k} = j | X_k = i) \\ p_{ij}^{(n+k)} &= \sum_{\ell \in \mathcal{S}} p_{i\ell}^{(n)} p_{\ell j}^{(k)} \quad (\text{Chapman-Kolmogoroff-Gleichungen}) \end{aligned}$$

für alle $i, j \in \mathcal{S}$, $n, k \in \mathbb{Z}^+$.

Für das Langzeitverhalten einer homogenen Markoff-Kette vom Typ I spielt der mögliche Grenzwert der Matrizen-Folge $\{\Pi^n\}_{n \in \mathbb{N}}$ eine wichtige Rolle. Hierfür betrachten wir die *Eigenwerte* der Ein-Schritt-Übergangsmatrix Π , für die im Falle der Endlichkeit von \mathcal{S} gilt:

- Für jeden (ggf. komplexen) Eigenwert λ von Π gilt $|\lambda| \leq 1$.
- $\lambda = 1$ ist stets ein Eigenwert von Π (zum Rechtseigenvektor $\mathbf{e} = (1, \dots, 1)^T$)
- $\lambda = 0$ ist ein Eigenwert von Π , wenn Π nicht vollen Rang hat.

Die ersten beiden Aussagen sind eine direkte Konsequenz aus der Tatsache, dass die Zeilensummen von Übergangsmatrizen den Wert 1 haben. Genauer gilt: Ist $\mathcal{S} = \{1, \dots, s\}$ mit $s = \#(\mathcal{S}) \in \mathbb{N}$ und $\mathbf{x} = (x_1, \dots, x_s)^T$ ein (nicht-trivialer) Eigenvektor zum Eigenwert $\lambda \in \mathbb{C}$ von Π , so gilt:

$$|\lambda| \cdot |x_i| = |\lambda \cdot x_i| = \left| \sum_{j=1}^s p_{ij} x_j \right| \leq \sum_{j=1}^s p_{ij} |x_j| \leq \sum_{j=1}^s p_{ij} \max_{k \in \mathcal{S}} \{|x_k|\} = \max_{k \in \mathcal{S}} \{|x_k|\} \quad \text{für alle } i \in \mathcal{S}$$

und damit auch

$$|\lambda| \cdot \max_{i \in \mathcal{S}} \{|x_i|\} \leq \max_{k \in \mathcal{S}} \{|x_k|\},$$

woraus wegen $\mathbf{x} \neq \mathbf{0}$ und damit auch $\max_{k \in \mathcal{S}} \{|x_k|\} \neq 0$ die erste Aussage folgt.

Satz 54 (Konvergenzsatz für homogene Markoff-Ketten mit endlichem Zustandsraum):

Ist $\lambda_1 = 1$ der einzige Eigenwert λ von Π mit $|\lambda| = 1$ und sind alle übrigen Eigenwerte von Π paarweise voneinander verschieden, so gilt

$$\Pi^* := \lim_{n \rightarrow \infty} \Pi^n = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_r \\ \pi_1 & \pi_2 & \cdots & \pi_r \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_r \end{pmatrix}, \quad \text{wobei } r = \#(\mathcal{S}),$$

mit dem eindeutig bestimmten normierten Linkseigenvektor $\mathbf{p}^* = (\pi_1 \ \pi_2 \ \cdots \ \pi_r)$ von Π . \mathbf{p}^* heißt auch *stationäre Verteilung* der Markoff-Kette.

IV. Stochastische Prozesse

Beweis: Jede Potenz Π^n der Übergangsmatrix kann berechnet werden über die Formel

$$\Pi^n = T \Delta^n T^{-1}, \quad n \in \mathbb{N},$$

wobei Δ die Matrix der Eigenwerte und T eine beliebige Matrix zugehöriger Rechtseigenvektoren

ist. Es folgt dann nach Voraussetzung $\lim_{n \rightarrow \infty} \Delta^n = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix}$ und somit

$$\Pi^* = \lim_{n \rightarrow \infty} \Pi^n = T \left(\lim_{n \rightarrow \infty} \Delta^n \right) T^{-1} = T \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 \end{pmatrix} T^{-1} = \begin{pmatrix} \pi_1 & \pi_2 & \dots & \pi_r \\ \pi_1 & \pi_2 & \dots & \pi_r \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \dots & \pi_r \end{pmatrix},$$

mit

Bemerkung: Unter den gemachten Annahmen folgt damit sofort auch:

Für jede beliebige Anfangsverteilung $\mathbf{p}(0)$ konvergiert die Markoff-Kette in Verteilung gegen die stationäre Verteilung \mathbf{p}^* . Ist schon $\mathbf{p}(0) = \mathbf{p}^*$, so sind alle Randverteilungen der Kette mit \mathbf{p}^* identisch (was den Namen erklärt).

Beispiel: Wir betrachten eine Markoff-Kette mit drei Zuständen. Für die Übergangswahrscheinlichkeiten sei angenommen:

$$\Pi = \begin{bmatrix} 0,8 & 0,2 & 0 \\ 0,1 & 0,5 & 0,4 \\ 0,1 & 0,2 & 0,7 \end{bmatrix}.$$

Das charakteristische Polynom lautet (\mathbf{I} = Einheitsmatrix)

$$\varphi(\lambda) = \det(\lambda \mathbf{I} - \Pi) = \lambda^3 - 2\lambda^2 + 1,21\lambda - 0,21,$$

mit den drei Nullstellen $\lambda_1 = 1$, $\lambda_2 = 0,7$, $\lambda_3 = 0,3$.

Die Voraussetzungen des Konvergenzsatzes sind also erfüllt; eine mögliche Wahl der T -Matrix ist z.B.:

$$T = \begin{bmatrix} 1 & -2 & 2 \\ 1 & 1 & -5 \\ 1 & 1 & 2 \end{bmatrix} \quad \text{mit} \quad T^{-1} = \frac{1}{21} \begin{bmatrix} 7 & 6 & 8 \\ -7 & 0 & 7 \\ 0 & -3 & 3 \end{bmatrix}.$$

Für die Grenzmatrix ergibt sich hieraus

IV. Stochastische Prozesse

$$\lim_{n \rightarrow \infty} \Pi^n = T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} T^{-1} = \frac{1}{21} \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 7 & 6 & 8 \\ -7 & 0 & 7 \\ 0 & -3 & 3 \end{bmatrix} = \frac{1}{21} \begin{bmatrix} 7 & 6 & 8 \\ 7 & 6 & 8 \\ 7 & 6 & 8 \end{bmatrix},$$

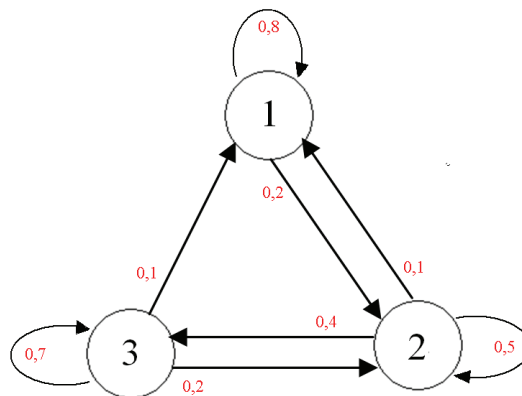
mit $\mathbf{p}^* = \left(\frac{1}{3} \frac{2}{7} \frac{8}{21}\right) \approx (0,33333 \ 0,28572 \ 0,38095)$. Allgemeiner erhält man noch für $n \in \mathbb{N}$:

$$\Pi^n = T \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0,7^n & 0 \\ 0 & 0 & 0,3^n \end{bmatrix} T^{-1} = \Pi^* + \begin{bmatrix} \frac{2}{3} \cdot 0,7^n & -\frac{2}{7} \cdot 0,3^n & \left(-\frac{2}{3} \cdot 0,7^n + \frac{2}{7} \cdot 0,3^n\right) \\ -\frac{1}{3} \cdot 0,7^n & \frac{5}{7} \cdot 0,3^n & \left(\frac{1}{3} \cdot 0,7^n - \frac{5}{7} \cdot 0,3^n\right) \\ -\frac{1}{3} \cdot 0,7^n & -\frac{2}{7} \cdot 0,3^n & \left(\frac{1}{3} \cdot 0,7^n + \frac{2}{7} \cdot 0,3^n\right) \end{bmatrix};$$

exemplarisch:

$$\Pi^2 = \begin{bmatrix} 0,66 & 0,26 & 0,08 \\ 0,17 & 0,35 & 0,48 \\ 0,17 & 0,26 & 0,57 \end{bmatrix}, \quad \Pi^{10} \approx \begin{bmatrix} 0,3520 & 0,2856 & 0,3620 \\ 0,3259 & 0,2857 & 0,3904 \\ 0,3240 & 0,2857 & 0,3903 \end{bmatrix}.$$

Häufig wird zur Veranschaulichung einer homogenen Markoff-Kette vom Typ I ein Graph verwendet, bei dem durch gerichtete Kanten die möglichen Übergänge zwischen den Zuständen angezeigt werden. Für das obige Beispiel wäre dies etwa folgender Graph (Zahlen in rot = jeweilige Übergangswahrscheinlichkeiten):



Bemerkung: Die Voraussetzungen des Konvergenzsatzes können nicht ohne weiteres abgeschwächt werden, wie die drei folgenden Beispiele zeigen:

1. Für $\Pi = \begin{bmatrix} 0,8 & 0,2 & 0 \\ 0,5 & 0,5 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ ist $\lambda_1 = \lambda_2 = 1$ ein doppelter Eigenwert von Π mit $\lambda_3 = 0,3$ und

$$\Pi^* = \lim_{n \rightarrow \infty} \Pi^n = \frac{1}{7} \begin{bmatrix} 5 & 2 & 0 \\ 5 & 2 & 0 \\ 0 & 0 & 7 \end{bmatrix}.$$

IV. Stochastische Prozesse

Ferner existieren wesentlich *verschiedene* stationäre Verteilungen, z.B. die linear unabhängigen Lösungen

$$\mathbf{p}_1^* = \frac{1}{7}(5 \ 2 \ 0) \text{ und } \mathbf{p}_2^* = (0 \ 0 \ 1).$$

2. Für $\Pi = \begin{bmatrix} 0,8 & 0,2 & 0 \\ 0 & 0,8 & 0,2 \\ 0 & 0 & 1 \end{bmatrix}$ ist $\lambda_2 = \lambda_3 = 0,8$ ein doppelter Eigenwert von Π mit $\lambda_1 = 1$, aber der

zu $\lambda_2 = \lambda_3 = 0,8$ gehörende Eigenraum ist eindimensional, so dass zwar

$\Pi^* = \lim_{n \rightarrow \infty} \Pi^n = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$ existiert, aber eine Darstellung für die Potenz Π^n der Übergangsmatrix über eine Formel

$$\Pi^n = T \Delta^n T^{-1}, \quad n \in \mathbb{N}$$

nicht möglich ist.

Man kann hier allerdings meist numerisch doch eine Lösung finden durch eine Implementierung kleiner Störungen, die man nach Berechnung der Eigenvektoren wieder rückgängig machen kann. Wir betrachten dazu im obigen Beispiel die gestörte Übergangsmatrix

$$\Pi_\varepsilon = \begin{bmatrix} 0,8 & 0,2 & 0 \\ 0 & 0,8 - \varepsilon & 0,2 + \varepsilon \\ 0 & 0 & 1 \end{bmatrix} \text{ für genügend kleine } \varepsilon > 0$$

mit den drei paarweise verschiedenen Eigenwerten $\lambda_1 = 1$, $\lambda_2 = \frac{4}{5}$ und $\lambda_3 = \frac{4}{5} - \varepsilon$. Hieraus ergibt sich z.B.

$$T_\varepsilon = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -5\varepsilon \\ 1 & 0 & 0 \end{bmatrix}, \quad \Delta_\varepsilon = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 4/5 & 0 \\ 0 & 0 & 4/5 - \varepsilon \end{bmatrix}, \quad T_\varepsilon^{-1} = \frac{1}{5\varepsilon} \begin{bmatrix} 0 & 0 & 5\varepsilon \\ 5\varepsilon & 1 & -(1+5\varepsilon) \\ 0 & -1 & 1 \end{bmatrix}$$

mit

$$\Pi_\varepsilon^n = T_\varepsilon \Delta_\varepsilon^n T_\varepsilon^{-1} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0,8^n & \frac{0,8^n - (0,8 - \varepsilon)^n}{5\varepsilon} & - \left\{ 0,8^n + \frac{0,8^n - (0,8 - \varepsilon)^n}{5\varepsilon} \right\} \\ 0 & (0,8 - \varepsilon)^n & -(0,8 - \varepsilon)^n \\ 0 & 0 & 0 \end{bmatrix} \text{ und}$$

IV. Stochastische Prozesse

$$\Pi^n = \lim_{\varepsilon \downarrow 0} \Pi_\varepsilon^n = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} + \begin{bmatrix} 0,8^n & \frac{n \cdot 0,8^{n-1}}{5} & -\left\{0,8^n + \frac{n \cdot 0,8^{n-1}}{5}\right\} \\ 0 & 0,8^n & -0,8^n \\ 0 & 0 & 0 \end{bmatrix}$$

für alle $n \in \mathbb{N}$.

Wenn der Zustandsraum nicht endlich ist, muss keine stationäre Verteilung existieren. Beispiel:

$$\mathcal{S} = \mathbb{Z}^+ \text{ mit } \Pi = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & \dots \\ 0 & 0 & 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & 0 & \ddots & \\ \vdots & \vdots & \vdots & \vdots & & \end{bmatrix}$$

Eine stationäre Verteilung wäre normierte Lösung der Gleichung

$$\mathbf{p}^* = (\pi_1 \ \pi_2 \ \pi_3 \ \dots) = \mathbf{p}^* \Pi = (0 \ \pi_1 \ \pi_2 \ \pi_3 \ \dots)$$

und damit $\mathbf{p}^* = \mathbf{0}$ (Nullvektor), was nicht möglich ist.

Als Vorbereitung auf die Markoff-Prozesse, die im nächsten Abschnitt behandelt werden, betrachten wir nun noch die Sprungzeitenfolge $\{S_n\}_{n \in \mathbb{Z}^+}$ der Markoff-Kette, die rekursiv definiert ist durch

$$S_0 := 0, \quad S_{n+1} := \inf \{k > S_n \mid X_k \neq X_{k-1}\} \text{ für alle } n \in \mathbb{Z}^+.$$

Dabei setzen wir voraus, dass $p_{ii} < 1$ für alle Zustände $i \in \mathcal{S}$ gilt; unter dieser Bedingung ist die Folge der Sprungzeiten f.s. wohldefiniert und messbar. Betrachtet man die ursprüngliche Markoff-Kette nur an den Sprungzeiten, so erhält man die Folge $\{\tilde{X}_n\}_{n \in \mathbb{Z}^+}$, die formal definiert ist durch

$$\tilde{X}_n := X_{S_n} \text{ für alle } n \in \mathbb{Z}^+.$$

Man kann zeigen, dass mit $\{X_n\}_{n \in \mathbb{Z}^+}$ auch $\{\tilde{X}_n\}_{n \in \mathbb{Z}^+}$ eine homogene Markoff-Kette bildet, mit der Übergangsmatrix $\tilde{\Pi} := [\tilde{p}_{ij}]_{i,j \in \mathcal{S}}$, die durch

$$\tilde{p}_{ij} := \begin{cases} \frac{p_{ij}}{1 - p_{ii}} & \text{falls } i \neq j \\ 0 & \text{falls } i = j \end{cases} \text{ für } i, j \in \mathcal{S}.$$

gegeben ist. Diese Markoff-Kette, die man die in $\{X_n\}_{n \in \mathbb{Z}^+}$ *eingebettete Markoff-Kette* nennt, unterscheidet sich von der originären Markoff-Kette nur dadurch, dass jeder besuchte Zustand im

IV. Stochastische Prozesse

nächsten Schritt sogleich wieder verlassen wird; man erhält sie also aus dieser, indem man die mehrfach in unmittelbarer Abfolge besuchten Zustände „streicht“.

In direktem Zusammenhang mit den Sprungzeiten der Markoff-Kette stehen die so genannten *Verweildauern* $\{V_n\}_{n \in \mathbb{Z}^+}$, die formal gegeben sind durch

$$V_n := S_{n+1} - S_n \text{ für alle } n \in \mathbb{Z}^+.$$

Ihre stochastische Struktur ist gegeben durch die Beziehung

$$P^{\{V_n\}_{n \in \mathbb{Z}^+}} \left(\cdot \mid \{ \tilde{X}_k = i_k \mid k \in \mathbb{Z}^+ \} \right) = \bigotimes_{n \in \mathbb{Z}^+} \mathcal{G}_{\mathbb{N}}(1 - p_{i_n i_n}),$$

d.h. die Verweildauern sind bedingt stochastisch unabhängig und über \mathbb{N} geometrisch verteilt, gegeben die Abfolge der sukzessiv besuchten *verschiedenen* Zustände. Anschaulich entwickelt sich eine homogene Markoff-Kette also folgendermaßen:

- Die Markoff-Kette startet in einem nach der Anfangsverteilung $\mathbf{p}(0)$ zufällig bestimmten Anfangszustand $\tilde{X}_0 = X_0 = i_0$.
- Sie verweilt dort eine zufällig lange Zeit V_0 , mit einer über \mathbb{N} gegebenen geometrischen Verteilung $\mathcal{G}_{\mathbb{N}}(1 - p_{i_0 i_0})$.
- Zum zufälligen Zeitpunkt $S_1 = V_0$ springt sie in den Zustand $\tilde{X}_1 = i_1$. Dieser Übergang erfolgt (bedingt) unabhängig von der im Zustand i_0 verbrachten Zeit, allein gesteuert durch die Übergangsmatrix $\tilde{\Pi}$.
- Sie verweilt nun eine zufällig lange Zeit V_1 im Zustand i_1 , mit einer über \mathbb{N} gegebenen geometrischen Verteilung $\mathcal{G}_{\mathbb{N}}(1 - p_{i_1 i_1})$. Diese Verweildauer ist bedingt stochastisch unabhängig von allen vorher verbrachten Verweildauern und angenommenen Zuständen; sie hängt über den Parameter $1 - p_{i_1 i_1}$ ausschließlich vom gerade erreichten Zustand ab.
- Zum zufälligen Zeitpunkt $S_2 = S_1 + V_1$ springt sie in den Zustand $\tilde{X}_2 = i_2$. Dieser Übergang erfolgt (bedingt) unabhängig von allen in den vorherigen Zuständen verbrachten Verweilzeiten, allein gesteuert durch die Übergangsmatrix $\tilde{\Pi}$.
- Diese Entwicklung - Verbleiben im gerade erreichten Zustand i mit einer nur von diesem Zustand abhängigen geometrischen Verteilung mit Parameter $1 - p_{ii}$, (bedingt) unabhängig von den vorher verbrachten Verweilzeiten und angenommenen Zuständen, mit nachfolgendem Übergang in einen neuen Zustand ausschließlich gesteuert durch $\tilde{\Pi}$ - vollzieht sich in gleicher Weise zu allen weiteren Zeitpunkten.

Die eingebettete Markoff-Kette zusammen mit der Sprungzeitenfolge bzw. äquivalent mit den Verweildauern beschreibt die originäre Markoff-Kette ebenfalls eindeutig. Dabei kann man auch noch die Abfolge der in einem festen Zustand $i \in \mathcal{S}$ sukzessiv verbrachten Verweildauern $\{V_m(i)\}_{m \in \mathbb{N}}$ betrachten, die nach den obigen Ausführungen sogar stochastisch unabhängig und jeweils wieder $\mathcal{G}_{\mathbb{N}}(1 - p_{ii})$ -verteilt sind.

IV. Stochastische Prozesse

Streng genommen muss man für eine exakte Definition der Verweildauern *gestoppte* Markoff-Ketten betrachten, mit den Stoppzeiten

$$J_0(i) := \inf \{k \in \mathbb{Z}^+ \mid X_k = i\},$$

$$J_1(i) := \inf \{k > J_0(i) \mid X_{k-1} \neq i, X_k = i\}, \quad J_{m+1}(i) := \inf \{k > J_m(i) \mid X_{k-1} \neq i, X_k = i\}$$

für $m \in \mathbb{N}$. $J_m(i)$ ist dann der Zeitpunkt, zu dem die Markoff-Kette zum m -ten Mal den Zustand i ohne vorherige Wiederholung erreicht. Man beachte dabei, dass die Folgen $\{X_{J_m(i)+k}\}_{k \in \mathbb{Z}^+}$, also die „Fortsetzungen“ der Markoff-Kette nach dem m -ten Erreichen des Zustands i ohne Wiederholung, selbst wieder homogene Markoff-Ketten sind mit der gleichen Übergangsmatrix Π .

Die Verweildauern im Zustand i sind dann formal korrekt gegeben durch

$$V_m(i) := \inf \{k > 0 \mid X_{J_m(i)+k} \neq i\} \quad \text{für } m \in \mathbb{Z}^+.$$

Für das obige Beispiel ergibt sich etwa:

$$\tilde{\Pi} = \begin{bmatrix} 0 & 1 & 0 \\ 1/5 & 0 & 4/5 \\ 1/3 & 2/3 & 0 \end{bmatrix}$$

mit den Verweildauerverteilungen

$$P^{V_m(1)} = \mathcal{G}_{\mathbb{N}}(0,2) \quad P^{V_m(2)} = \mathcal{G}_{\mathbb{N}}(0,5) \quad P^{V_m(3)} = \mathcal{G}_{\mathbb{N}}(0,3) \quad \text{für alle } m \in \mathbb{N}.$$

Die Eigenwerte der Matrix $\tilde{\Pi}$ sind $\lambda_1 = 1$, $\lambda_{2,3} = -\frac{1}{2} \pm \frac{i}{30} \sqrt{15}$ (zwei konjugiert-komplexe Eigenwerte). Die eingebettete Markoff-Kette $\{\tilde{X}_n\}_{n \in \mathbb{Z}^+}$ besitzt demnach auf Grund des Konvergenzsatzes eine eindeutige stationäre (Grenz-)Verteilung $\tilde{\mathbf{p}}^*$, die hier gegeben ist durch

$$\tilde{\mathbf{p}}^* = \left(\frac{7}{34} \quad \frac{15}{34} \quad \frac{6}{17} \right) \approx (0,2059 \quad 0,4412 \quad 0,3529).$$

Es ist aber zu beachten, dass eine eingebettete Markoff-Kette auch unter den Voraussetzungen des Konvergenzsatzes für die originäre Markoff-Kette nicht unbedingt ein ähnliches Konvergenzverhalten besitzt.

IV. Stochastische Prozesse

Beispiel: $\Pi = \begin{bmatrix} 0,8 & 0,2 \\ 0,5 & 0,5 \end{bmatrix}$ mit den Eigenwerten $\lambda_1 = 1$, $\lambda_2 = 0,3$ und

$$\Pi^n = \frac{1}{7} \begin{bmatrix} 5 & 2 \\ 5 & 2 \end{bmatrix} + \frac{1}{7} \begin{bmatrix} 2 \cdot 0,3^n & -2 \cdot 0,3^n \\ -5 \cdot 0,3^n & 5 \cdot 0,3^n \end{bmatrix}, \quad \mathbf{p}^* = \begin{pmatrix} 5 & 2 \\ 7 & 7 \end{pmatrix}.$$

Dagegen ist $\tilde{\Pi} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$ mit den Eigenwerten $\lambda_{1,2} = \pm 1$ und

$$\tilde{\Pi}^n = \begin{cases} \tilde{\Pi} & \text{falls } n \text{ ungerade} \\ \mathbf{I} & \text{falls } n \text{ gerade,} \end{cases}$$

d.h. $\{\tilde{\Pi}^n\}_{n \in \mathbb{N}}$ konvergiert *nicht*.

Es existiert aber trotzdem eine eindeutig bestimmte stationäre Verteilung, nämlich

$$\tilde{\mathbf{p}}^* = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}.$$

Man kann eine eingebettete Markoff-Kette auch noch dann definieren, wenn $p_{ii} = 1$ gilt für gewisse $i \in \mathcal{S}$. Solche Zustände heißen *absorbierend*, denn wenn sie einmal erreicht wurden, können sie fast sicher nicht mehr verlassen werden, d.h. die Verweildauer in diesem Zustand ist dann unendlich groß. In diesem Fall ist auch $\tilde{p}_{ii} = 1$ zu setzen.

IV.2 Markoff-Prozesse

Bei den Markoff-Prozessen unterscheidet man im Wesentlichen wieder zwei Typen: zum einen diejenigen mit *abzählbarem* Zustandsraum, welche sich ähnlich beschreiben lassen wie die Markoff-Ketten vom Typ I im vorangehenden Abschnitt (z.B. der Poisson-Prozess und seine Varianten als berühmtester Vertreter), zum anderen Markoff-Prozesse mit *überabzählbarem* Zustandsraum (Typ II, z.B. der Wiener-Prozess [auch Brown'sche Bewegung genannt] und seine Varianten als wichtigster Vertreter). Eine rigorose Behandlung der letztgenannten Prozesse setzt allerdings einen erheblichen maßtheoretischen Aufwand voraus, den wir hier zu Gunsten einer mehr plakativen Behandlung zurückstellen.

Definition 51. Unter einem *Markoff-Prozess vom Typ I* versteht man eine Familie $\{X_t\}_{t \geq 0}$ reellwertiger Zufallsvariablen mit Werten in einer höchstens abzählbaren Menge \mathcal{S} , die durch folgende fundamentale Eigenschaft charakterisiert ist:

- Für jede streng aufsteigende Folge $\{t_n\}_{n \in \mathbb{Z}^+} \subset \mathbb{R}^+$ von „Zeitpunkten“ ist die Folge $\{X_{t_n}\}_{n \in \mathbb{Z}^+}$ eine Markoff-Kette vom Typ I.

Man nennt einen solchen Markoff-Prozess (zeitlich) *homogen*, wenn zusätzlich gilt:

- Für alle Folgen $\{t_n\}_{n \in \mathbb{Z}^+} \subset \mathbb{R}^+$ mit $t_{n+1} - t_n = c > 0$, $n \in \mathbb{Z}^+$ ist die Markoff-Kette $\{X_{t_n}\}_{n \in \mathbb{Z}^+}$ homogen, mit identischen Übergangsmatrizen für gleiche $c > 0$.

IV. Stochastische Prozesse

Ähnlich wie bei den Markoff-Ketten stellt sich auch hier die Frage, ob das stochastische Verhalten eines Markoff-Prozesses vom Typ I bereits durch die Angabe geeigneter Übergangswahrscheinlichkeiten charakterisiert ist. Bei zeitlich *homogenen* Prozessen benötigt man dazu eine geeignete Familie $\{\Pi^t\}_{t \geq 0}$ stochastischer Matrizen mit Elementen $p_{ij}^{(t)}$, $i, j \in \mathcal{S}$, $t \geq 0$, die das Übergangsverhalten des Prozesses vermöge

$$p_{ij}^{(t-s)} = P(X_t = j | X_s = i), \quad i, j \in \mathcal{S}, \quad 0 \leq s \leq t$$

beschreiben, wobei wie üblich noch $\Pi^0 = \mathbf{I}$ (Einheitsmatrix) zu setzen ist.

Dies ist konsistent zur Definition eines homogenen Markoff-Prozesses, weil für ganzzahlige Werte von t die Übergangsmatrix Π^t genau mit der t -ten Potenz der Einschritt-Übergangsmatrix der Markoff-Kette $\{X_t\}_{t \in \mathbb{Z}^+}$ zusammenfällt. Es bleibt also lediglich zu klären, wie hier „gebrochene“ Potenzen der Einschritt-Übergangsmatrix $\Pi = \Pi^1$ zu definieren sind.

Definition 52: Es sei $\mathbf{Q} = [q_{ij}]_{i, j \in \mathcal{S}}$ eine Matrix mit der Eigenschaft

$$e^{\mathbf{Q}} := \sum_{k=0}^{\infty} \frac{1}{k!} \mathbf{Q}^k = \mathbf{I} + \mathbf{Q} + \frac{1}{2} \mathbf{Q}^2 + \dots = \Pi.$$

Eine solche Matrix heißt *Fundamental-* bzw. *Intensitätsmatrix*. Setze dann

$$\Pi^t := e^{t\mathbf{Q}} := \sum_{k=0}^{\infty} \frac{t^k}{k!} \mathbf{Q}^k = \mathbf{I} + t\mathbf{Q} + \frac{t^2}{2} \mathbf{Q}^2 + \dots \quad \text{für } t \geq 0.$$

Bemerkung: Die notwendige Konvergenz der Matrix-Reihe orientiert sich dabei an einer geeigneten Norm; bei endlichen Zustandsräumen sind die entsprechenden Vektorräume alle endlich-dimensional und daher nach einem Satz der Funktionalanalysis alle Normen dort äquivalent.

Die Fundamentalmatrix wird auch gelegentlich mit $\mathbf{Q} = \ln(\Pi)$ bezeichnet mit der zum reellen Fall analogen formalen Reihendarstellung

$$\mathbf{Q} = \ln(\mathbf{I} - (\mathbf{I} - \Pi)) = -\sum_{k=1}^{\infty} \frac{1}{k} (\mathbf{I} - \Pi)^k.$$

Die für homogene Markoff-Ketten gültigen Chapman-Kolmogoroff-Gleichungen erweitern sich hier natürlicherweise zu der Gleichung

$$\Pi^{s+t} = \Pi^s \Pi^t, \quad s, t \geq 0$$

bzw. in expliziter Form:

$$p_{ij}^{(s+t)} = \sum_{\ell \in \mathcal{S}} p_{i\ell}^{(s)} p_{\ell j}^{(t)}, \quad s, t \geq 0.$$

Für die Verteilung von X_t gilt hier wie im zeitdiskreten Fall:

$$\mathbf{p}(t) = \mathbf{p}(0) \cdot \Pi^t, \quad t \geq 0.$$

IV. Stochastische Prozesse

Leider besitzen nicht alle Übergangsmatrizen Π eine Fundamentalmatrix; dies gilt z.B. schon für $\Pi = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}$ mit $0 \leq p < \frac{1}{2}$. Es müsste sonst nämlich die Gleichung $\Pi = (\Pi^{1/2})^2$ und damit

$$\left\{ \det(\Pi^{1/2}) \right\}^2 = \det(\Pi) = p^2 - (1-p)^2 = 2p - 1 < 0$$

erfüllt sein, was nicht geht, da $\Pi^{1/2}$ als stochastische Matrix nur reelle Zahlen enthalten darf und damit auch $\det(\Pi^{1/2}) \in \mathbb{R}$ gelten muss.

Für $\frac{1}{2} \leq p \leq 1$ existiert dagegen immer eine Fundamentalmatrix, nämlich

$$Q = \frac{1}{2} \ln(2p-1) \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}.$$

Setzt man etwa $A := \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$, so erhält man $A^n = \frac{1}{2} \begin{bmatrix} 2^n & -2^n \\ -2^n & 2^n \end{bmatrix}$ für alle $n \in \mathbb{N}$ (Nachweis z.B. mit Induktion) und damit für beliebige $s \geq 0$

$$e^{sA} = \mathbf{I} + \sum_{k=1}^{\infty} \frac{s^k}{k!} A^k = \mathbf{I} + \frac{1}{2} \begin{bmatrix} e^{2s} - 1 & -e^{2s} + 1 \\ -e^{2s} + 1 & e^{2s} - 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 + e^{2s} & 1 - e^{2s} \\ 1 - e^{2s} & 1 + e^{2s} \end{bmatrix}.$$

Mit der Wahl $s = \frac{1}{2} \ln(2p-1)$ ergibt sich das gewünschte Ergebnis.

Sind die Übergangsmatrizen $\{\Pi^t\}_{t \geq 0}$ eines homogenen Markoff-Prozesses vom Typ I explizit bekannt, so erhält man die Fundamentalmatrix Q als Matrix-Ableitung:

$$Q = \lim_{t \downarrow 0} \frac{1}{t} (\Pi^t - \mathbf{I})$$

bzw. elementweise

$$q_{ij} = \begin{cases} \lim_{t \downarrow 0} \frac{1}{t} p_{ij}^{(t)} = \lim_{t \downarrow 0} \frac{1}{t} P(X_t = j | X_0 = i) & \text{für } i \neq j \\ \lim_{t \downarrow 0} \frac{1}{t} (p_{ii}^{(t)} - 1) = -\lim_{t \downarrow 0} \frac{1}{t} P(X_t \neq i | X_0 = i) & \text{für } i = j, \end{cases} \quad i, j \in \mathcal{S}.$$

Insbesondere gilt für jede Fundamentalmatrix Q stets

$$q_{ii} \leq 0, \quad q_{ij} \geq 0 \text{ für } i \neq j \in \mathcal{S} \quad \text{und} \quad \sum_{j \in \mathcal{S}} q_{ij} = 0 \quad \text{für alle } i \in \mathcal{S}; \quad (*)$$

letzteres folgt aus der Beobachtung, dass alle Zeilensummen von $\Pi^t - \mathbf{I}$ für jedes $t \geq 0$ den Wert Null ergeben müssen.

IV. Stochastische Prozesse

Ein weiteres Gleichungssystem (Kolmogoroff'sche Rückwärtsgleichungen) erhält man aus der Beziehung

$$\frac{d}{dt} \Pi^t = \frac{d}{dt} e^{t\mathbf{Q}} = \mathbf{Q}e^{t\mathbf{Q}} = \mathbf{Q} \cdot \Pi^t \text{ für } t \geq 0$$

bzw. explizit, mit $\lambda_i := -q_{ii}$ für $i \in \mathcal{S}$:

$$p_{ij}'(t) = \sum_{k \in \mathcal{S}} q_{ik} p_{kj}(t) = \sum_{\substack{k \in \mathcal{S} \\ k \neq i}} q_{ik} p_{kj}(t) - \lambda_i p_{ij}(t) \text{ für } i, j \in \mathcal{S}.$$

Allerdings existiert nicht für jede beliebige Matrix \mathbf{Q} mit diesen beiden Eigenschaften stets ein passender homogener Markoff-Prozess. Dazu müssen die Diagonalelemente von \mathbf{Q} in der Regel einer bestimmten Beschränktheitseigenschaft genügen.

Satz 55 (Struktursatz für homogene Markoff-Prozesse vom Typ I): Es sei \mathbf{Q} eine Matrix mit der Eigenschaft (*) derart, dass die Bedingungen

$$0 < \nu := \inf_{i \in \mathcal{S}} \{\lambda_i\}, \quad \mu := \sup_{i \in \mathcal{S}} \{\lambda_i\} < \infty \text{ mit } \lambda_i = -q_{ii} \text{ für } i \in \mathcal{S}$$

gelten. Dann existiert ein homogener Markoff-Prozess $\{X_t\}_{t \geq 0}$ vom Typ I auf einem geeigneten Wahrscheinlichkeitsraum derart, dass seine Übergangsmatrizen durch $\Pi^t = e^{t\mathbf{Q}}$ für alle $t \geq 0$ gegeben sind. Der Markoff-Prozess kann kanonisch wie folgt dargestellt werden (ähnlich den homogenen Markoff-Ketten über die eingebettete Markoff-Kette und die Sprung-/Verweilzeiten):

Es sei $\{Y_n\}_{n \in \mathbb{Z}^+}$ eine homogene Markoff-Kette mit demselben Zustandsraum \mathcal{S} und der Übergangsmatrix $\tilde{\Pi}$ mit den Übergangswahrscheinlichkeiten

$$P(Y_{n+1} = j | Y_n = i) = \begin{cases} \frac{q_{ij}}{\lambda_i} & \text{für } j \neq i \\ 0 & \text{für } j = i, \end{cases} \quad i, j \in \mathcal{S}$$

und der Anfangsverteilung $P^{X_0} = P^{X_0}$. [Hier sieht man, warum die Bedingung $q_{ij} \geq 0$ für $i \neq j \in \mathcal{S}$ wesentlich ist.] Dann erfolgt die zeitliche Entwicklung des Markoff-Prozesses in folgender Weise, wobei wieder $S_0 := 0$ zusetzen ist:

- Der Markoff-Prozess startet in einem nach der Anfangsverteilung $\mathbf{p}(0)$ zufällig bestimmten Anfangszustand $X_0 = Y_0 = i_0$.
- Er verweilt dort eine zufällig lange Zeit V_0 , mit einer $\mathcal{E}(\lambda_{i_0})$ -Exponentialverteilung.
- Zum zufälligen Zeitpunkt $S_1 = V_0$ springt er in den Zustand $Y_1 = i_1$. Dieser Übergang erfolgt (bedingt) unabhängig von der im Zustand i_0 verbrachten Zeit, allein gesteuert durch die Übergangsmatrix $\tilde{\Pi}$.
- Er verweilt nun eine zufällig lange Zeit V_2 im Zustand i_1 , mit einer $\mathcal{E}(\lambda_{i_1})$ -Exponentialverteilung. Diese Verweildauer ist bedingt stochastisch unabhängig von allen vorher verbrachten Verweildauern und angenommenen Zuständen; sie hängt über den Parameter λ_{i_1} ausschließlich vom gerade erreichten Zustand ab.

IV. Stochastische Prozesse

- Zum zufälligen Zeitpunkt $S_2 = S_1 + V_1$ springt er in den Zustand $Y_2 = i_2$. Dieser Übergang erfolgt (bedingt) unabhängig von allen in den vorherigen Zuständen verbrachten Verweilzeiten, allein gesteuert durch die Übergangsmatrix $\tilde{\Pi}$.
- Diese Entwicklung - Verbleiben im gerade erreichten Zustand i mit einer nur von diesem Zustand abhängigen Exponentialverteilung mit Parameter λ_i , (bedingt) unabhängig von den vorher verbrachten Verweilzeiten und angenommenen Zuständen, mit nachfolgendem Übergang in einen neuen Zustand ausschließlich gesteuert durch $\tilde{\Pi}$ - vollzieht sich in gleicher Weise zu allen weiteren Zeitpunkten.

Die Tatsache, dass die jeweiligen Verweildauern in den sukzessiv besuchten Zuständen exponentialverteilt sind, hängt dabei damit zusammen, dass die geometrischen Verteilungen, die die Verweildauern bei homogenen Markoff-Ketten steuern, bei stetiger Zeit in Exponentialverteilungen übergehen.

Die hier verwendete Markoff-Kette $\{Y_n\}_{n \in \mathbb{Z}^+}$ heißt analog wie im zeitdiskreten Fall die in $\{X_t\}_{t \in \mathbb{Z}^+}$ eingebettete Markoff-Kette.

Bemerkung: Auf Grund der obigen Konstruktion ist ersichtlich, dass an die Parameter λ_i Bedingungen, wie der Struktursatz sie vorgibt, gemacht werden müssen. Ist beispielsweise $\mathcal{S} = \mathbb{Z}^+$ und wählt man

$$q_{ij} = \begin{cases} \lambda_j & \text{für } j = i + 1 \\ -\lambda_j & \text{für } j = i \\ 0 & \text{sonst,} \end{cases} \quad i \in \mathcal{S},$$

sowie als Anfangsverteilung ε_0 (d.h. fast sicherer Start im Nullpunkt), so gilt $E(S_m) = \sum_{k=0}^{m-1} \frac{1}{\lambda_k}$ für

alle $m \in \mathbb{N}$ wegen $S_m = \sum_{k=0}^{m-1} V_k$. Für den Fall, dass $\sum_{k=0}^{\infty} \frac{1}{\lambda_k} < \infty$ ist (und damit $\mu = \sup_{i \in \mathcal{S}} \{\lambda_i\} = \infty$ gilt), „explodiert“ der Prozess also fast sicher bereits nach endlicher Zeit, so dass nicht für alle Zeitpunkte $t > 0$ in konsistenter Weise ein Wert für X_t bestimmt werden kann.

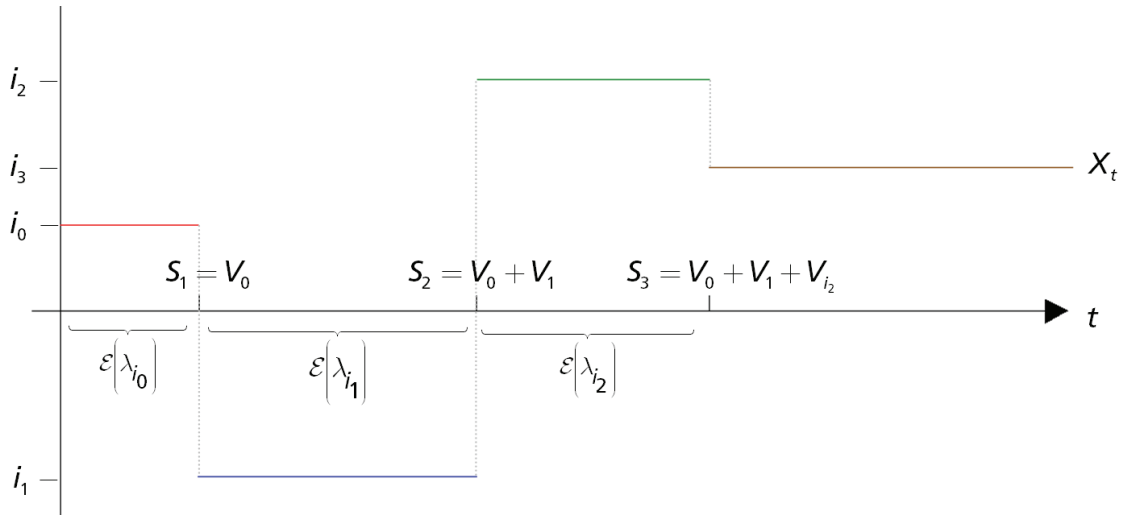
Die andere Bedingung $\nu = \inf_{i \in \mathcal{S}} \{\lambda_i\} > 0$ garantiert, dass alle λ_i positiv und damit die Übergangswahrscheinlichkeiten der eingebetteten Markoff-Kette $\{Y_n\}_{n \in \mathbb{Z}^+}$ wohldefiniert sind, also insbesondere keine absorbierenden Zustände auftreten.

Ist dagegen ein $\lambda_i = 0$, so enthält \mathbf{Q} eine Nullzeile; in diesem Fall wäre für die eingebettete Markoff-Kette

$$P(Y_{n+1} = i | Y_n = i) = 1$$

zu setzen. Der Zustand $i \in \mathcal{S}$ ist also für diese Kette absorbierend, mit unendlicher Verweildauer. Der Struktursatz gilt dann entsprechend, d.h. bei Erreichen irgendeines absorbierenden Zustands ist die zugehörige Verweildauer des Markoff-Prozesses ebenfalls unendlich.

IV. Stochastische Prozesse



zeitliche Entwicklung eines homogenen Markoff-Prozesses

Satz 56 (Konvergenzsatz für homogene Markoff-Prozesse mit endlichem Zustandsraum):

Unter der Bedingung $0 < \nu := \min_{i \in \mathcal{S}} \{\lambda_i\}$, $\mu := \max_{i \in \mathcal{S}} \{\lambda_i\} < \infty$ gilt:

Alle Eigenwerte von $\Pi := \Pi^1$ sind nicht-negativ; sind alle Eigenwerte von Π paarweise voneinander verschieden, so gilt

$$\Pi^* := \lim_{t \rightarrow \infty} \Pi^t = \begin{pmatrix} \pi_1 & \pi_2 & \cdots & \pi_r \\ \pi_1 & \pi_2 & \cdots & \pi_r \\ \vdots & \vdots & \vdots & \vdots \\ \pi_1 & \pi_2 & \cdots & \pi_r \end{pmatrix}, \text{ wobei } r = \#(\mathcal{S}),$$

mit dem eindeutig bestimmten normierten Linkseigenvektor (*stationäre Verteilung*) $\mathbf{p}^* = (\pi_1 \ \pi_2 \ \cdots \ \pi_r)$ von Π .

Beweis: Ist λ ein Eigenwert von Π , so ist λ^t ein Eigenwert von Π^t für alle $t > 0$. Die zugehörigen Eigenräume sind identisch. Π^t kann alternativ berechnet werden über die Formel

$$\Pi^t = T e^{t\Delta} T^{-1}, \quad t \geq 0,$$

wobei Δ die Matrix der Eigenwerte von \mathbf{Q} und T eine beliebige Matrix zugehöriger Rechtseigenvektoren ist. Vgl. auch den Beweis zu Satz 51. ■

Bemerkungen:

- Ist ν ein Eigenwert von \mathbf{Q} , so ist $e^{\nu t}$ ein Eigenwert von Π^t für alle $t > 0$. Die zugehörigen Eigenräume sind identisch.
- Für jede beliebige Anfangsverteilung $\mathbf{p}(0)$ konvergiert der Markoff-Prozess in Verteilung gegen die stationäre Verteilung \mathbf{p}^* . Ist schon $\mathbf{p}(0) = \mathbf{p}^*$, so sind alle Randverteilungen des Prozesses mit \mathbf{p}^* identisch.
- Die stationäre Verteilung \mathbf{p}^* ist zugleich eindeutige Lösung des Gleichungssystems $\mathbf{p} \cdot \mathbf{Q} = \mathbf{0}$.

IV. Stochastische Prozesse

Erklärung: Die stationäre Verteilung erfüllt die Gleichung

$$\mathbf{p} \cdot \Pi^t = \mathbf{p} \text{ für alle } t \geq 0.$$

Differenzieren nach t ergibt:

$$\mathbf{0} = \frac{d}{dt} \mathbf{p} = \frac{d}{dt} \mathbf{p} \cdot \Pi^t = \mathbf{p} \cdot \frac{d}{dt} \Pi^t = \mathbf{p} \cdot \frac{d}{dt} e^{t\mathbf{Q}} = \mathbf{p} \cdot \mathbf{Q} \cdot e^{t\mathbf{Q}}$$

für alle $t \geq 0$. Für $t = 0$ erhält man hieraus $\mathbf{p} \cdot \mathbf{Q} = \mathbf{0}$.

Beispiel: Wir modellieren das Krankheitsverhalten einer Person durch den homogenen Markoff-Prozess $\{X_t\}_{t \geq 0}$ mit Zustandsraum $S = \{0, 1\}$ ($0 = \text{krank}$, $1 = \text{gesund}$) und Fundamentalmatrix

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda \\ \nu & -\nu \end{bmatrix}$$

für geeignete Parameter $\lambda, \nu > 0$.

Dann ist $\varphi(x) = x^2 + (\lambda + \nu)x$ das charakteristische Polynom von \mathbf{Q} mit den Nullstellen (Eigenwerten) 0 und $-(\lambda + \nu)$.

Es folgt

$$T = \begin{bmatrix} 1 & -\lambda \\ 1 & \nu \end{bmatrix}, \quad T^{-1} = \frac{1}{\lambda + \nu} \begin{bmatrix} \nu & \lambda \\ -1 & 1 \end{bmatrix}, \quad e^{t\Delta} = \begin{bmatrix} 1 & 0 \\ 0 & e^{-t(\lambda + \nu)} \end{bmatrix}$$

mit

$$\Pi^t = e^{t\mathbf{Q}} = T e^{t\Delta} T^{-1} = \frac{1}{\lambda + \nu} \begin{bmatrix} \nu + \lambda \cdot e^{-(\lambda + \nu)t} & \lambda \cdot (1 - e^{-(\lambda + \nu)t}) \\ \nu \cdot (1 - e^{-(\lambda + \nu)t}) & \lambda + \nu \cdot e^{-(\lambda + \nu)t} \end{bmatrix}$$

und Grenzwert

$$\lim_{t \rightarrow \infty} \Pi^t = \frac{1}{\lambda + \nu} \begin{bmatrix} \nu & \lambda \\ \nu & \lambda \end{bmatrix}.$$

Die stationäre Verteilung ist also gegeben durch

$$\mathbf{p}^* = \left(\frac{\nu}{\lambda + \nu}, \frac{\lambda}{\lambda + \nu} \right),$$

was zugleich die normierte Lösung von

$$\mathbf{p} \cdot \mathbf{Q} = \mathbf{p} \cdot \begin{bmatrix} -\lambda & \lambda \\ \nu & -\nu \end{bmatrix} = \mathbf{0}$$

ist.

IV. Stochastische Prozesse

Die Verweildauer im Zustand 0 (krank) ist $\mathcal{E}(\lambda)$ -verteilt, die im Zustand 1 (gesund) ist $\mathcal{E}(\nu)$ -verteilt. Die erwartete Krankheitsdauer beträgt also $\frac{1}{\lambda}$, die erwartete Gesundheitsdauer $\frac{1}{\nu}$.

Definition 53 (homogener Poisson-Prozess mit Parameter $\lambda > 0$). Dieser Prozess ist charakterisiert durch die Wahl $\mathcal{S} = \mathbb{Z}^+$ mit

$$q_{ij} = \begin{cases} \lambda & \text{für } j = i + 1 \\ -\lambda & \text{für } j = i \\ 0 & \text{sonst,} \end{cases} \quad i \in \mathcal{S}$$

und Anfangsverteilung ε_0 (fast sicherer Start im Nullpunkt). Die zugehörigen Verweildauern $\{V_n\}_{n \in \mathbb{Z}^+}$ sind damit stochastisch unabhängige, jeweils $\mathcal{E}(\lambda)$ -verteilte Zufallsvariablen. Hieraus ergibt sich auch die folgende namensgebende Eigenschaft:

Lemma 52. Der homogene Poisson-Prozess $\{X_t\}_{t \geq 0}$ mit Parameter $\lambda > 0$ besitzt folgende Eigenschaften:

1. Es gilt

$$P(X_t = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n \in \mathbb{Z}^+, t \geq 0,$$

d.h. jedes X_t , $t \geq 0$, ist $\mathcal{P}(\lambda t)$ -verteilt.

2. Für jede aufsteigende Folge $\{t_n\}_{n \in \mathbb{Z}^+} \subset \mathbb{R}^+$ von Zeitpunkten gilt:

$$X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \dots \text{ sind stochastisch unabhängig mit } P^{X_{t_n} - X_{t_{n-1}}} = \mathcal{P}(\lambda(t_n - t_{n-1}))$$

für alle $n \in \mathbb{N}$, d.h. der homogene Poisson-Prozess besitzt unabhängige, Poisson-verteilte Zuwächse.

Beweis: Zu 1: Für $n \in \mathbb{Z}^+$, $t \geq 0$ gilt:

$$P(X_t = n) = P(S_n \leq t < S_{n+1}) = P(S_n \leq t) - P(S_{n+1} \leq t)$$

mit (partielle Integration)

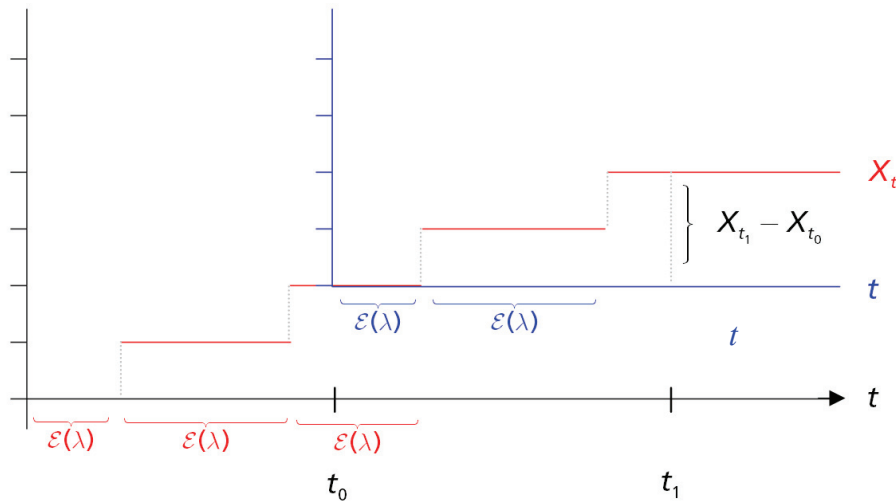
$$P(S_n \leq t) = \int_0^t \underbrace{\frac{\lambda^n}{(n-1)!} x^{n-1}}_{=u'(x)} \underbrace{e^{-\lambda x}}_{=v(x)} dx = u(x)v(x) \Big|_0^t + \int_0^t \frac{\lambda^{n+1}}{n!} x^n e^{-\lambda x} dx = \frac{(\lambda t)^n}{n!} e^{-\lambda t} + P(S_{n+1} \leq t)$$

für $n \in \mathbb{N}$ und

$$P(X_t = 0) = P(t < S_1) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t}.$$

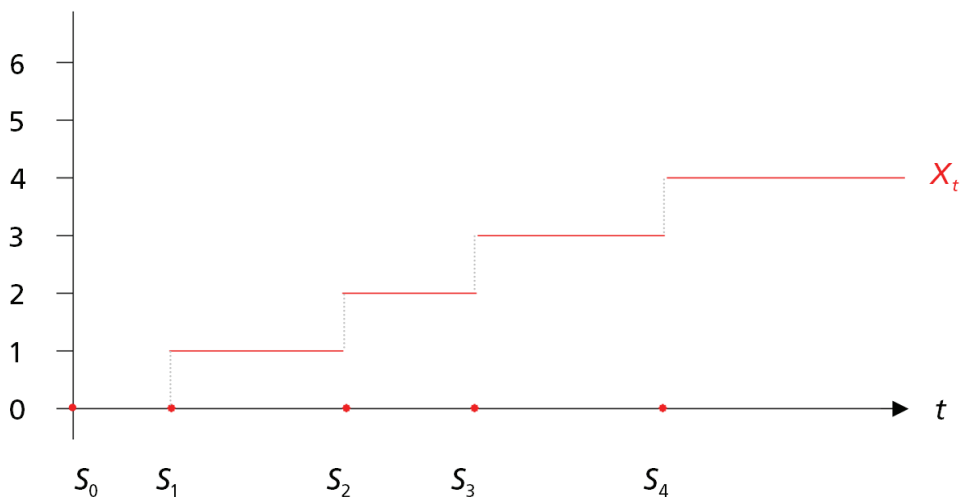
Zu 2:

IV. Stochastische Prozesse



Veranschaulichung der Unabhängigkeit der Zuwächse

Auf Grund der Struktur des Zustandsraums identifiziert man häufig den homogenen Poisson-Prozess auch mit der „zufälligen Menge“ der Sprungzeiten („Ankunftszeiten“).



Lemma 53. Homogene Poisson-Prozesse sind durch vier weitere, in der Praxis oft angewandte Eigenschaften ausgezeichnet:

- *Überlagerungseigenschaft:*

Sind $\{X_t\}_{t \geq 0}$ und $\{Y_t\}_{t \geq 0}$ voneinander unabhängige homogene Poisson-Prozesse mit Parametern $\lambda > 0$ und $\nu > 0$, so ist auch der Summenprozess $\{X_t + Y_t\}_{t \geq 0}$ ein homogener Poisson-Prozess, mit Parameter $\lambda + \nu > 0$. Bezeichnen $\{T_n\}_{n \in \mathbb{Z}^+}$ und $\{S_n\}_{n \in \mathbb{Z}^+}$ die zugehörigen Ankunftszeiten, so lässt sich die Ankunftszeitenfolge $\{U_n\}_{n \in \mathbb{Z}^+}$ des Summenprozesses rekursiv folgendermaßen darstellen:

$$U_0 := 0, U_n := \min \{T_k, S_k > U_{n-1} \mid k \in \mathbb{N}\}, n \in \mathbb{N}.$$

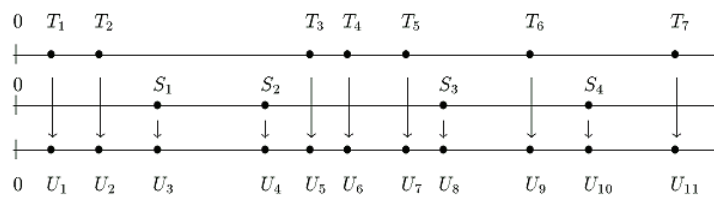
IV. Stochastische Prozesse

- *Aufteilungseigenschaft:*

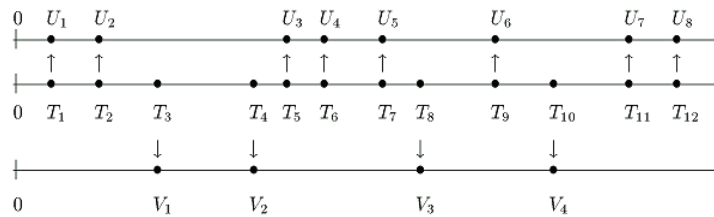
Ist $\{X_t\}_{t \geq 0}$ ein homogener Poisson-Prozess mit Parameter $\lambda > 0$ und $\{I_k\}_{k \in \mathbb{N}}$ eine (auch von $\{X_t\}_{t \geq 0}$) unabhängige Folge $B(1, p)$ -verteilter Zufallsvariablen mit $p = 1 - q \in (0, 1)$, und teilt man die Ankunftszeitenfolge $\{T_n\}_{n \in \mathbb{Z}^+}$ des Poisson-Prozesses zufällig auf gemäß

$$U_0 := V_0 := 0, \quad \begin{aligned} U_n &:= \min \{T_k > U_{n-1} \mid I_k = 1, k \in \mathbb{N}\} \\ V_n &:= \min \{T_k > V_{n-1} \mid I_k = 0, k \in \mathbb{N}\} \end{aligned} \quad \text{für } n \in \mathbb{N},$$

so bilden $\{U_n\}_{n \in \mathbb{Z}^+}$ und $\{V_n\}_{n \in \mathbb{Z}^+}$ die Ankunftszeiten zweier unabhängiger homogener Poisson-Prozesse mit Parametern λp und λq .



Überlagerung homogener Poisson-Prozesse



Aufteilung eines homogenen Poisson-Prozesses

- *Gleichverteilungseigenschaft:*

Ist $\{X_t\}_{t \geq 0}$ ein homogener Poisson-Prozess mit Parameter $\lambda > 0$ und $\{S_n\}_{n \in \mathbb{Z}^+}$ die zugehörige Ankunftszeitenfolge, so entsprechen die in das Intervall $[a, b]$ (mit $0 \leq a < b$) fallenden Ankunftszeiten unter der Bedingung $N = X_b - X_a = n \in \mathbb{N}$ genau einer geordneten Stichprobe vom Umfang n aus über $[a, b]$ stetig gleichverteilten Zufallsvariablen.

Diese Eigenschaft kann insbesondere zur Simulation von Poisson-Prozessen benutzt werden, wenn lediglich ein definierter beschränkter Zeithorizont von Interesse ist. Man beachte, dass dabei N Poisson-verteilt ist mit Parameter $\lambda(b - a)$.

- *Unabhängigkeitseigenschaft:*

Ist $\{X_t\}_{t \geq 0}$ ein homogener Poisson-Prozess mit Parameter $\lambda > 0$ und sind $I_i = (a_i, b_i]$ mit $0 \leq a_i < b_i$ für $i = 1, \dots, n \in \mathbb{N}$ paarweise disjunkte Zeitintervalle, so gilt:

- Die Anzahlen $N_i = X_{b_i} - X_{a_i}$ der Ankunftszeiten des Prozesses in den Intervallen I_i , $i = 1, \dots, n$ sind stochastisch unabhängig.
- Gegeben $N_i = n_i \in \mathbb{Z}^+$ sind die n_i Ankunftszeiten in den Intervallen I_i , $i = 1, \dots, n$ (bedingt) stochastisch unabhängig.

IV. Stochastische Prozesse

Zum Abschluss dieses Abschnitts wollen wir uns noch etwas genauer mit dem Themenkreis *Brown'sche Bewegung / Wiener-Prozess* befassen. Dazu benutzen wir folgendes fundamentale Resultat, das wir in ähnlicher Form schon beim Poisson-Prozess kennen gelernt haben:

Ist $\{X_t\}_{t \geq 0}$ eine beliebige Familie reellwertiger Zufallsvariablen mit unabhängigen Zuwächsen, so ist $\{X_t\}_{t \geq 0}$ ein Markoff-Prozess.

Denn für jede streng aufsteigende Folge $\{t_n\}_{n \in \mathbb{Z}^+} \subset \mathbb{R}^+$ von Zeitpunkten ist die Folge $\{X_{t_0}, X_{t_n} - X_{t_{n-1}}\}_{n \in \mathbb{N}}$ stochastisch unabhängig, also $\{X_{t_n}\}_{n \in \mathbb{Z}^+}$ eine Markoff-Kette nach dem Konstruktionssatz für Markoff-Ketten.

Neben dem homogenen Poisson-Prozess als prominentem Vertreter eines solchen Markoff-Prozesses (vom Typ I) ist die Standard-Brown'sche Bewegung / der Standard-Wiener-Prozess $\{X_t\}_{t \geq 0}$ durch folgende Eigenschaften charakterisiert:

Definition 54: Ein (Markoff-)Prozess $\{X_t\}_{t \geq 0}$ heißt Standard-Brown'sche Bewegung / Standard-Wiener-Prozess, wenn gilt:

1. $\{X_t\}_{t \geq 0}$ hat unabhängige Zuwächse.
2. $\{X_t\}_{t \geq 0}$ hat stationäre Zuwächse, d.h. die Verteilung von $X_t - X_s$ für $0 \leq s < t$ hängt nur von $t - s$ ab.
3. $\{X_t\}_{t \geq 0}$ hat Gauß'sche Zuwächse, d.h. die Verteilung von $X_t - X_s$ für $0 \leq s < t$ ist eine $\mathcal{N}(0, t - s)$ -Verteilung.
4. $\{X_t\}_{t \geq 0}$ hat [fast sicher] stetige Pfade.
5. Die Anfangsverteilung ist ε_0 , d.h. der Prozess startet [fast sicher] in 0.

Bemerkung: Die Bedingungen 3 und 4 in Definition 60 sind *äquivalent*, d.h. eine von beiden kann ersatzlos gestrichen werden!

Lemma 54. Die Standard-Brown'sche Bewegung / der Standard-Wiener-Prozess $\{X_t\}_{t \geq 0}$ besitzt folgende, sich hieraus ergebende Eigenschaften:

- $\{X_t\}_{t \geq 0}$ ist ein homogener Markoff-Prozess vom Typ II.
- Jedes X_t ist $\mathcal{N}(0, t)$ -verteilt für $t \geq 0$ (mit $\mathcal{N}(0, 0) = \varepsilon_0$).
- Jeder Pfad von $\{X_t\}_{t \geq 0}$ ist [fast sicher] nirgends differenzierbar und von unbeschränkter Variation in jedem endlichen Intervall.
- Es gilt $Cov(X_s, X_t) = \min\{s, t\}$ für $s, t \geq 0$.
- Für jede streng aufsteigende endliche Folge $\{t_k\}_{k=1}^n \subset \mathbb{R}^+$ ist $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ multivariat normalverteilt nach $\mathcal{N}(\mathbf{0}, \Sigma)$ mit

$$\sigma_{ij} = \min\{t_i, t_j\}, \quad 1 \leq i, j \leq n \in \mathbb{N}.$$

IV. Stochastische Prozesse

- Mit $\{X_t\}_{t \geq 0}$ sind auch die Prozesse $\left\{ \frac{1}{\sqrt{c}} X_{ct} \right\}_{t \geq 0}$ für alle $c > 0$ und $\{tX_{1/t}\}_{t > 0}$ Standard-Wiener-Prozesse.
- $\{X_t\}_{t \geq 0}$ ist Grenzwert (in Verteilung) einer skalierten homogenen Markoff-Kette $\{Y_n\}_{n \in \mathbb{Z}^+}$ („Irrfahrt“, „Random Walk“) folgender Form:

$$Y_n := \sum_{k=1}^n (2I_k - 1)$$

mit unabhängigen, je $B\left(1, \frac{1}{2}\right)$ -verteilten Zufallsvariablen gemäß

$$X_t \stackrel{d}{=} \lim_{m \rightarrow \infty} \frac{1}{\sqrt{m}} Y_{\lfloor mt \rfloor} \text{ für alle } t \geq 0.$$

Definition 55. Ein(e) allgemeine Brown'sche Bewegung / allgemeiner Wiener-Prozess $\{B_t\}_{t \geq 0}$ entsteht aus dem Standard-Prozess $\{X_t\}_{t \geq 0}$ durch lineare Transformation:

$$B_t = B_0 + \sigma X_t + \mu t, \quad t \geq 0 \quad (\sigma > 0, \mu \in \mathbb{R})$$

mit beliebigem Anfangswert $B_0 \in \mathbb{R}$. Im allgemeinen Sprachgebrauch heißt

- σ der *Diffusionskoeffizient* oder (vor allem in finanzmathematischen Anwendungen) die *Volatilität*
- μ die *Drift*.

Lemma 55. Ein(e) allgemeine Brown'sche Bewegung / allgemeiner Wiener-Prozess $\{X_t\}_{t \geq 0}$ besitzt folgende Eigenschaften:

- Jedes X_t ist $\mathcal{N}(B_0 + \mu t, \sigma^2 t)$ -verteilt für $t \geq 0$ (mit $\mathcal{N}(B_0, 0) = \varepsilon_{B_0}$).
- Jeder Pfad von $\{X_t\}_{t \geq 0}$ ist [fast sicher] nirgends differenzierbar und von unbeschränkter Variation in jedem endlichen Intervall.
- Es gilt $\text{Cov}(X_s, X_t) = \sigma^2 \min\{s, t\}$ für $s, t \geq 0$.
- Für jede streng aufsteigende endliche Folge $\{t_k\}_{k=1}^n \subset \mathbb{R}^+$ ist $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ multivariat normalverteilt nach $\mathcal{N}(\mathbf{v}, \Sigma)$ mit $\mathbf{v} = (B_0 + t_1 \mu, \dots, B_0 + t_n \mu)^T$ und

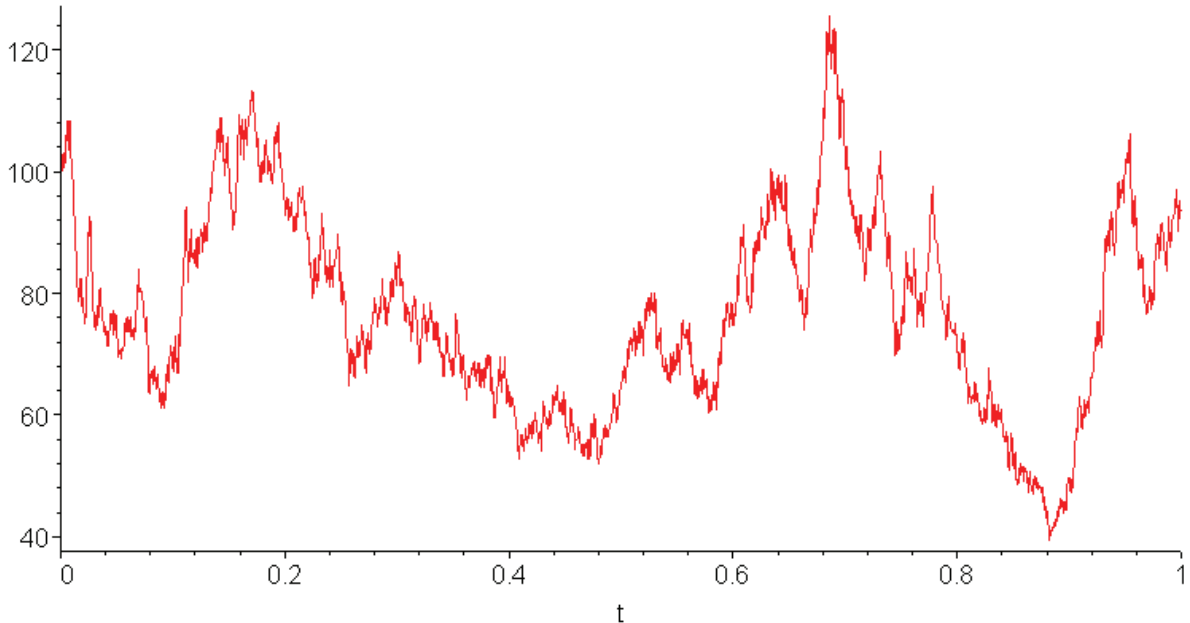
$$\sigma_{ij} = \sigma^2 \min\{t_i, t_j\}, \quad 1 \leq i, j \leq n \in \mathbb{N}.$$

IV. Stochastische Prozesse

Bemerkung: In den klassischen stochastischen Finanzmarktmodellen wird auf Grund der Nichtnegativitätsbedingungen an die Kurse (z.B. Aktienkurse, festverzinsliche Wertpapiere) statt der allgemeinen Brown'schen Bewegung $\{B_t\}_{t \geq 0}$ die *geometrische Brown'sche Bewegung* $\{S_t\}_{t \geq 0}$ betrachtet:

$$S_t := \exp(B_t), \quad t \geq 0.$$

Diese besitzt ähnliche Eigenschaften wie die Brown'sche Bewegung (insbesondere die Stetigkeit, aber Nichtdifferenzierbarkeit der Pfade). Sie ist z.B. Grundlage der *Black-Scholes*-Markttheorie.



Simulierter Pfad eines Standard-Black-Scholes-Aktienkursprozesses mit Anfangswert 100 €

Bemerkung: Die (Standard-)Brown'sche Bewegung $\{B_t\}_{t \geq 0}$ ist auch Grundlage zahlreicher stochastischer Differenzialgleichungen, die formal so aussehen:

$$dX_t = \mu(t, X_t)dt + \sigma(t, X_t)dB_t$$

mit geeigneten (in der Regel stetigen) Funktionen $\mu, \sigma: \mathbb{R}^2 \rightarrow \mathbb{R}$ ($\sigma \geq 0$) und bekanntem Anfangswert $X_0 \in \mathbb{R}$. Im strengen Sinne handelt sich hier eigentlich um eine (Itô-)Integralgleichung der Art

$$X_t = \int_0^t \mu(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s.$$

Wir wollen hier nicht genauer auf die zu Grunde liegende Theorie eingehen, sondern nur die Möglichkeit einer approximativ rekursiven Auflösung in einem vorgegebenen Zeitintervall $[0, T]$ angeben. Dazu geben wir zunächst eine heuristische Erklärung des „Differenzial“-Terms dB_t :

IV. Stochastische Prozesse

Für stetige Funktionen σ und X_\bullet gilt approximativ:

$$\int_t^{t+h} \sigma(s, X_s) dB_s \approx \sigma(t, X_t) \cdot (B_{t+h} - B_t) = \sigma(t, X_t) \cdot \sqrt{h} \cdot Z \quad \text{für kleine } h > 0,$$

wobei Z eine standard-normalverteilte Zufallsgröße ist (weil $B_{t+h} - B_t$ ja gerade $\mathcal{N}(0, h)$ -verteilt ist mit Standardabweichung \sqrt{h}).

Dieser Ansatz erklärt heuristisch folgende Vorgehensweise:

Durch Zerlegung des Zeitintervalls $[0, T]$ in $n \in \mathbb{N}$ äquidistante Teilintervalle erhält man die Stützpunkte $t_k = k \cdot h_n$ mit $h_n = \frac{T}{n}$, $k = 0, \dots, n$ und damit näherungsweise die Rekursion

$$X_{t_{k+1}}^{[n]} - X_{t_k}^{[n]} = h_n \cdot \mu(t_k, X_{t_k}^{[n]}) + \sqrt{h_n} \cdot \sigma(t_k, X_{t_k}^{[n]}) Z_{k+1}^{[n]}, \quad k = 0, 1, \dots, n-1$$

mit unabhängigen standard-normalverteilten Zufallsvariablen $Z_1^{[n]}, \dots, Z_n^{[n]}$.

Nach dem Konstruktionssatz 53 liefert dies das Anfangsstück einer Markoff-Kette $\{X_{t_k}^{[n]}\}_{k \in \mathbb{N}}$, die homogen ist, wenn die Abbildungen μ und σ nicht von der ersten (Zeit-)Komponente abhängen. Die "echte" Lösung $\{X_t | t \geq 0\}$ der stochastischen Differentialgleichung ist in der Regel ein Markoff-Prozess.

An dieser Rekursion ist auch ersichtlich, warum „stochastische Differenziale“ im klassischen Sinn der Analysis nicht existieren (können). Nach Division durch den Term $h_n = t_1 - t_0 = t_1 = \frac{T}{n}$ erhält man nämlich

$$\frac{X_{t_1}^{[n]} - X_{t_0}^{[n]}}{t_1 - t_0} = \mu(t_0, X_0) + \sqrt{\frac{n}{T}} \sigma(t_0, X_0) Z_1^{[n]}, \quad k = 0, 1, \dots, n-1;$$

ein „Grenzwert“ für $n \rightarrow \infty$ auf der rechten Seite existiert aber nicht für den zweiten Summanden.

Eine besondere stochastische Differentialgleichung ist

$$dX_t = \mu X_t dt + \sigma X_t dB_t$$

mit Konstanten $\mu \in \mathbb{R}$, $\sigma > 0$. Die Lösung mit dem Itô-Integral (geometrische Brown'sche Bewegung) lautet hier, abweichend von der Intuition:

$$X_t = X_0 \cdot \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t\right), \quad t \geq 0.$$

Für vertiefende Literatur verweisen wir auf die Monographie von ETHERIDGE (2002).

Verzeichnis der Definitionen, Sätze und Lemmata

Definition	Satz	Lemma	Seite
		1	6
1			7
		2	7
2			8
3			8
	1		9
		3	10
4			11
		4	11
	2		12
	3		14
5			15
	4		16
		5	16
6			17
		6	18
7			19
	5		19
	6		22
8			23
		7	23
9			25
	7		25
	8		27
10			29
11			31
		8	31
		9	33
		10	33
		11	34
		12	34
	9		35
	10		36
12			37
13			37
	11		38
	12		38
14			40
		13	40
	13		41
15			43
		14	43
	14		43
		15	47
	15		47
16			49

Definition	Satz	Lemma	Seite
17			49
		16	49
	16		50
		17	51
	17		52
18			52
		18	53
		19	54
19			55
20			55
		20	56
	18		57
		21	58
		22	58
	19		58
21			60
	20		60
	21		61
	22		61
	23		62
	24		64
	25		64
	26		65
22			66
		23	67
23			72
24			73
25			73
		24	74
		25	74
26			75
		26	75
	27		77
27			77
		27	78
		28	79
	28		81
28			83
29			83
		29	83
	29		84
	30		86
30			87
		30	87
	31		90
31			94

Definition	Satz	Lemma	Seite
		31	95
		32	96
32			98
		33	99
	32		99
		34	103
		35	103
		36	104
33			105
		37	106
34			110
35			112
		38	112
		39	112
36			118
	33		118
	34		123
		40	123
		41	125
	35		126
		42	127
		43	129
		44	129
37			134
38			135
39			135
	36		136
	37		137
		45	138
		46	139
	38		140
		47	142
	39		142
40			143
		48	143
	40		145
	41		146
	42		148
	43		150
	44		151
41			153
42			154
	45		154
	46		155
	47		156
	48		158

Definition	Satz	Lemma	Seite
	49		160
43			161
		49	162
	50		165
44			166
	51		169
	52		169
		50	170
45			171
46			172
47			178
48			181
49			201
		51	202
50			202
	53		203
	54		204
51			211
52			212
	55		214
	56		216
53			218
		52	218
		53	219
54			221
		54	221
55			222
		55	222

Literatur

- [1] F. BARTH, R. HALLER (1998): *Stochastik Leistungskurs*. 12. Aufl., Oldenbourg Verlag, München.
- [2] H. BAUER (2002): *Wahrscheinlichkeitstheorie*. 5. Aufl., Walter de Gruyter, Berlin.
- [3] H. BAUER (1992): *Maß- und Integrationstheorie*. 2. Aufl., Walter de Gruyter, Berlin.
- [4] K. BEHNEN, G. NEUHAUS (2003): *Grundkurs Stochastik*. 4. Aufl., PD-Verlag, Heidenau (ISBN 3-930737-69-8)
- [5] J.J. BENEDETTO (1976): *Real Variable and Integration*. Teubner, Stuttgart.
- [6] P. BILLINGSLEY (1986): *Probability and Measure*. 2nd ed., Wiley, N.Y.
- [7] A.A. BOROVKOV (2013): *Probability Theory*. Springer, London.
- [8] C. CARATHÉODORY (1918): *Vorlesungen über reelle Funktionen*. Teubner Verlag, Leipzig.
- [9] T. COVER, B. GOPINATH (EDS.) (1987): *Open Problems in Communication and Computation*. Springer, N.Y.
- [10] C. CZADO, T. SCHMIDT (2011): *Mathematische Statistik*. Springer, Berlin.
- [11] B. DE FINETTI (1990): *Theory of Probability: A Critical Introductory Treatment*. Vol. I and II. Wiley Classics Library Edition, Wiley, New York.
- [12] B. DE FINETTI (1981): *Wahrscheinlichkeitstheorie*. Einführende Synthese mit kritischem Anhang. Oldenbourg Verlag, München.
- [13] H. DEHLING, B. HAUPT (2003): *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Springer, Berlin.
- [14] O. DEISER (2010): *Einführung in die Mengenlehre*. 3. Aufl., Springer, Berlin.
- [15] H. DINGES, H. ROST (1982): *Prinzipien der Stochastik*. B.G. Teubner, Stuttgart.
- [16] R.G. DOWNEY, D.R. HIRSCHFELDT (2010): *Algorithmic Randomness and Complexity*. Springer, New York.
- [17] J. ELSTRODT (1996): *Maß- und Integrationstheorie*. Springer, Berlin.
- [18] A. ETHERIDGE (2002): *A course in Financial Calculus*. Cambridge University Press, Cambridge.
- [19] L. FAHRMEIR, A. HAMERLE, G. TUTZ (Hrsg.) (1996): *Multivariate statistische Verfahren*. 2. Aufl., de Gruyter, Berlin.
- [20] W. FELLER (1968, 1971): *An Introduction to Probability Theory and Its Applications*. Vol. I and II. 2nd and 3rd ed., Wiley, New York.
- [21] D. FOATA, A. FUCHS (1999): *Wahrscheinlichkeitsrechnung*. Birkhäuser, Basel.
- [22] A. KLENKE (2008): *Wahrscheinlichkeitstheorie*. 2. Aufl., Springer, Berlin.
- [23] P. GÄNSSLER, W. STUTE (1977): *Wahrscheinlichkeitstheorie*. Springer, Berlin.
- [24] H.-O. GEORGII (2002): *Stochastik*. Einführung in die Wahrscheinlichkeitstheorie und Statistik. Walter de Gruyter, Berlin.
- [25] R. HAFNER (1989): *Wahrscheinlichkeitsrechnung und Statistik*. Springer, Berlin.
- [26] F. HAUSDORFF (1914): *Grundzüge der Mengenlehre*. Veit & Comp., Leipzig.

- [27] N. HENZE (2003): *Stochastik für Einsteiger*. 4. Aufl., Vieweg, Braunschweig.
- [28] H. HEUSER (1988): *Lehrbuch der Analysis*, Teil 2. Teubner Verlag, Stuttgart.
- [29] A. IRLE (2000): *Wahrscheinlichkeitstheorie und Statistik*. B.G. Teubner, Stuttgart.
- [30] H. KNÖPFEL, M. LÖWE (2007): *Stochastik - Struktur im Zufall*. Oldenbourg Verlag, München.
- [31] E.T. JAYNES (2006). *Probability Theory. The Logic of Science*. 6th ed., Cambridge University Press, Cambridge.
- [32] U. KNAUER (2001): *Diskrete Strukturen – kurz gefasst*. Spektrum Akademischer Verlag, Heidelberg.
- [33] U. KRENGEL (2003): *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 7. Aufl., Vieweg, Braunschweig.
- [34] K. KRICKEBERG, H. ZIEZOLD (1995): *Stochastische Methoden*. 4. Aufl., Springer, Berlin.
- [35] R. LEADBTTER, S. CAMBANIS, V. PIPIRAS (2014): *A Basic Course in Measure and Probability. Theory for Applications*. Cambridge University Press, Cambridge.
- [36] J. LEHN, H. WEGMANN (2000): *Einführung in die Statistik*. 3. Aufl., B.G. Teubner, Stuttgart.
- [37] R. MATHAR, D. PFEIFER (1990): *Stochastik für Informatiker*. B.G. Teubner, Stuttgart.
- [38] K. NAWROTZKI (1994): *Lehrbuch der Stochastik*. Verlag Harri Deutsch, Frankfurt / Main.
- [39] R. B. NELSEN (1999): *An Introduction to Copulas*. Lecture Notes in Statistics 139, Springer, N.Y.
- [40] A. NIES (2009): *Computability and Randomness*. Oxford University Press, Oxford.
- [41] J. PFANZAGL (1991): *Elementare Wahrscheinlichkeitsrechnung*. 2. Aufl., Walter de Gruyter, Berlin.
- [42] D. PLACHKY (1980): *Stochastik II*. Akademische Verlagsgesellschaft, Wiesbaden.
- [43] H. PRUSCHA (2000): *Vorlesungen über Mathematische Statistik*. B.G. Teubner, Stuttgart.
- [44] S.I. RESNICK (2001): *A Probability Path*. 2nd ed., Birkhäuser, Basel.
- [45] K.D. SCHMIDT (2011): *Maß und Wahrscheinlichkeit*. 2. Aufl., Springer, Berlin.
- [46] I. SCHNEIDER (Hrsg.) (1988): *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Wiss. Buchgesellschaft, Darmstadt.
- [47] K. SCHÜRGER (1998): *Wahrscheinlichkeitstheorie*. Oldenbourg Verlag, München.
- [48] E. WARMUTH, W. WARMUTH (1998): *Elementare Wahrscheinlichkeitsrechnung*. B.G. Teubner, Stuttgart.
- [49] J. WENGENROTH (2008): *Wahrscheinlichkeitstheorie*. de Gruyter, Berlin.

Danksagung

Ohne die tatkräftige Unterstützung von Dr. Christian Mohn, Dr. Doreen Straßburger, Dipl.-Math. Hero Wanders, Dipl.-Math. Martin Hampel und M.Sc. Dominic Lauterbach wäre die Ausgestaltung des Textes in dieser Form nicht möglich gewesen. Dafür sei ihnen allen an dieser Stelle herzlich gedankt.