Julius-Maximilians-
UNIVERSITÄT
WÜRZBURG

Skript, Wintersemester 2024/25

Stochastik I

Prof. Dr. Markus Bibinger
Institut für Mathematik
Lehrstuhl für Mathematik VIII (Angewandte Stochastik)

Adrian Grüber
Assistent

Homepage
www.mathematik.uni-wuerzburg.de/appliedstochastics

Julius-Maximilians-Universität Würzburg
Lehrstuhl für Angewandte Stochastik (Mathematik VIII)
Fakultät für Mathematik und Informatik, Institut für Mathematik

Emil-Fischer-Straße 30, D-97074 Würzburg
Tel. +49 931 31-87610
markus.bibinger@uni-wuerzburg.de
adrian.grueber@uni-wuerzburg.de
https://www.mathematik.uni-wuerzburg.de/appliedstochastics

Version: 4. Oktober 2024

Inhaltsverzeichnis

Vorwort:

Das vorliegende Skript ist begleitend zur 4+2 SWS Veranstaltung "Stochastik I" im Wintersemester 2024/25 an der Universität Würzburg. Skript und Vorlesung wenden sich an Studierende in den mathematischen Bachelorstudiengängen. Mathematische Grundkenntnisse aus der Analysis und der linearen Algebra werden vorausgesetzt.

Vorlesung und Skript orientieren sich unter anderem an folgenden drei deutschen Standardlehrbüchern:

- Norbert Henze (2010) *Stochastik für Einsteiger*. 8. Auflage. Vieweg, Wiesbaden.
- Ulrich Krengel (2005) *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 8. Auflage. Vieweg, Wiesbaden.
- Herold Dehling und Beate Haupt (2004) *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. 2. Auflage. Springer, Berlin.

Das Buch von Henze ist sehr gut zugänglich und das Buch von Krengel enthält - auf vergleichbarem Schwierigkeitsgrad wie die Vorlesung - noch mehr Material über Stochastik, und ist daher als Ergänzung geeignet. Jedoch werden wir einige tiefere Grundlagen der Maß- und Integrationstheorie entwickeln als diese Einführungen in die Stochastik.

Empfehlenswert ist auch das Buch

- H. O. Georgii (2009) *Stochastik*. 4. Auflage. De Gruyter Verlag,

welches gut zugänglich ist, und dabei auch weiterführende Inhalte auf höherem mathematischen Niveau behandelt.

Ein empfehlenswertes, umfassenderes Buch zur Stochastik mit Grundlagen der Maß- und Integrationstheorie ist

- A. Klenke (2008) *Wahrscheinlichkeitstheorie.* Berlin: Springer.

Zur Maß- und Integrationstheorie sind unter anderem folgende Bücher empfehlenswert:

- Elstrodt, J. (1996) *Maß- und Integrationstheorie*, Springer.
- Bauer, H. (1992) *Maß- und Integrationstheorie*, 2. Auflage, de Gruyter.
- Billingsley, P. (2012) *Probability and Measure*, 4. Auflage, Wiley.

Hyperlinks und Links sind blau markiert, wichtige Begriffe im Text werden *blau und kursiv hervorgehoben* und in Sätzen und Definitionen durch **breite Schrift**. Wichtige Feststellungen außerhalb von Sätzen werden rotbraun hervorgehoben. Das Skript wird in erster Version neu erstellt und kapitelweise vorlesungsbegleitend veröffentlicht. Für Hinweise zu eventuellen Fehlern und Anmerkungen bin ich dankbar.

1 Laplace-Wahrscheinlichkeit und Kombinatorik

1.1 Historische Entwicklung

Der Begriff *Stochastik* kommt aus dem Altgriechischen von *stochastike* = zum Erraten gehörende Kunst, und steht als Sammelbegriff für Wahrscheinlichkeitsheorie und Statistik.

Die Ursprünge der Wahrscheinlichkeitstheorie liegen im 16. Jahrhundert, als man sich zunehmend für die beim Würfelspiel beobachteten Zufallsgesetze interessierte. Hervorzuheben sind hier das Werk von CARDANO (1501 - 1576), mit Titel "Liber de ludo aleae", sowie ein Briefwechsel zwischen BLAISE PASCAL (1623-1682) und PIERRE FERMAT (1601 - 1665) aus dem Jahr 1654 über Gewinnaussichten in Spielsituationen, in dem diese bereits Begriffe wie *Wahrscheinlichkeit* und *Erwartungswert* erfassen. CHRISTIAN HUYGENS (1629 - 1695) verfasste dann 1657 eine vollständige Theorie des Würfelspiels.

Um eine Vorstellung von den zu dieser Zeit diskutierten Problemen zu bekommen, betrachten wir folgende Situation: Bei zwei Würfelspielen wird

- mit einem Würfel geworfen, es gibt 6 mögliche Ausgänge. Wie groß ist die Wahrscheinlichkeit, in 4 Würfen mindestens eine Sechs zu erhalten?
- mit zwei Würfeln geworfen, hier gibt es $6 \cdot 6 = 36$ mögliche Ausgänge. Wie groß ist die Wahrscheinlichkeit, in $6 \cdot 4 = 24$ Würfen mindestens eine Doppelsechs zu erhalten?

Einige Zeit wurde diesen beiden Ereignissen fälschlicherweise die gleiche Wahrscheinlichkeit zugeordnet, da die Verhältnisse $6 : 4$ und $36 : 24$ gleich sind.

Weitere wichtige Stationen bei der Entwicklung der Wahrscheinlichkeitstheorie waren:

- 1713: JAKOB BERNOULLI (1654 - 1705) verfasst das Werk "Ars conjectandi". Dies war das erste grundlegende Werk über Wahrscheinlichkeitsrechnung und enthält bereits das Gesetz der großen Zahlen.
- 1812 PIERRE SIMON DE LAPLACE (1749 - 1827) verfasst das Werk "Theorie analytique des probabilites". Er definiert die Wahrscheinlichkeit eines Ereignisses als das Verhältnis der günstigen Fälle zu den möglichen Fällen.
- 1919 RICHARD VON MISES (1883 - 1953) verfasst das Werk "Grundlagen der Wahrscheinlichkeitsrechnung". Er definiert Wahrscheinlichkeit als Grenzwert der relativen Häufigkeit.
- 1933 ANDREI NIKOLAJEWITSCH KOLMOGOROFF (1903 - 1987) verfasst das Werk "Grundbegriffe der Wahrscheinlichkeitsrechnung", in dem er die bis heute übliche axiomatische Begründung der Wahrscheinlichkeitstheorie darlegt.

Moderne Wahrscheinlichkeitstheorie umfasst auch die Theorie der stochastischen Prozesse, d.h. der zufälligen Phänome, die sich dynamisch über die Zeit entwickeln. Ein grundlegendes Beispiel ist die *Brownsche Bewegung* bzw. der *Wiener Prozess*. Diese wurde eingeführt von dem britischen Botaniker ROBERT BROWN (1773-1858) als Modell für die Bewegung schwimmender Partikel in Flüssigkeiten. ALBERT EINSTEIN (1879 - 1955) benutzte die Brownsche Bewegung zur Erklärung der molekularen Struktur von Wasser. LOUIS BACHELIER promovierte 1900 in Paris über die Theorie der Spekulation an Finanzmärkten und begründete damit die Finanzmathematik, die auf stochastischen Prozessen basiert. NORBERT WIENER (1894 - 1964) gab schließlich eine formale mathematische Konstruktion – und damit einen Existenzbeweis – der Brownschen Bewegung als stochastischer Prozess. Die Brownsche Bewegung ist von grundlegender Bedeutung in verschiedenen Wissenschaften wie Physik, Biologie, Finanzmathematik, und hat auch Anwendungen in der reinen Mathematik.
Ein anderer wichtiger Teilbereich der Stochastik ist die (mathematische) Statistik, mit der sich stochastische Modelle anhand von Daten kalibrieren lassen.

1.2 Was ist Wahrscheinlichkeit?

Was ist Wahrscheinlichkeit? Empirisch kann man Wahrscheinlichkeit in einem Zufallsexperiment als *(relative) Häufigkeit* eines Ereignisses eines physikalischen Prozesses betrachten, dessen Ausgang nicht (exakt) vorhersagbar ist. Man unterscheidet zwischen

- *deterministischen Prozessen*, die bei vollständiger Information prinzipiell vorhersagbar sind (Wetter), sowie
- *nichtdeterministischen Prozessen*, die prinzipiell nicht vorhersagbar ist (radioaktiver Zerfall).

Wir betrachte zwei Beispiele.

Beispiel 1.1 (Roulette)**.** Der *Ergebnisraum* dieses Zufallsexperimentes ist

$$\Omega = \{0, 1, 2, \dots, 35, 36\}.$$

Elemente $\omega \in \Omega$ heißen *Elementarereignisse*, allgemeine Teilmengen $A \subseteq \Omega$ heißen *Ereignisse*. Beispiele beim Roulette sind

- Ereignis "rot": $A = \{1,3,5,7,9,12,14,16,18,19,21,23,25,27,30,32,34,36\}$ oder "schwarz" (die anderen bis auf 0).
- Ereignisse "gerade" (ohne 0) oder "ungerade".
- Ereignis "klein" $(1-18)$ oder "groß" $(19-36)$.

Im fairen Roulette ist jedes Elementarereignis $\omega \in \Omega$ gleich wahrscheinlich, also

$$\mathbb{P}\big(\{\omega\}\big) = \frac{1}{37}, \qquad \omega \in \Omega.$$

Wir sprechen dann von *Laplace-Wahrscheinlichkeiten* und einem *Laplace-Raum*. Für die Wahrscheinlichkeiten von allgemeinen Ereignissen $A \subseteq \Omega$ erhält man

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0,1], \qquad \mathbb{P}(A) = \frac{\operatorname{card} A}{37}.$$

Wir schreiben $\mathcal{P}(\Omega)$ für die *Potenzmenge*, die Menge aller Teilmengen von Ω, manchmal auch *Ereignismenge* genannt. Für $A \in \mathcal{P}(\Omega)$ bezeichnet $\operatorname{card} A$, oder $|A|$, die Kardinalität von A, also für endliche Mengen Ω die Anzahl der Elementarereignisse $\omega \in A$. Wir bemerken, dass die oben definierte "Wahrscheinlichkeitsfunktion" $\mathbb{P}$ *additiv* ist: Sind $A, B \subseteq \Omega$ disjunkt, also $A \cap B = \emptyset$, so ist

$$\mathbb{P}\big(A \cup B\big) = \frac{\operatorname{card}(A \cup B)}{37} = \frac{\operatorname{card} A}{37} + \frac{\operatorname{card} B}{37} = \mathbb{P}(A) + \mathbb{P}(B).$$

Beispiel 1.2 (Reißzwecke werfen)**.** Wir werfen wiederholt eine Reißzwecke und notieren

- 0, falls die Spitze schräg nach unten liegt,
- 1, falls die Spitze nach oben zeigt.

In diesem Zufallsexperiment sind die Wahrscheinlichkeiten nicht wie in dem obigen Beispiel berechenbar. Man nähert sie dann durch *relative Häufigkeiten* an. Sei Ω der Ergebnisraum, beim Wurf der Reißzwecke etwa $\Omega = \{0, 1\}$. Wir führen nun das Zufallsexperiment mit Ergebnissen in Ω unter gleichen Bedingungen n mal durch und erhalten ein n-Tupel $\mathbf{a} = (a_1 \dots, a_n)$, $a_i \in \Omega$.

Für ein Ereignis $A \subseteq \Omega$ setzt man dann

$$r_{n,\mathbf{a}}(A) = \frac{1}{n} \operatorname{card}\{j : a_j \in A,\ j = 1, \dots, n\},$$

also ist $r_{n,\mathbf{a}}(A)$ die *relative Häufigkeit* des Ereignisses A im n-Tupel $\mathbf{a}$. Relative Häufigkeiten erfüllen $0 \leq r_{n,\mathbf{a}}(A) \leq 1$, $A \subseteq \Omega$, $r_{n,\mathbf{a}}(\Omega) = 1$ sowie

$$r_{n,\mathbf{a}}(A \cup B) = r_{n,\mathbf{a}}(A) + r_{n,\mathbf{a}}(B) \qquad \text{falls} \quad A \cap B = \emptyset,$$

sind also auch additiv. Bei 300 Würfen einer Reißzwecke ergab sich etwa $r_{n,a}(\{1\}) = \frac{124}{300}$.

Für relative Häufigkeiten beobachtet man nun ein *Empirisches Gesetz der großen Zahlen*: Die relativen Häufigkeiten "stabilisieren" sich für wachsendes n. Dies interpretiert man wie folgt: Relative Häufigkeiten konvergieren gegen die zugrundeliegenden Wahrscheinlichkeiten, diese behalten als Grenzwerte die obigen Eigenschaften. Man definiert nun Wahrscheinlichkeiten aber nicht als solche Grenzwerte, sondern axiomatisch über die Eigenschaften, die sich oben ergeben haben. Ein *endlicher Wahrscheinlichkeitsraum* $(\Omega, \mathbb{P})$ besteht aus einem endlichen nicht-leeren Ergebnisraum Ω und einer Abbildung

$$\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$$

mit den Eigenschaften
1. $\mathbb{P}(\Omega) = 1$,
2. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ für $A, B \subseteq \Omega$, $A \cap B = \emptyset$.

$\mathbb{P}$ ist ein *Wahrscheinlichkeitsmaß* auf Ω. Wir werden diese Begriffe unten noch allgemeiner definieren. Ist $(\Omega, \mathbb{P})$ so ein endlicher Wahrscheinlichkeitsraum, so gilt für paarweise disjunkte Ereignisse $A_1, \ldots, A_n \subseteq \Omega$, dass

$$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathbb{P}(A_k).$$

Wir bezeichnen dies als *(endliche) Additivität*. Dies folgt durch Induktion. Der Induktionsanfang gilt nach Voraussetzung. Für den Induktionsschritt von n nach $n+1$ rechnen wir

$$\begin{aligned}
\mathbb{P}\Big(\bigcup_{k=1}^{n+1} A_k\Big) &= \mathbb{P}\Big(\Big(\bigcup_{k=1}^{n} A_k\Big) \cup A_{n+1}\Big) \\
&= \mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) + \mathbb{P}(A_{n+1}) = \sum_{k=1}^{n} \mathbb{P}(A_k) + \mathbb{P}(A_{n+1}).
\end{aligned}$$

Insbesondere erhält man aus der endlichen Additivität bei endlichen Wahrscheinlichkeitsräumen für $A \subseteq \Omega$

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\}). \tag{1.1}$$

Das Wahrscheinlichkeitsmaß ist also durch Auswertung auf den Elementarereignissen bereits eindeutig bestimmt.

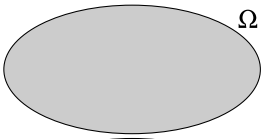
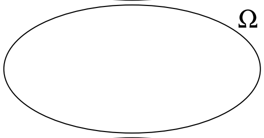
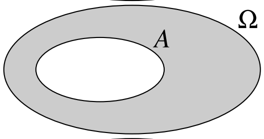
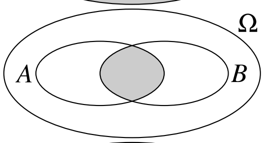
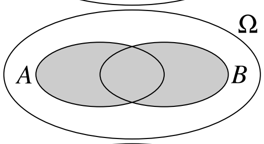
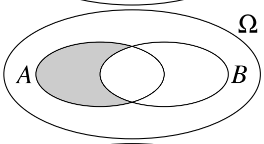
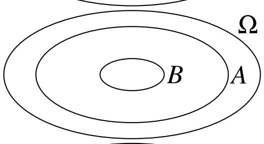
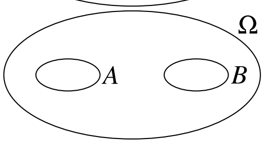
Wir wiederholen einige grundsätzliche Aussagen über Mengen aus der Analysis 1. Die wichtigsten Mengenoperationen sind in Tabelle 1.1 dargestellt.

Definition 1.3. *Ist A eine Menge, so heißt die Anzahl der Elemente von A die* ***Mächtigkeit*** *oder* ***Kardinalität*** *von A. Übliche Schreibweisen sind* $\operatorname{card} A$, $\#A$ *oder* $|A|$.

Satz 1.4 (Rechenregeln für Mengen). *Sind* $A, B, C, A_1, \ldots, B_1, \ldots$ *Mengen und* $I \subseteq \mathbb{N}$ *eine beliebige Indexmenge, so gilt:*

1. $A \setminus B = A \cap B^{\complement}$,
2. *Distributivgesetze:*
   a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$,

Tabelle 1.1: Darstellung der wichtigsten Mengennotationen

Darstellung	Notation	Bedeutung
	Ω	sicheres Ereignis
	$\emptyset$	unmögliches Ereignis
	$A^{\complement}$	*A* nicht eingetreten
	$A \cap B$	*A* und *B* eingetreten
	$A \cup B$	*A* oder *B* eingetreten
	$A \setminus B$	*A*, aber nicht *B* eingetreten
	$B \subseteq A$	Wenn *B* eintritt, dann auch *A*
	$A \cap B = \emptyset$	*A* und *B* schließen sich aus

b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$,

oder allgemein

a) $A \cap \left(\bigcup_{i \in I} B_i\right) = \bigcup_{i \in I} (A \cap B_i)$,

b) $A \cup \left(\bigcap_{i \in I} B_i\right) = \bigcap_{i \in I} (A \cup B_i)$,

3. DE MORGAN*sche Regeln:*

a) $(A \cap B)^{\complement} = A^{\complement} \cup B^{\complement}$,

b) $(A \cup B)^{\complement} = A^{\complement} \cap B^{\complement}$,

oder allgemein

a) $(\bigcap_{i \in I} A_i)^{\complement} = \bigcup_{i \in I} A_i^{\complement}$,

b) $(\bigcup_{i \in I} A_i)^{\complement} = \bigcap_{i \in I} A_i^{\complement}$.

1.3 Urnenmodelle

Urnenmodelle liefern Abzählmethoden zur Bestimmung von Wahrscheinlichkeiten in bestimmten LAPLACE-Räumen. Dabei stellen wir uns eine Urne mit n durchnummerierten Kugeln vor, aus der k-mal gezogen wird. Dabei bestehen folgende Möglichkeiten:

- mit Zurücklegen mit Beachtung der Reihenfolge:

$$\Omega_{\mathrm{I}}^{k,n} = \{1, \dots, n\}^k \text{ mit } \operatorname{card} \Omega_{\mathrm{I}} = n^k.$$

- ohne Zurücklegen mit Beachtung der Reihenfolge:

$$\Omega_{\mathrm{II}}^{k,n} = \{\omega = (\omega_1, \dots, \omega_k) \in \{1, \dots, n\}^k \mid \omega_i \neq \omega_j \text{ für } i \neq j \in \{1, \dots, k\}\}$$

mit

$$\operatorname{card} \Omega_{\mathrm{II}}^{k,n} = n \cdot (n-1) \cdot \dots \cdot (n-k+1) = \frac{n!}{(n-k)!} = (n)_k. \quad (\text{“}n \text{ unten } k\text{”})$$

- ohne Zurücklegen ohne Beachtung der Reihenfolge:

$$\Omega_{\mathrm{III}}^{k,n} = \{A \subseteq \{1, \dots, n\} \mid \operatorname{card} A = k\}$$

ist die Menge der k-elementigen Teilmengen von $\{1, \dots, n\}$. Sind $k, n \in \mathbb{N}$ mit $k \leq n$, so gilt

$$\operatorname{card} \Omega_{\mathrm{III}}^{k,n} = \frac{(n)_k}{k!} = \frac{n!}{(n-k)!k!} = \binom{n}{k}. \quad (\text{“}n \text{ über } k\text{”})$$

Wir definieren eine Abbildung $\varphi : \Omega_{\mathrm{II}}^{k,n} \to \Omega_{\mathrm{III}}^{k,n}$ mit $\varphi(\omega_1, \dots, \omega_k) = \{\omega_1, \dots, \omega_k\}$. Da alle Elemente gewählt werden können, ist φ surjektiv. Außerdem, da die Urbilder $\varphi^{-1}(\{\omega_1, \dots, \omega_k\})$ genau alle Permutationen von $\omega_1, \dots, \omega_k$ enthalten, gilt stets

$$\operatorname{card} \varphi^{-1}(\{\omega_1, \dots, \omega_k\}) = k!. \tag{1.2}$$

Es folgt direkt

$$\operatorname{card} \Omega_{\mathrm{II}}^{k,n} = k! \operatorname{card} \Omega_{\mathrm{III}}^{k,n} \Leftrightarrow \operatorname{card} \Omega_{\mathrm{III}}^{k,n} = \frac{\operatorname{card} \Omega_{\mathrm{II}}^{k,n}}{k!} = \frac{(n)_k}{k!}.$$

Tabelle 1.2: Urnenmodell und Murmelmodell

Urnenmodell	mit Zurücklegen	ohne Zurücklegen	
mit Reihenfolge	Ω_{I}	Ω_{II}	Murmeln unterscheidbar
ohne Reihenfolge	Ω_{IV}	Ω_{III}	Murmeln ununterscheidbar
	mit Mehrfachbesetzung der Zellen	ohne Mehrfachbesetzung der Zellen	Murmelmodell

Bemerkung. Wir können eine Äquivalenzrelation $\sim$ zwischen Tupeln definieren mit

$$\begin{aligned}(\omega_1,\dots,\omega_k)\sim(\tilde\omega_1,\dots,\tilde\omega_k) :\Leftrightarrow {} & \{\omega_1,\dots,\omega_k\}=\{\tilde\omega_1,\dots,\tilde\omega_k\}\\ \Leftrightarrow {} & \exists\pi\in\mathfrak{S}_k:\forall i\in\{1,\dots,k\}:\omega_i=\tilde\omega_{\pi(i)}.\end{aligned}$$

Diese Äquivalenzklassen haben je $\operatorname{card}\mathfrak{S}_k=k!$ Elemente und jeweils eindeutige Repräsentanten $\omega=(\omega_1,\dots,\omega_k)$ mit $\omega_1<\dots<\omega_k$. Man könnte also $\Omega_{\mathrm{III}}^{k,n}$ auch definieren als

$$\Omega_{\mathrm{III}'}^{k,n}=\{\omega=(\omega_1,\dots,\omega_k)\in\{1,\dots,n\}^k\mid\omega_1<\dots<\omega_k\}.$$

- mit Zurücklegen ohne Beachtung der Reihenfolge:

  $$\Omega_{\mathrm{IV}}^{k,n}=\{\omega=(\omega_1,\dots,\omega_k)\in\{1,\dots,n\}^k\mid\omega_1\le\dots\le\omega_k\}.$$

  Als Ergebnis erhalten wir k Ziffern aus $\{1,\dots,n\}$, die nicht notwendig verschieden und beliebig angeordnet sind, etwa aufsteigend.

  Sind $n,k\in\mathbb{N}$, so gilt

  $$\operatorname{card}\Omega_{\mathrm{IV}}^{k,n}=\binom{k+n-1}{k}=\binom{k+n-1}{n-1}.$$

  Wir definieren eine Abbildung $\psi:\Omega_{\mathrm{IV}}^{k,n}\to\Omega_{\mathrm{III}}^{k,k+n-1}$ mit $(\psi(\omega))_i=\omega_i+i-1$ für $i\in\{1,\dots,k\}$. Die Abbildung ψ ist wohldefiniert, da für $i\in\{2,\dots,k\}$

  $$(\psi(\omega))_{i-1}=\omega_{i-1}+i-2\le\omega_i+i-2<\omega_i+i-1=(\psi(\omega))_i$$

  gilt und für $i\in\{1,\dots,k\}$ außerdem $1\le(\psi(\omega))_i\le n+k-1$ gilt. ψ ist sogar bijektiv, da mit $(\psi^{-1}(\omega))_i=\omega_i-i+1$ eine Umkehrabbildung angegeben werden kann. Also gilt

  $$\operatorname{card}\Omega_{\mathrm{IV}}^{k,n}=\operatorname{card}\Omega_{\mathrm{III}}^{k,k+n-1}=\binom{k+n-1}{k}.$$

Bemerkung. Die Mengen $\Omega_{\mathrm{I}}^{k,n}$ und $\Omega_{\mathrm{II}}^{k,n}$ sind, wenn jede Kugel gleich wahrscheinlich ist, in natürlicher Weise LAPLACE-Räume für das jeweilige Zufallsexperiment. Auch auf $\Omega_{\mathrm{III}}^{k,n}$ beschreiben LAPLACE-Wahrscheinlichkeiten in diesem Fall das Zufallsexperiment, da die Urbilder unter der Abbildung φ die gleiche Kardinalität besitzen, siehe (1.2). Anschaulich zieht man die Kugeln zunächst in Reihenfolge in $\Omega_{\mathrm{II}}^{k,n}$, und vergisst dann die Reihenfolge mit φ. Dagegen ist $\Omega_{\mathrm{IV}}^{k,n}$ kein LAPLACE-Raum für das Zufallsexperiment, denn die Urbilder unter der Abbildung $\tilde\varphi:\Omega_{\mathrm{I}}^{k,n}\to\Omega_{\mathrm{IV}}^{k,n}$, die das übergebene Tupel ordnet und bei der somit die ursprüngliche Reihenfolge vergessen wird, sind nicht gleichmächtig: Beispielsweise gilt für $n=k=2$

$$\tilde\varphi^{-1}(1,1)=\{(1,1)\},\quad\tilde\varphi^{-1}(1,2)=\{(1,2),(2,1)\},\quad\tilde\varphi^{-1}(2,2)=\{(2,2)\}.$$

Bemerkung (Murmelmodell). Äquivalentes Alternativszenario: Verteile k Murmeln auf n Zellen, die von $1,\ldots,n$ nummeriert sind. Dann entspricht die Nummer der Zelle im Urnenmodell der Nummer der Kugel und die Nummer der Murmel der Nummer der Ziehung. Die Mengen können den einzelnen Varianten dann nach Tabelle 1.2 zugeordnet werden.

1.4 Stimmzettelproblem

Problem 1.5. *Es gibt eine Wahl zwischen den Kandidaten A und B. Am Ende hat Kandidat A a Stimmen und Kandidat B b Stimmen, dabei gilt $a > b$. Die Stimmen werden sukzessive ausgezählt. Wie wahrscheinlich ist es, dass A während der gesamten Auszählung vorne lag?*

Modell. Setze

$$n = a + b, \quad \omega_j = \begin{cases} 1, & \text{falls } j\text{-te Stimme für A,} \\ -1, & \text{falls } j\text{-te Stimme für B.} \end{cases}$$

Dann modellieren wir das Zufallsexperiment durch den LAPLACE-Raum

$$\begin{aligned} \Omega &= \Big\{\omega = (\omega_1,\ldots,\omega_n) \in \{-1,1\}^n : \sum_{j=1}^n 1_{\{\omega_j=1\}} = a\Big\} \\ &= \Big\{\omega = (\omega_1,\ldots,\omega_n) \in \{-1,1\}^n : \sum_{j=1}^n 1_{\{\omega_j=-1\}} = b\Big\} \end{aligned}$$

mit $\operatorname{card}\Omega = \binom{n}{a}$. Für $\omega = (\omega_1,\ldots,\omega_n) \in \Omega$, $1 \le k \le n$ setzen wir

$$s_k(\omega) = \omega_1 + \cdots + \omega_k,$$

die Stimmendifferenz von A zu B zum Zeitpunkt k. Das gesuchte Ereignis ist

$$D = \{\omega = (\omega_1,\ldots,\omega_n) \in \Omega \mid s_k(\omega) > 0 \text{ für alle } k = 1,\ldots,n\}.$$

Jetzt bleibt noch $\operatorname{card} D$ zu bestimmen. Dazu betrachten wir die Mengen

$$\begin{aligned} E &= \{\omega = (\omega_1,\ldots,\omega_n) \in \Omega \mid \omega_1 = -1\}, \\ F &= \{\omega = (\omega_1,\ldots,\omega_n) \in \Omega \mid \omega_1 = 1, s_k(\omega) \le 0 \text{ für ein } k = 2,\ldots,n\}. \end{aligned}$$

Dann ist $\Omega = D \cup E \cup F$ eine disjunkte Vereinigung mit $\operatorname{card}\Omega = \operatorname{card} D + \operatorname{card} E + \operatorname{card} F$. Zunächst ist $\operatorname{card} E = \binom{n-1}{a}$, da $\omega_2,\ldots,\omega_n$ frei sind.

Satz 1.6 (Spiegelungsprinzip)**.** *Beim Stimmzettelproblem wie oben beschrieben gilt*

$$\operatorname{card} E = \operatorname{card} F.$$

Beweis. Es sei $\tau : E \cup F \to \{2,\ldots,n\}$ mit $\tau(\omega) = \min\{k \in \{2,\ldots,n\} \mid s_k(\omega) = 0\}$. τ ist wohldefiniert, da irgendwo $s_k(\omega) \le 0$ gilt, aber $s_n(\omega) = a - b > 0$ und $s_{i+1}(\omega) = s_i(\omega) \pm 1$. Wir definieren nun eine Abbildung

$$\psi : E \to F, \quad (\psi(\omega))_i := \begin{cases} -\omega_i, & \text{falls } i \le \tau(\omega), \\ \omega_i, & \text{falls } i > \tau(\omega). \end{cases}$$

Sie ist wohldefiniert: Wegen $\tau(\omega) \ge 2$ ist stets $(\psi(\omega))_1 = -\omega_1 = 1$. Weiter ist

$$s_{\tau(\omega)}(\psi(\omega)) = s_{\tau(\omega)}(\omega) = 0$$

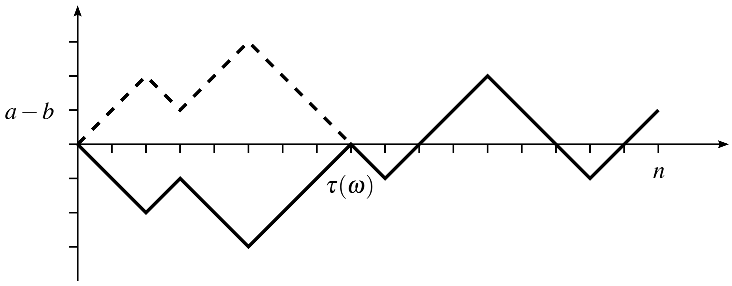


Abbildung 1.1: Mögliche Auszählung ω mit gespiegeltem Pfad $\psi(\omega)$.

und da $(\psi(\omega))_i = \omega_i$, $i > \tau(\omega)$, gilt $(\psi(\omega)) \in F$. Die Umkehrabbildung von ψ lässt sich mit derselben Formel definieren. Um nachzuweisen, dass es sich um die Umkehrabbildung handelt, muss man nur $\tau(\psi(\omega)) = \tau(\omega)$ beachten. Also ist ψ bijektiv und es gilt $\operatorname{card} F = \operatorname{card} E = \binom{n-1}{a}$. ■

Mit diesem Satz können wir nun die Kardinalität und die Wahrscheinlichkeit von D bestimmen und damit das Problem lösen:

$$\operatorname{card} D = \operatorname{card} \Omega - \operatorname{card} E - \operatorname{card} F = \operatorname{card} \Omega - 2 \cdot \operatorname{card} E = \binom{n}{a} - 2 \cdot \binom{n-1}{a},$$
$$\mathbb{P}(D) = \frac{\operatorname{card} D}{\operatorname{card} \Omega} = 1 - 2 \cdot \frac{\binom{n-1}{a}}{\binom{n}{a}} = 1 - 2 \cdot \frac{n-a}{n} = \frac{2a}{n} - 1.$$

2 Wahrscheinlichkeitsräume

2.1 Maßproblem und Wahrscheinlichkeitsräume

Sei Ω eine nicht-leere Menge und $\mathcal{P}(\Omega) := \{A \mid A \subseteq \Omega\}$ die *Potenzmenge* von Ω. Ω heißt *diskrete Menge*, wenn Ω endlich oder abzählbar unendlich ist. Mit einer Abbildung $\mathbb{P} : \mathcal{P}(\Omega) \to [0,1]$ die

1. (Normiertheit) $\quad \mathbb{P}(\Omega) = 1$,
2. (σ-Additivität) $\quad$ Für paarweise disjunkte Mengen $A_1, A_2, A_3, \ldots \in \mathcal{P}(\Omega)$ gilt

$$\mathbb{P}\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mathbb{P}(A_k), \tag{2.1}$$

erfüllt, erhält man einen *diskreten Wahrscheinlichkeitsraum*. $\mathbb{P}$ ist ein *diskretes Wahrscheinlichkeitsmaß* auf Ω. Die linke Seite in (2.1) hängt nicht von der Reihenfolge der A_k ab, die rechte Seite auch nicht (in Reihen mit positiven Gliedern darf nach dem Umordnungssatz beliebig umgeordnet werden). Um auch überabzählbare Ergebnisräume Ω zu betrachten, ist jedoch eine allgemeinere Definition von Wahrscheinlichkeitsräumen notwendig.

Problem 2.1 (Maßproblem)**.** Frage: *Kann über $[0,1)^n \subset \mathbb{R}^n$ ein Wahrscheinlichkeitsmaß definiert werden, so dass dieses wichtige wünschenswerte Eigenschaften besitzt?*

Satz 2.2. *Es existiert keine Abbildung $\mu : \mathcal{P}([0,1)^n) \to [0,1]$ mit folgenden Eigenschaften:*
1.) $\mu([0,1)^n) = 1$;
2.) Ist $A \subset [0,1)^n$ und $A + x \subset [0,1)^n$, $x \in \mathbb{R}^n$, so gilt $\mu(A) = \mu(A+x)$ (Translationsinvarianz);
3.) Sind $A_1, A_2, \ldots \subset [0,1)^n$ paarweise disjunkt, so gilt

$$\mu\Big(\bigcup_{n=1}^{\infty} A_n\Big) = \sum_{n=1}^{\infty} \mu(A_n) \quad (\sigma\text{-Additivität}).$$

Beweis. Für die Aussage des Satzes ist die Gültigkeit des Auswahlaxioms Voraussetzung. Wir betrachten eine sogenannte Vitali-Menge: Für $x, y \in [1/3, 2/3)^n$ definieren wir die Äquivalenzrelation

$$x \sim y : \quad \Leftrightarrow \quad x - y \in \mathbb{Q}^n.$$

Sei A ein Vertretersystem mit genau einem Repräsentanten in jeder Äquivalenzklasse. Dieses existiert nach dem Auswahlaxiom. Die Menge

$$B = \bigcup_{r \in \left[-\frac{1}{3}, \frac{1}{3}\right]^n \cap \mathbb{Q}^n} \{r\} + A$$

ist die Vereinigung abzählbar vieler disjunkter Teilmengen und ist eine solche Vitali-Menge. Es ist die Disjunktheit der Mengen $\{r\} + A, r \in [-1/3, 1/3]^n \cap \mathbb{Q}^n$ zu zeigen.

Seien $r_1 \neq r_2$ Elemente aus $[-1/3, 1/3]^n \cap \mathbb{Q}^n$ und $r_1 + a = r_2 + \tilde{a}$ mit $a, \tilde{a} \in A$, dann ist

$$(r_2 - r_1) = (a - \tilde{a}) \ \in \mathbb{Q}^n.$$

Daraus folgt aber $a = \tilde{a}$.

Es gilt außerdem $[1/3,2/3)^n \subseteq B \subseteq [0,1]^n$, da für $x \in [1/3,2/3)^n$ ein $y \in A$ existiert, so dass $x - y \in \mathbb{Q}^n$. Dann folgt einerseits

$$\mu\left([\tfrac{1}{3},\tfrac{2}{3})^n\right) = 3^{-n} \tag{2.2a}$$

aus 1.), 2.) und 3.), da wir den Einheitsquader als Vereinigung von 3^n disjunkten, zu $[1/3,2/3)^n$ kongruenten Quadern darstellen können. Andererseits folgt mit 2.) und 3.):

$$3^{-n} \leq \mu(B) = \sum_{r\in[-1/3,1/3]^n\cap\mathbb{Q}^n} \mu\left((A+\{r\})\right) = \sum_{r\in[-1/3,1/3]^n\cap\mathbb{Q}^n} \mu(A) = \begin{cases} 0, & \text{falls } \mu(A) = 0, \\ \infty, & \text{falls } \mu(A) > 0. \end{cases} \tag{2.2b}$$

Damit führen (2.2a) und (2.2b) zum Widerspruch. ■

Die oben gestellte Frage muss also mit „Nein“ beantwortet werden. Dadurch wird motiviert, möglichst große geeignete Mengensysteme zu finden, deren Elemente wir eine Wahrscheinlichkeit zuordnen können. Die nicht anschaulichen Vitali-Mengen müssen wir aus solchen Mengensystem ausschließen. Ein ähnliches, negatives Resultat über die Nicht-Existenz von einem Wahrscheinlichkeitsmaß mit gewünschten Eigenschaften unter der Annahme der Gültigkeit des Auswahlaxioms gibt es für das Modell des unendlichen Münzwurfs mit $\Omega = \{0,1\}^{\mathbb{N}}$. Diese Resultate sind jeweils als Satz von Vitali bekannt. Der Satz oben motiviert genauso die Konstruktion des Lebesguemaßes zur Definition einer Volumenfunktion auf $\mathbb{R}^n$ in der Analysis. Ein weiteres negatives Resultat was die Notwendigkeit zeigt, Wahrscheinlichkeiten auf Mengensystemen zu konstruieren die echte Teilmengen der Potenzmenge sind, ist der Satz von Banach und Kuratowski (1929).

Um dieses Dilemma zu lösen, definiert man Wahrscheinlichkeitsmaße für überabzählbare Ergebnismengen Ω nicht mehr auf der gesamten Potenzmenge $\mathcal{P}(\Omega)$, sondern nur für ein System von Teilmengen. Dieses soll zwei wesentliche Eigenschaften haben:

1. Es soll grundlegende Mengen enthalten, denen in jedem Fall Wahrscheinlichkeiten zugeordnet werden sollen. Im Fall von $\Omega = \mathbb{R}$, oder $\Omega = [0,1)$, sind dies etwa Intervalle oder diskrete Teilmengen, im Fall von $\Omega = \{0,1\}^{\mathbb{N}}$ Zylindermengen der Form $\{i_1\} \times \ldots \times \{i_k\} \times \{0,1\} \times \{0,1\} \times \ldots$ für $k \in \mathbb{N}$, $i_1,\ldots,i_k \in \{0,1\}$, bei denen der Ausgang des Experimentes also an endlich vielen Koordinaten spezifiziert wird.

2. Es soll unter abzählbar vielen Mengenoperationen abgeschlossen sein.

Die Anforderung 2. führt auf folgende Definition.

Definition 2.3. *Sei Ω eine nicht-leere Menge. Ein System von Teilmengen $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ heißt **σ-Algebra** über Ω, falls gilt:*

- $\Omega \in \mathcal{A}$.
- $A \in \mathcal{A} \Rightarrow A^{\complement} \in \mathcal{A}$.
- $A_1, A_2, \ldots \in \mathcal{A} \Rightarrow \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$.

Bemerkung. Ist $\mathcal{A}$ eine σ-Algebra über Ω, dann sind auch $\emptyset = \Omega^{\complement} \in \mathcal{A}$, und sind $A_1, A_2, \ldots \in \mathcal{A}$, so folgen für $n \in \mathbb{N}$

$$A_1 \cup \ldots \cup A_n = A_1 \cup \ldots \cup A_n \cup \emptyset \cup \emptyset \ldots \in \mathcal{A},$$

sowie

$$\bigcap_{k\in\mathbb{N}} A_k = \Big(\bigcup_{k\in\mathbb{N}} A_k^{\complement}\Big)^{\complement} \in \mathcal{A}.$$

Ein Tupel $(\Omega, \mathcal{A})$ heißt *Messraum*, oder *messbarer Raum*.

Beispiel 2.4. a. $\mathcal{A}_1 = \mathcal{P}(\Omega)$ sowie $\mathcal{A}_2 = \{\emptyset, \Omega\}$ sind σ-Algebren, $\mathcal{A}_2$ nennt man die *triviale σ-Algebra*.
b. Eine diskrete Teilmenge $A \subseteq \Omega$ wird auch *abzählbar* genannt, also falls A endlich oder abzählbar unendlich ist. Es ist

$$\mathcal{A} := \left\{A \subseteq \Omega : \quad A \text{ abzählbar oder } A^{\complement} \text{ abzählbar}\right\}$$

eine σ-Algebra.
c. Ist $E \subseteq \Omega$ und $\mathcal{A}$ eine σ-Algebra über Ω, so ist

$$\mathcal{A}(E) = \{A \cap E, \quad A \in \mathcal{A}\}$$

eine σ-Algebra über E, die sogenannte *Spur σ-Algebra* von $\mathcal{A}$ in E. Ist $E \in \mathcal{A}$, so gilt

$$\mathcal{A}(E) = \{A \in \mathcal{A} : A \subseteq E\}.$$

Lemma 2.5. *Sei $\mathcal{E} \subseteq \mathcal{P}(\Omega)$. Es existiert genau eine σ-Algebra $\sigma(\mathcal{E})$ über Ω mit den Eigenschaften*
a. $\mathcal{E} \subseteq \sigma(\mathcal{E})$,
b. Ist $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine σ-Algebra mit $\mathcal{E} \subseteq \mathcal{A}$, dann ist auch $\sigma(\mathcal{E}) \subseteq \mathcal{A}$.

Beweis. Der Schnitt über beliebig viele σ-Algebren ist wieder eine σ-Algebra. Daher erfüllt offenbar

$$\sigma(\mathcal{E}) = \bigcap_{\mathcal{A} \subseteq \mathcal{P}(\Omega)\ \sigma\text{-Algebra}, \mathcal{E} \subseteq \mathcal{A}} \mathcal{A}$$

das Gewünschte. ∎

$\sigma(\mathcal{E})$ heißt die *von $\mathcal{E}$ erzeugte σ-Algebra*, $\mathcal{E}$ heißt der *Erzeuger* der σ-Algebra $\sigma(\mathcal{E})$. Die folgende Eigenschaft wird besonders für Erzeuger relevant sein.

Definition 2.6. *Ein Mengensystem $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ heißt* ***durchschnittsstabil*** *(kurz $\cap$-stabil), falls für $A, B \in \mathcal{E}$ gilt $A \cap B \in \mathcal{E}$.*

Ein $\cap$-stabiles Mengensystem ist abgeschlossen unter endlichen Durchschnitten.
Ein Wahrscheinlichkeitsmaß ist nun eine normierte, σ-additive Abbildung auf einer σ-Algebra.

Definition 2.7. *Sei Ω eine nicht-leere Menge und $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ eine σ-Algebra über Ω. Eine Abbildung*

$$\mathbb{P} : \mathcal{A} \to [0,1]$$

heißt ***Wahrscheinlichkeitsmaß*** *über $(\Omega, \mathcal{A})$, falls gelten*

*1. (*Normiertheit*) $\mathbb{P}(\Omega) = 1$.*

2. (σ-Additivität*) Für paarweise disjunkte Mengen $A_1, A_2, \ldots \in \mathcal{A}$ gilt*

$$\mathbb{P}\Big(\bigcup_{k=1}^{\infty} A_k\Big) = \sum_{k=1}^{\infty} \mathbb{P}(A_k).$$

Das Tripel $(\Omega, \mathcal{A}, \mathbb{P})$ heißt ein Wahrscheinlichkeitsraum.

Teilmengen $A \subseteq \Omega$, für welche $A \in \mathcal{A}$ gilt, werden nachfolgend auch als *messbar* (bzgl. $\mathbb{P}$) bezeichnet. Eine Abbildung $\mu : \mathcal{A} \to [0, \infty]$ heißt allgemeiner ein *Maß* (über $\mathcal{A}$), falls μ σ-additiv ist mit $\mu(\emptyset) = 0$. Ein Maß ist endlich, falls $\mu(\Omega) < \infty$.

Definition 2.8. *Ein Maß $\mu : \mathcal{A} \to [0, \infty]$ über dem Messraum $(\Omega, \mathcal{A})$ heißt* ***σ-endlich****, falls $E_n \in \mathcal{A}$ existieren mit $\mu(E_n) < \infty$ und $\bigcup_{n=1}^{\infty} E_n = \Omega$.*

Das in der Vertiefung Analysis eingeführte Lebesguemaß, die Verallgemeinerung des elementaren Volumens im $\mathbb{R}^n$, ist nicht endlich, aber σ-endlich.

Satz 2.9. *Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B, A_i \in \mathcal{A}$, so gelten:*

1. $\mathbb{P}(\emptyset) = 0$,
2. *endliche Additivität: Sind $A_1, \ldots, A_n \in \mathcal{A}$ paarweise disjunkt, so ist*
$$\mathbb{P}\Big(\bigcup_{k=1}^{n} A_k\Big) = \sum_{k=1}^{n} \mathbb{P}(A_k),$$
3. $\mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$,
4. *Subtraktivität: $A \subseteq B \Rightarrow \mathbb{P}(B \setminus A) = \mathbb{P}(B) - \mathbb{P}(A)$, insb. $\mathbb{P}(A) \leq \mathbb{P}(B)$,*
5. *im diskreten Wahrscheinlichkeitsraum gilt $\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$.*

Beweis.
1. Setze in (2.1) $A_k = \emptyset$, dann $\mathbb{P}(\emptyset) = \lim_{k \to \infty}(k \cdot \mathbb{P}(\emptyset))$, also $\mathbb{P}(\emptyset) = 0$.
2. Setze $A_{n+1} = A_{n+2} = \ldots = \emptyset$, dann klar mit 1. und (2.1).
3. $A \cup A^{\complement} = \Omega$, $A \cap A^{\complement} = \emptyset$, also folgt mit 2., dass $\mathbb{P}(A) + \mathbb{P}(A^{\complement}) = 1$.
4. $B = A \cup (B \setminus A)$ (da $A \subseteq B$) und $A \cap (B \setminus A) = \emptyset$, also mit 2. $\mathbb{P}(B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$.
5. direkt aus 2. falls A endlich bzw. (2.1), falls A abzählbar unendlich.

■

Bemerkung. 5. zeigt, dass ein diskretes Wahrscheinlichkeitsmaß durch
$$\mathrm{p} : \Omega \to [0,1], \qquad \mathrm{p}(\omega) = \mathbb{P}(\{\omega\}), \qquad \omega \in \Omega,$$
eindeutig bestimmt ist. Wir nennen p *Zähldichte*, alternativ gibt es den Begriff *Wahrscheinlichkeitsfunktion*. Ist umgekehrt $\mathrm{p} : \Omega \to [0,1]$ gegeben mit $\sum_{\omega \in \Omega} \mathrm{p}(\omega) = 1$, so ist $\mathbb{P}(A) := \sum_{\omega \in A} \mathrm{p}(\omega)$ ein Wahrscheinlichkeitsmaß mit p als Wahrscheinlichkeitsfunktion. In diskreten Wahrscheinlichkeitsräumen ist es nützlich, $\mathbb{P}$ durch p zu charakterisieren.

Beispiel 2.10. Ein fairer Würfel wird n-mal geworfen. Wir betrachten den LAPLACE-Raum mit
$$\Omega = \{1, \ldots, 6\}^n, \quad \operatorname{card} \Omega = 6^n$$

1. Wie wahrscheinlich ist es, mindestens eine 6 zu würfeln?
$$\begin{aligned} A &= \{\omega = (\omega_1, \ldots, \omega_n) \in \Omega \mid \text{mindestens ein } \omega_i = 6\}, \\ A^{\complement} &= \{\omega = (\omega_1, \ldots, \omega_n) \in \Omega \mid \text{alle } \omega_i \in \{1, \ldots, 5\}\}. \end{aligned}$$
Es gilt $\operatorname{card} A^{\complement} = 5^n$, somit $\mathbb{P}(A^{\complement}) = \frac{5^n}{6^n} = (\frac{5}{6})^n$ und $\mathbb{P}(A) = 1 - (\frac{5}{6})^n$.
2. Wie wahrscheinlich ist es, dass nie zweimal die gleiche Ziffer in Folge gewürfelt wird?
$$A = \{\omega \in \Omega \mid \omega_i \neq \omega_{i+1} \text{ für } i = 1, \ldots, n-1\}$$
Es gilt $\operatorname{card} A = 6 \cdot 5^{n-1}$, also $\mathbb{P}(A) = \frac{6 \cdot 5^{n-1}}{6^n} = (\frac{5}{6})^{n-1}$.

2.2 Einschluss-Ausschluss-Formel

Problem 2.11. *Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \mathcal{A}$ nicht notwendigerweise disjunkt. Berechne $\mathbb{P}(A_1 \cup \dots \cup A_n)$, falls die $\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})$ bekannt sind.*

Beispiel 2.12. Zunächst:

1. Für zwei messbare Ereignisse $A, B \subseteq \Omega$: $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$

   Beweis. Aus $\mathbb{P}(B) = \mathbb{P}(B \setminus A) + \mathbb{P}(A \cap B)$ folgt wegen $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B \setminus A)$ die obige Formel. ■

2. Für drei messbare Ereignisse $A_1, A_2, A_3 \subseteq \Omega$:

$$
\begin{aligned}
\mathbb{P}(A_1 \cup A_2 \cup A_3) &= \mathbb{P}(A_1) + \mathbb{P}(A_2 \cup A_3) - \mathbb{P}(\underbrace{A_1 \cap (A_2 \cup A_3)}_{(A_1 \cap A_2) \cup (A_1 \cap A_3)}) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) - \mathbb{P}(A_2 \cap A_3) \\
&\quad - (\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) - \mathbb{P}(A_1 \cap A_2 \cap A_3)) \\
&= \mathbb{P}(A_1) + \mathbb{P}(A_2) + \mathbb{P}(A_3) \\
&\quad - (\mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_1 \cap A_3) + \mathbb{P}(A_2 \cap A_3)) + \mathbb{P}(A_1 \cap A_2 \cap A_3)\,.
\end{aligned}
$$

Satz 2.13 (Inklusions-Exklusions-Formel)**.** *Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A_1, \dots, A_n \in \mathcal{A}$, so gilt*

$$
\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}).
$$

Beweis. Induktion ab $n = 2$ (bereits gezeigt). Angenommen, die Formel gilt für alle Vereinigungen von n Mengen. Es gilt, dass

$$
\mathbb{P}(A_1 \cup \dots \cup A_n \cup A_{n+1)} = \mathbb{P}(A_1 \cup \dots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}((A_1 \cup \dots \cup A_n) \cap A_{n+1})\,.
$$

Nach Induktionsvoraussetzung gilt

$$
\mathbb{P}(A_1 \cup \dots \cup A_n) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k})
$$

und wegen $(A_1 \cup \dots \cup A_n) \cap A_{n+1} = (A_1 \cap A_{n+1}) \cup \dots \cup (A_n \cap A_{n+1})$ mit $B_i = A_i \cap A_{n+1}$

$$
\begin{aligned}
\mathbb{P}((A_1 \cup \dots \cup A_n) \cap A_{n+1}) &= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{P}(B_{i_1} \cap \dots \cap B_{i_k}) \\
&= \sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k} \cap A_{n+1}).
\end{aligned}
$$

Es bleibt zu zeigen:

$$
\sum_{k=1}^{n} (-1)^{k-1} \sum_{1 \le i_1 < \dots < i_k \le n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) + \mathbb{P}(A_{n+1})
$$

$$-\sum_{k=1}^{n}(-1)^{k-1}\sum_{1\le i_1<\ldots<i_k\le n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}\cap A_{n+1})$$
$$\stackrel{!}{=}\sum_{k=1}^{n+1}(-1)^{k-1}\sum_{1\le i_1<\ldots<i_k\le n+1}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}).$$

Die erste Summe auf der linken Seite entspricht allen Termen auf der rechten Seite mit $i_k \le n$. Also bleibt zu zeigen:

$$\mathbb{P}(A_{n+1})-\sum_{k=1}^{n}(-1)^{k-1}\sum_{1\le i_1<\ldots<i_k\le n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}\cap A_{n+1})$$
$$\stackrel{!}{=}\sum_{k=1}^{n+1}(-1)^{k-1}\sum_{1\le i_1<\ldots<i_k=n+1}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}).$$

Das Glied der rechten Seite mit $k=1$ ist gerade $\mathbb{P}(A_{n+1})$, also bleibt:

$$-\sum_{k=1}^{n}(-1)^{k-1}\sum_{1\le i_1<\ldots<i_k\le n}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_k}\cap A_{n+1})$$
$$\stackrel{!}{=}\sum_{j=2}^{n+1}(-1)^{j-1}\sum_{1\le i_1<\ldots<i_j=n+1}\mathbb{P}(A_{i_1}\cap\cdots\cap A_{i_j}),$$

was mit $k=j-1$ klar ist. ■

Beispiel 2.14 (Rencontre-Problem)**.** Dieses Problem lässt sich folgendermaßen beschreiben:

1. Verteile n Briefe zufällig auf n Umschläge, sodass in jedem Umschlag ein Brief steckt. Mit welcher Wahrscheinlichkeit gelangt mindestens ein Brief in den richtigen Umschlag?
2. Mische $n=13$ Karten mit Ziffern von 1 bis 13. Decke diese Karten nacheinander auf. Mit welcher Wahrscheinlichkeit stimmt die Ziffer auf mindestens einer Karte mit der Ziehungsnummer überein?

Dabei betrachten wir insbesondere die Konvergenz für $n\to\infty$. Ist $\mathfrak{S}_n$ die Menge der Permutationen von $\{1,\ldots,n\}$ mit $\operatorname{card}\mathfrak{S}_n = n!$, so ist das gesuchte Ereignis

$$A=\{\pi\in\mathfrak{S}_n \mid \pi(i)=i \text{ für mindestens ein } i\in\{1,\ldots,n\}\}.$$

Es stellt die Menge der Permutationen mit mindestens einem Fixpunkt dar. Betrachten wir nun für $j=1,\ldots,n$ die Mengen der Permutationen mit Fixpunkt in j

$$A_j=\{\pi\in\mathfrak{S}_n \mid \pi(j)=j\},$$

so gilt $A=A_1\cup\cdots\cup A_n$. Für $1\le j_1<\ldots<j_k\le n$ ist

$$A_{j_1}\cap\cdots\cap A_{j_k}=\{\pi\in\mathfrak{S}_n \mid \pi(j_i)=j_i \text{ für } i=1,\ldots,k\}.$$

Auf den übrigen $(n-k)$ Positionen kann beliebig permutiert werden, daher gilt

$$\operatorname{card}(A_{j_1}\cap\cdots\cap A_{j_k})=(n-k)!.$$

Da wir uns in einem LAPLACE-Raum befinden, gilt also

$$\mathbb{P}(A_{j_1}\cap\cdots\cap A_{j_k})=\frac{(n-k)!}{n!}.$$

Mit der Einschluss-Ausschlussformel gilt somit

$$\mathbb{P}(A) = \sum_{k=1}^{n} (-1)^{k-1} \underbrace{\sum_{1 \leq j_1 < \ldots < j_k \leq n}}_{\binom{n}{k} \text{ Summanden}} \frac{(n-k)!}{n!} = \sum_{k=1}^{n} (-1)^{k-1} \frac{(n-k)!}{n!} \frac{n!}{k!(n-k)!}$$

$$= \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!} = -\sum_{k=1}^{n} \frac{(-1)^k}{k!} = 1 - \sum_{k=0}^{n} \frac{(-1)^k}{k!} \overset{n\to\infty}{\to} 1 - \frac{1}{e}$$

mit der EULERschen Zahl e mit

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

Betrachten wir nun für $k = 0, \ldots, n$ die Wahrscheinlichkeit, dass genau k Fixpunkte auftreten. Die Wahrscheinlichkeiten für mindestens einen Fixpunkt bzw. keinen Fixpunkt sind

$$p_n = \sum_{k=1}^{n} \frac{(-1)^{k-1}}{k!} \text{ bzw. } 1 - p_n = \sum_{k=0}^{n} \frac{(-1)^k}{k!}.$$

Es gibt somit $a_n = n!(1-p_n)$ Permutationen ohne Fixpunkt. Um die Anzahl der Permutationen $a_{k,n}$ aus n Elementen mit genau k Fixpunkten zu bestimmen, wählen wir mit $\binom{n}{k}$ Möglichkeiten k Stellen für Fixpunkte aus und permutieren die restlichen $(n-k)$ Stellen ohne Fixpunkte mit a_{n-k} Möglichkeiten frei. Wir erhalten also

$$a_{k,n} = \binom{n}{k} \cdot a_{n-k} = \binom{n}{k}(n-k)! \sum_{j=0}^{n-k} \frac{(-1)^j}{j!}$$

Permutationen aus n Elementen mit genau k Fixpunkten. Die Wahrscheinlichkeit ist also

$$\mathbb{P}(\text{genau } k \text{ Fixpunkte in } n) = \frac{a_{k,n}}{n!} = \frac{\binom{n}{k}(n-k)!}{n!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} = \frac{1}{k!} \sum_{j=0}^{n-k} \frac{(-1)^j}{j!} \overset{n\to\infty}{\to} e^{-1} \frac{1}{k!}.$$

2.3 Bedingte Wahrscheinlichkeiten

Problem 2.15. *Im Ergebnisraum Ω ist die Wahrscheinlichkeit eines Ereignisses A gesucht. Dabei ist schon bekannt, dass B eingetreten ist. Wie wahrscheinlich ist A, gegeben dass B eingetreten ist? Gesucht ist die bedingte Wahrscheinlichkeit von A gegeben B, geschrieben $\mathbb{P}(A \mid B)$.*

Beispiel 2.16. Sei Ω ein LAPLACE-Raum und seien $A, B \subseteq \Omega$ mit $B \neq \emptyset$ Ereignisse. Angenommen, es sei schon bekannt, dass B eingetreten ist. Wie wahrscheinlich ist dann, dass A eintritt? Da bereits bekannt ist, dass B eingetreten ist, sind für das Eintreten von A nur noch die Ergebnisse in $A \cap B$ möglich. Insgesamt sind noch alle Ergebnisse in B möglich, und da ursprünglich jedes Ergebnis in B gleich wahrscheinlich war, muss dies hier immer noch so sein. Daher muss gelten:

$$\mathbb{P}(A \mid B) = \frac{\operatorname{card}(A \cap B)}{\operatorname{card} B}$$

und somit

$$\mathbb{P}(A \mid B) = \frac{\frac{\operatorname{card}(A\cap B)}{\operatorname{card}\Omega}}{\frac{\operatorname{card} B}{\operatorname{card}\Omega}} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}.$$

Bemerkung (Motivation über relative Häufigkeiten). Seien A, B Ereignisse. Wir führen nun ein Zufallsexperiment n-mal durch. Der empirische Gewissheitsgrad von A unter der Bedingung B ist

$$r_n(A \mid B) = \frac{\text{Anzahl der Versuche, in denen } A \text{ und } B \text{ eintreten}}{\text{Anzahl der Versuche, in denen } B \text{ eintritt}}$$

$$= \frac{n \cdot r_n(A \cap B)}{n \cdot r_n(B)} = \frac{r_n(A \cap B)}{r_n(B)} \overset{n\to\infty}{\to} \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)},$$

wobei die Grenzwertbildung dem empirischen Gesetz der großen Zahlen folgt.

Definition 2.17. *Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A, B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$, so heißt*

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

***bedingte Wahrscheinlichkeit** von A unter der Bedingung B.*

Bemerkung. Diese ist nur für $\mathbb{P}(B) > 0$ definiert.
Für Elementarereignisse $\omega \in \Omega$ in einem diskreten Wahrscheinlichkeitsraum ist

$$\mathbb{P}(\{\omega\} \mid B) = \begin{cases} \frac{\mathbb{P}(\{\omega\})}{\mathbb{P}(B)}, & \text{falls } \omega \in B, \\ 0, & \text{sonst.} \end{cases}$$

Lemma 2.18. *Es ist $\tilde{\mathbb{P}}_B := \mathbb{P}(\cdot \mid B)$ wieder ein Wahrscheinlichkeitsmaß auf $(\Omega, \mathcal{A})$, also $(\Omega, \mathcal{A}, \tilde{\mathbb{P}}_B)$ ein Wahrscheinlichkeitsraum.*

Beweis. Die Normiertheit folgt, da

$$\tilde{\mathbb{P}}_B(\Omega) = \mathbb{P}(\Omega \mid B) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1\,.$$

Für disjunkte $(A_i)_{i \geq 1}$ folgt

$$\tilde{\mathbb{P}}_B\Big(\bigcup_{i\geq 1} A_i\Big) = \frac{\mathbb{P}\Big(\big(\bigcup_i A_i\big) \cap B\Big)}{\mathbb{P}(B)} = \frac{\mathbb{P}\Big(\bigcup_i (A_i \cap B)\Big)}{\mathbb{P}(B)} = \sum_i \frac{\mathbb{P}(A_i \cap B)}{\mathbb{P}(B)} = \sum_i \tilde{\mathbb{P}}_B(A_i)\,.$$

∎

Satz 2.19 (Satz von der totalen Wahrscheinlichkeit). *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $B_1, \ldots, B_n$ eine disjunkte Zerlegung von Ω, d.h. $\bigcup_{k=1}^n B_k = \Omega$, $B_i \in \mathcal{A}$ für jedes i, und $B_i \cap B_j = \emptyset$ für alle $i \neq j$. Weiterhin sei $\mathbb{P}(B_j) > 0$, für alle j. Dann gilt für alle $A \in \mathcal{A}$, dass*

$$\mathbb{P}(A) = \sum_{j=1}^n \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j).$$

Beweis. Da für $j = 1, \ldots, n$ auch die $(A \cap B_j)$ paarweise disjunkt sind, gilt

$$\begin{aligned} \sum_{j=1}^n \mathbb{P}(A \mid B_j) \cdot \mathbb{P}(B_j) &= \sum_{j=1}^n \frac{\mathbb{P}(A \cap B_j)}{\mathbb{P}(B_j)} \cdot \mathbb{P}(B_j) = \sum_{j=1}^n \mathbb{P}(A \cap B_j) \\ &= \mathbb{P}\big(\bigcup_{j=1}^n (A \cap B_j)\big) = \mathbb{P}\big(A \cap \bigcup_{j=1}^n B_j\big) = \mathbb{P}(A \cap \Omega) = \mathbb{P}(A). \end{aligned}$$

∎

Lemma 2.20. *Sind $\mathbb{P}$ und $\tilde{\mathbb{P}}_B$ wie oben, so gilt für $A \in \mathcal{A}$, dass*

$$\tilde{\mathbb{P}}_B(A) > 0 \Leftrightarrow \mathbb{P}(A \cap B) > 0, \tag{2.3}$$

$$\tilde{\mathbb{P}}_B(C \mid A) = \mathbb{P}(C \mid A \cap B). \tag{2.4}$$

Beweis. 1. Beweis für (2.3):

$$\tilde{\mathbb{P}}_B(A) > 0 \Leftrightarrow \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(B)} > 0 \Leftrightarrow \mathbb{P}(A \cap B) > 0.$$

2. Beweis für (2.4):

$$\tilde{\mathbb{P}}_B(C \mid A) = \frac{\tilde{\mathbb{P}}_B(A \cap C)}{\tilde{\mathbb{P}}_B(A)} = \frac{\frac{\mathbb{P}(B \cap C \cap A)}{\mathbb{P}(B)}}{\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}} = \frac{\mathbb{P}(B \cap C \cap A)}{\mathbb{P}(A \cap B)} = \mathbb{P}(C \mid A \cap B).$$ ∎

Bemerkung (Formel von der totalen Wahrscheinlichkeit für $\tilde{\mathbb{P}}_B$). Ist $C_1, \ldots, C_n$ eine disjunkte Zerlegung von Ω mit $\tilde{\mathbb{P}}_B(C_j) > 0$ ($\Leftrightarrow \mathbb{P}(B \cap C_j) > 0$), so gilt

$$\tilde{\mathbb{P}}_B(A) = \sum_{j=1}^{n} \tilde{\mathbb{P}}_B(A \mid C_j) \cdot \tilde{\mathbb{P}}_B(C_j)$$

oder nach (2.4) äquivalent

$$\mathbb{P}(A \mid B) = \sum_{j=1}^{n} \mathbb{P}(A \mid C_j \cap B) \cdot \mathbb{P}(C_j \mid B). \tag{2.5}$$

Beispiel 2.21 (SIMPSON-Paradoxon)**.** Betrachten wir eine Universität mit zwei zulassungsbeschränkten Fächern und den Bewerberdaten aus Tabelle 2.1.

Tabelle 2.1: Verteilung der Zulassungen auf Fächer und Geschlechter

	Frauen			Männer		
	Bewerber	zugelassen	Anteil	Bewerber	zugelassen	Anteil
Fach 1	900	720	80%	200	180	90%
Fach 2	100	20	20%	800	240	30%
gesamt	1000	740	74%	1000	420	42%

Zwar ist die Zulassungsquote bei Männern in jedem Fach höher als bei Frauen, aber insgesamt ist die Quote bei Frauen deutlich höher. Nach den Gesamtzahlen könnte der Eindruck entstehen, Männer wären diskriminiert, obwohl in den einzelnen Fächern der gegenteilige Eindruck entsteht. Das liegt an der unterschiedlichen Verteilung der Bewerbungen über die Fächer. Seien

B: Person männlich, A: Person zugelassen,

C_1: Bewerbung in Fach 1, C_2: Bewerbung in Fach 2.

Dann ist einerseits

$$\mathbb{P}(A \mid B) < \mathbb{P}(A \mid B^{\complement}) \quad \text{(für Frauen besser)},$$

aber andererseits für $j = 1, 2$

$$\mathbb{P}(A \mid B \cap C_j) > \mathbb{P}(A \mid B^{\complement} \cap C_j) \quad \text{(für Männer besser)}.$$

Die Gesamtquoten setzen sich nach (2.5) wie folgt zusammen (Gewichtungen jeweils kursiv):

Frauen: $0.74 = 0.8 \cdot \mathit{0.9} + 0.2 \cdot \mathit{0.1}$

Männer: $\quad 0.42 = 0.9 \cdot \mathit{0.2} + 0.3 \cdot \mathit{0.8}$

Wir sehen, dass die niedrigere Quote bei den Männern stärker gewichtet wird.

Satz 2.22 (**Bayes-Formel**)**.** *Sind $A, B \in \mathcal{A}$ mit $\mathbb{P}(A) > 0$ und $\mathbb{P}(B) > 0$, so ist*

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

Beweis. Es gilt

$$\frac{\mathbb{P}(A \mid B) \cdot \mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \cdot \mathbb{P}(B)}{\mathbb{P}(A)} = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B \mid A). \qquad \blacksquare$$

Bemerkung. Ist $B = B_i$ Teil einer Zerlegung $B_1, \ldots, B_n$ von Ω, so ergibt sich, wenn rechts im Nenner für $\mathbb{P}(A)$ die Formel von der totalen Wahrscheinlichkeit angewendet wird, dass

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i) \cdot \mathbb{P}(B_i)}{\sum_{k=1}^n \mathbb{P}(A \mid B_k) \cdot \mathbb{P}(B_k)}.$$

Satz 2.23 (Multiplikationsregel)**.** *Sind $A_1, \ldots, A_n$ Ereignisse, $A_i \in \mathcal{A}$ für jedes i, und $\mathbb{P}(A_1 \cap \cdots \cap A_{n-1}) \neq 0$, so ist*

$$\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \cdots \cdot \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}).$$

Beweis. Unter den angegebenen Bedingungen sind alle Terme definiert, also gilt

$$\begin{aligned}\mathbb{P}(A_1) \cdot \mathbb{P}(A_2 \mid A_1) \cdot \cdots \cdot \mathbb{P}(A_n \mid A_1 \cap \cdots \cap A_{n-1}) &= \mathbb{P}(A_1) \cdot \prod_{k=2}^n \mathbb{P}\Big(A_k \Big| \bigcap_{i=1}^{k-1} A_i\Big) \\ = \mathbb{P}(A_1) \cdot \prod_{k=2}^n \frac{\mathbb{P}(A_k \cap \bigcap_{i=1}^{k-1} A_i)}{\mathbb{P}(\bigcap_{i=1}^{k-1} A_i)} = \frac{\prod_{k=1}^n \mathbb{P}(\bigcap_{i=1}^k A_i)}{\prod_{k=1}^{n-1} \mathbb{P}(\bigcap_{i=1}^k A_i)} &= \mathbb{P}\Big(\bigcap_{i=1}^n A_i\Big) = \mathbb{P}(A_1 \cap \cdots \cap A_n). \qquad \blacksquare\end{aligned}$$

Beispiel 2.24 (Erkennung von Krankheiten)**.** Bei einem Test auf eine Krankheit können zwei Fehler auftreten.

- falsch positiv: Der Test zeigt eine Krankheit an, obwohl die Person nicht krank ist.
- falsch negativ: Der Test zeigt keine Krankheit an, obwohl die Person krank ist.

Dabei treten zwei Wahrscheinlichkeiten auf:

- Sensitivität: Wahrscheinlichkeit p_{se}, mit der eine kranke Person als krank erkannt wird.
- Spezifität: Wahrscheinlichkeit p_{sp}, mit der eine gesunde Person als gesund erkannt wird.

Für Standardtests liegen Schätzwerte für Sensitivität und Spezifität aufgrund großer Studien vor. Beim ELISA-Test auf HIV-Antikörper gilt etwa $p_{\text{se}} = p_{\text{sp}} = 0{,}998$.

Angenommen, eine Person macht den Test und dieser zeigt ein positives Ergebnis. Wie wahrscheinlich ist es dann, dass die Person auch wirklich krank ist? Dabei sei die sogenannte *a-priori-Wahrscheinlichkeit* q für eine Krankheit gegeben, also etwa der Anteil der Kranken in der Bevölkerung oder einer Risikogruppe. Diese ist schwer festzulegen, falls etwa aus speziellem Anlass getestet wird.

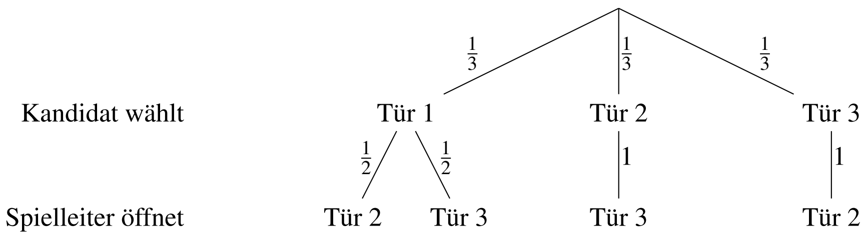


Abbildung 2.1: Entscheidungsbaum bis zur Wechselentscheidung

Modell. Der Ergebnisraum ist

$$\Omega = \{0,1\}^2 = \{(\omega_1,\omega_2) \mid \omega_1,\omega_2 \in \{0,1\}\},$$

wobei gilt

$$\omega_1 = \begin{cases} 0, & \text{falls Patient gesund,} \\ 1, & \text{sonst,} \end{cases} \qquad \omega_2 = \begin{cases} 0, & \text{falls Test negativ,} \\ 1, & \text{falls Test positiv.} \end{cases}$$

Weiterhin definieren wir die Ereignisse

$$K = \{1\} \times \{0,1\} = \{(1,0),(1,1)\}, \quad P = \{0,1\} \times \{1\} = \{(0,1),(1,1)\},$$

die anzeigen, dass der Patient krank ist bzw. dass der Test positiv ist. Wir kennen nun die Wahrscheinlichkeiten

$$\mathbb{P}(K) = q, \quad \mathbb{P}(P \mid K) = p_{\mathrm{se}}, \quad \mathbb{P}(P^{\complement} \mid K^{\complement}) = p_{\mathrm{sp}}.$$

Daraus können wir berechnen

$$\mathbb{P}(P) = \mathbb{P}(P \mid K)\cdot\mathbb{P}(K) + \mathbb{P}(P \mid K^{\complement})\cdot\mathbb{P}(K^{\complement}) = p_{\mathrm{se}}\cdot q + (1-p_{\mathrm{sp}})\cdot(1-q).$$

Nach der BAYES-Formel erhalten wir also

$$\mathbb{P}(K \mid P) = \frac{\mathbb{P}(P \mid K)\cdot\mathbb{P}(K)}{\mathbb{P}(P)} = \frac{p_{\mathrm{se}}\cdot q}{p_{\mathrm{se}}\cdot q + (1-p_{\mathrm{sp}})\cdot(1-q)}.$$

Beim ELISA-Test erhalten wir also für $q = 0{,}001$ den Wert $\mathbb{P}(K \mid P) = \frac{1}{3}$.

Beispiel 2.25 (Ziegenproblem)**.** In der amerikanischen Spielshow “Let’s make a deal” kann ein Kandidat als Hauptpreis ein Auto gewinnen. Im Studio sind drei verschiedene Türen, wobei hinter einer der Hauptpreis und hinter den anderen Ziegen stehen. Der Kandidat wählt eine Tür, diese bleibt verschlossen. Dann öffnet der Spielleiter (weiß, wo der Hauptpreis ist) eine der anderen Türen. Hinter dieser steht immer eine Ziege. Dann darf der Kandidat die Tür wechseln. Er erhält den Preis, der hinter der zuletzt gewählten Tür steht. In Abbildung 2.1 werden die einzelnen Entscheidungen genauer dargestellt. Dabei nehmen wir ohne Einschränkung an, dass das Auto hinter Tür 1 ist.

1. Behalte die gewählte Tür.

   Tür 1 bleibt genau dann gewählt, wenn sie im ersten Schritt gewählt wurde. Die Gewinnwahrscheinlichkeit ist also $1/3$.

2. Wechsle die Tür.

   Der Kandidat gewinnt, wenn er im ersten Schritt Tür 2 oder Tür 3 gewählt hatte, da er dann stets zu Tür 1 wechselt. Die Gewinnwahrscheinlichkeit ist also $2/3$.

Es ist also ratsam, die Tür zu wechseln.

2.4 Unabhängigkeit

Bemerkung. Gilt für messbare Ereignisse $A, B \subseteq \Omega$ mit $\mathbb{P}(B) > 0$

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

so hängt die Wahrscheinlichkeit von A nicht davon ab, ob der Eintritt von B bekannt ist oder nicht.

Definition 2.26. *Zwei messbare Ereignisse $A, B \subseteq \Omega$ heißen **unabhängig**, falls*

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

Bemerkung.
1. Falls $\mathbb{P}(B) > 0$, so ist das äquivalent zu $\mathbb{P}(A \mid B) = \mathbb{P}(A)$; falls $\mathbb{P}(A) > 0$, so ist das äquivalent zu $\mathbb{P}(B \mid A) = \mathbb{P}(B)$.
2. Wird zweimal eine faire Münze geworfen, und ist A das Ereignis, dass im ersten Wurf Zahl fällt, und entsprechend B das Ereignis, dass im zweiten Wurf Zahl fällt, so sind diese Ereignisse unabhängig. Dies ist auch unmittelbar plausibel, da ja auch die Experimente nichts miteinander zu tun haben. Dies ist jedoch nicht der Kern des Begriffs der *stochastischen* Unabhängigkeit von zwei Ereignissen.
   Beispiel: Bei einem zweifachen Würfelwurf, $\Omega = \{1, \ldots, 6\}^2$, definieren wir die Ereignisse

   $$A = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 + \omega_2 \text{ ungerade}\}, \quad B = \{(\omega_1, \omega_2) \in \Omega \mid \omega_1 \text{ gerade}\}.$$

   Dann gilt

   $$\mathbb{P}(A) = \mathbb{P}(B) = \frac{1}{2}, \quad \mathbb{P}(A \cap B) = \mathbb{P}(\omega_1 \text{ gerade}, \omega_2 \text{ ungerade}) = \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(B),$$

   also sind A und B wahrscheinlichkeitstheoretisch unabhängig, obwohl die Ereignisse ihr gegenseitiges Eintreten beeinflussen. Unabhängigkeit im wahrscheinlichkeitstheoretischen Sinn ist also nicht äquivalent zu keiner physischen Beeinflussung bei der Durchführung des Zufallsexperiments.
3. Unabhängigkeit ist nicht Disjunktheit.

   Im Gegenteil: Ist $A \cap B = \emptyset$, so gilt $0 = \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$ nur, wenn $\mathbb{P}(A) = 0$ oder $\mathbb{P}(B) = 0$ gilt.
4. A ist unabhängig von sich selbst genau dann, wenn $\mathbb{P}(A)^2 = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(A) \in \{0, 1\}$.

Bemerkung (Erweiterung auf mehr als zwei Ereignisse). Seien nun $A_1, \ldots, A_n$ Ereignisse mit $\mathbb{P}(A_1 \cap \ldots \cap A_n) > 0$. Diese sollen unabhängig heißen, wenn durch Bedingen auf eine Auswahl der anderen Ereignisse die Wahrscheinlichkeiten nicht verändert werden. Formal bedeutet dies, dass für alle $I \subseteq \{1, \ldots, n\}$, $I \neq \emptyset$ mit $\operatorname{card} I \leq n - 1$ und $j \in \{1, \ldots, n\} \setminus I$ gilt

$$\mathbb{P}(A_j) = \mathbb{P}\Big(A_j \,\Big|\, \bigcap_{i \in I} A_i\Big).$$

Ein einfaches Argument mit Induktion über die Kardinalität von I zeigt, dass dies im Fall $\mathbb{P}(A_1 \cap \ldots \cap A_n) > 0$ äquivalent zu folgender Definition ist.

Definition 2.27. *Messbare Ereignisse $A_1, \ldots, A_n \subseteq \Omega$ heißen **unabhängig**, falls $\forall I \subseteq \{1, \ldots, n\}$ gilt*

$$\mathbb{P}\Big(\bigcap_{i \in I} A_i\Big) = \prod_{i \in I} \mathbb{P}(A_i).$$

Wir können ohne Einschränkung $\operatorname{card} I \geq 2$ *setzen, d.h. es sind* $2^n - n - 1$ *Bedingungen.*
Falls für alle i, j *gilt* $\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j)$*, nennen wir die Ereignisse* ***paarweise unabhängig****.*

Bemerkung (Illustration für drei Ereignisse). Seien $A, B, C \subseteq \Omega$ messbare Ereignisse. Für die Unabhängigkeit von A, B, C werden neben

$$\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(B) \cdot \mathbb{P}(C) \tag{2.6}$$

noch

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B), \quad \mathbb{P}(A \cap C) = \mathbb{P}(A) \cdot \mathbb{P}(C), \quad \mathbb{P}(B \cap C) = \mathbb{P}(B) \cdot \mathbb{P}(C).$$

gefordert. Das folgende Beispiel zeigt, dass (2.6) allein nicht ausreicht.

Beispiel 2.28. Im LAPLACE-Raum mit dem Ereignisraum $\Omega = \{1, \ldots, 8\}$ seien die Ereignisse

$$A = B = \{1, \ldots, 4\}, \quad C = \{1, 5, 6, 7\}.$$

Dann gilt $\mathbb{P}(A) = \mathbb{P}(B) = \mathbb{P}(C) = \frac{1}{2}$ und $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\{1\}) = \frac{1}{8}$, also ist (2.6) erfüllt, aber

$$\begin{aligned}
\mathbb{P}(A \cap B) &= \frac{1}{2} \neq \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(B), \\
\mathbb{P}(A \cap C) &= \frac{1}{8} \neq \frac{1}{4} = \mathbb{P}(A) \cdot \mathbb{P}(C), \\
\mathbb{P}(B \cap C) &= \frac{1}{8} \neq \frac{1}{4} = \mathbb{P}(B) \cdot \mathbb{P}(C).
\end{aligned}$$

Lemma 2.29 (Unabhängigkeit und Komplementbildung). *Messbare Ereignisse* $A, B \subseteq \Omega$ *sind genau dann unabhängig, wenn* $A, B^\complement$ *unabhängig sind.*

Beweis. Für A und B unabhängig gilt

$$\begin{aligned}
\mathbb{P}(A \cap B^\complement) &= \mathbb{P}(A) - \mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A) \cdot \mathbb{P}(B) \\
&= \mathbb{P}(A) \cdot (1 - \mathbb{P}(B)) = \mathbb{P}(A) \cdot \mathbb{P}(B^\complement).
\end{aligned} \tag{2.7}$$

Die andere Beweisrichtung verläuft analog. ■

Satz 2.30 (Verallgemeinerung). *Seien* $A_1, \ldots, A_n \subseteq \Omega$ *messbare Ereignisse. Dann sind äquivalent:*

1. $A_1, \ldots, A_n$ *sind unabhängig.*
2. *Für alle* $I, J \subseteq \{1, \ldots, n\}$ *mit* $I \cap J = \emptyset$ *gilt*

$$\mathbb{P}\left(\left(\bigcap_{i \in I} A_i\right) \cap \left(\bigcap_{j \in J} A_j^\complement\right)\right) = \prod_{i \in I} \mathbb{P}(A_i) \cdot \prod_{j \in J} \mathbb{P}(A_j^\complement).$$

Bemerkung. Dabei gilt

$$\bigcap_{i \in \emptyset} A_i = \Omega, \quad \prod_{i \in \emptyset} \mathbb{P}(A_i) = 1.$$

Beweis.
1. Punkt 2. ist ein Spezialfall von Punkt 1. mit $J = \emptyset$.
2. Aus Punkt 1. folgt Punkt 2. per Induktion über $\operatorname{card} J$ analog zu (2.7). ■

2.5 Die Borelsche σ-Algebra

Sei Ω eine nicht-leere Menge. Wir erinnern an den Begriff der *Topologie* auf Ω aus der Analysis II. Eine Topologie ist ein System *offener Teilmengen*, $\mathcal{O} \subseteq \mathcal{P}(\Omega)$ mit $\Omega, \emptyset \in \mathcal{O}$ sowie:

T1. Ist $(U_i)_{i\in I} \subseteq \mathcal{O}$ eine beliebige Familie, so ist auch $\bigcup_{i\in I} U_i \in \mathcal{O}$;

T2. Sind $U_1, \dots, U_n \in \mathcal{O}$, so auch $U_1 \cap \dots \cap U_n \in \mathcal{O}$.

Dann heißt $(\Omega, \mathcal{O})$ ein *topologischer Raum*. Für einen topologischen Raum $(\Omega, \mathcal{O})$ heißt die von den offenen Mengen erzeugte σ-Algebra $\sigma(\mathcal{O})$ die *Borel-σ-Algebra*, für die wir $\mathcal{B}(\Omega)$ schreiben. Offenbar ist $\mathcal{O}$ ein $\cap$-stabiler Erzeuger von $\mathcal{B}(\Omega)$. Für uns sind hier in erster Linie *metrische Räume* von Interesse und jede Metrik induziert eine Topologie (Analysis II). Wir erinnern daran, dass eine Metrik $d(x,y)$ auf Ω eine Abbildung

$$d : \Omega \times \Omega \to [0, \infty),$$

ist, für die gilt:
1. $d(x,y) = d(y,x)$ für alle $x, y \in \Omega$, (*Symmetrie*);
2. $d(x,z) \leq d(x,y) + d(y,z)$ für alle $x,y,z \in \Omega$ (*Dreiecksungleichung*);
3. $d(x,y) = 0$, genau dann wenn $x = y$, (*Definitheit*).

Satz 2.31. *Sei $(\Omega, \mathcal{O})$ ein topologischer Raum und sei $\mathcal{F} \subseteq \mathcal{O}$ eine abzählbare Menge, so dass*

$$\forall\, U \in \mathcal{O} : \qquad U = \bigcup_{F\in\mathcal{F},\ F\subseteq U} F, \tag{2.8}$$

falls also jede offene Menge als Vereinigung von Mengen aus $\mathcal{F}$ dargestellt werden kann. Dann gilt $\mathcal{B}(\Omega) = \sigma(\mathcal{F})$.

Bemerkung: Für den Beweis (und nachfolgend) benutzen wir folgendes Prinzip: Sind $\mathcal{E}_1, \mathcal{E}_2 \subseteq \mathcal{P}(\Omega)$ und gilt $\mathcal{E}_1 \subseteq \sigma(\mathcal{E}_2)$, so folgt $\sigma(\mathcal{E}_1) \subseteq \sigma(\mathcal{E}_2)$. Ist darüber hinaus $\mathcal{E}_2 \subseteq \sigma(\mathcal{E}_1)$, so ist $\sigma(\mathcal{E}_1) = \sigma(\mathcal{E}_2)$.

Beweis von Satz 2.31. Offenbar ist $\mathcal{F} \subseteq \mathcal{O} \subseteq \mathcal{B}(\Omega)$, also $\sigma(\mathcal{F}) \subseteq \mathcal{B}(\Omega)$.
Für die Umkehrung müssen wir zeigen $\mathcal{O} \subseteq \sigma(\mathcal{F})$. Sei dazu $U \in \mathcal{O}$. In der Darstellung (2.8) ist die Vereinigung über höchstens abzählbar viele Mengen gebildet, da $\mathcal{F}$ abzählbar ist. Dies bedeutet aber $U \in \sigma(\mathcal{F})$, da $\sigma(\mathcal{F})$ abgeschlossen unter abzählbaren Vereinigungen ist. ■

Die Frage ist, wann eine solche abzählbare Menge $\mathcal{F}$ existiert. Eine Teilmenge $M \subseteq \Omega$ eines topologischen Raums heißt *dicht*, falls für alle offenen Mengen $U \in \mathcal{O}$ ein $x \in M$ existiert mit $x \in U$. Der topologische Raum $(\Omega, \mathcal{O})$ heißt *separabel*, falls Ω eine abzählbare, dichte Teilmenge enthält.

Beispiel. $\mathbb{R}^p$ ist separabel, denn $\mathbb{Q}^p$ liegt dicht.

Tatsächlich kann für einen separablen metrischen Raum (Ω, d), mit einer abzählbaren dichten Teilmenge $M \subset \Omega$, gezeigt werden, dass

$$\mathcal{F} = \{B_\varepsilon(x), \quad x \in M,\ \varepsilon > 0, \varepsilon \in \mathbb{Q}\}$$

stets eine solche abzählbare Menge wie im Satz darstellt. Für unser Standardbeispiel $\mathbb{R}^p$ nutzen wir typischerweise jedoch speziellere Erzeuger. Wir betrachten nun den $\mathbb{R}^p$. Auf einem endlich-dimensionalen Vektorraum sind bekanntlich alle Normen äquivalent, d.h. sind $\|\cdot\|_1$, $\|\cdot\|_2$ zwei Normen, so existieren $0 < c < C < \infty$ mit

$$c\,\|x\|_1 \leq \|x\|_2 \leq C\,\|x\|_1, \qquad \forall\, x \in \mathbb{R}^p.$$

Eine Norm $\|\cdot\|$ auf $\mathbb{R}^p$ induziert eine Metrik mittels $d(x,y) = \|x - y\|$, $x, y \in \mathbb{R}^p$. Die Äquivalenz der Normen impliziert, dass alle Metriken (aus einer Norm) dieselbe Topologie induzieren, die nach der euklidischen Norm auch als *euklidische Topologie* bezeichnet wird. Es gibt eine Vielzahl von abzählbaren

Mengen, welche die Borel-σ-Algebra für die euklidische Topologie erzeugen. Da bei überabzählbaren Ergebnisräumen Ω die Fälle $\Omega \subseteq \mathbb{R}^p$, vor allem für $p = 1$, die wichtigsten sind, bezeichnet *Borel-σ-Algebra* in der Literatur oft nur diesen spezifischen Fall, obwohl obige Definition allgemein zu verstehen ist. Wir bezeichnen die Borel-σ-Algebra (der euklidischen Topologie) des $\mathbb{R}^p$ nachfolgend kurz mit $\mathcal{B}^p$. Insbesondere verwenden wir später $\mathcal{B}$ als Kurznotation für $p = 1$. Wir geben einen der wichtigsten Erzeuger für $\mathcal{B}^p$ an. Für $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$, $\mathbf{a} = (a_1, \ldots, a_p)^T$, $\mathbf{b} = (b_1, \ldots, b_p)^T$ mit $a_i < b_i$, $i = 1, \ldots, p$, setzen wir

$$(\mathbf{a}, \mathbf{b}] = \left\{x = (x_1, \ldots, x_p)^T \in \mathbb{R}^p : \ a_j < x_j \leq b_j, \ j = 1, \ldots, p\right\}.$$

und analog $(\mathbf{a}, \mathbf{b}), [\mathbf{a}, \mathbf{b}], [\mathbf{a}, \mathbf{b})$.

Satz 2.32. *Die Mengensysteme*

$$\mathcal{J}^p = \left\{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^p\right\}, \qquad \mathcal{J}^p_{\mathbb{Q}} = \left\{(\boldsymbol{a}, \boldsymbol{b}] : \boldsymbol{a}, \boldsymbol{b} \in \mathbb{Q}^p\right\}$$

sind $\cap$-stabile Erzeuger von $\mathcal{B}^p$. ▪

Beweis. Sind $\mathbf{a}, \mathbf{b}, \tilde{\mathbf{a}}, \tilde{\mathbf{b}} \in \mathbb{R}^p$ (bzw. $\mathbb{Q}^p$), so sind

$$(\mathbf{a}, \mathbf{b}] \cap (\tilde{\mathbf{a}}, \tilde{\mathbf{b}}] = \big(\big(\max(a_1, \tilde{a}_1), \ldots, \max(a_p, \tilde{a}_p)\big), \big(\min(b_1, \tilde{b}_1), \ldots, \min(b_p, \tilde{b}_p)\big)\big] \in \mathcal{J}^p \quad \big(\in \mathcal{J}^p_{\mathbb{Q}}\big),$$

also gilt $\cap$-Stabilität. Offenbar ist $\left\{(\mathbf{a}, \mathbf{b}) : \mathbf{a}, \mathbf{b} \in \mathbb{Q}^p\right\}$ ein Erzeuger für $\mathcal{B}^p$ (es sind z. Bsp. offene Kugeln für die Maximum-Norm $\|x\|_{\max} = \max_{i=1,\ldots,p} |x_i|$ mit Radien $\in \mathbb{Q}$ enthalten). Die Behauptung, dass die halb-offenen Quader auch Erzeuger sind, ergibt sich dann mit

$$(\mathbf{a}, \mathbf{b}) = \bigcup_{n \geq 1} (\mathbf{a}, \mathbf{b} - 1/n] \, .$$

■

3 Zufallsvariablen und Verteilung

3.1 Erste Beispiele und messbare Abbildungen

Häufig betrachten wir Abbildungen $X : \Omega \to Y$, von unserem Ergebnisraum in eine andere Menge Y. Oft sind den Ergebnissen eines Zufallsexperimentes zum Beispiel reelle Zahlen zugeordnet, $X(\Omega) \subseteq \mathbb{R}$, etwa der Zahlenkombination im Lotto der Gewinn eines individuellen Spielers. Wir formalisieren dies durch den Begriff der *Zufallsvariablen*. Sei im Folgenden $(\Omega, \mathbb{P})$ zunächst ein diskreter oder sogar endlicher Wahrscheinlichkeitsraum.

Notation (Beschreibung von Ereignissen). Ist $X : \Omega \to Y$, $A \subseteq X(\Omega)$ und $x \in X(\Omega)$, so schreiben wir

$$\begin{aligned}
\{X \in A\} &:= X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}\,,\\
\{X = x\} &:= X^{-1}(\{x\}) = \{\omega \in \Omega \mid X(\omega) = x\}\,.
\end{aligned}$$

Für $X(\Omega) \subseteq \mathbb{R}$ schreiben wir analog

$$\{X \leq x\} := X^{-1}((-\infty, x]) = \{\omega \in \Omega \mid X(\omega) \leq x\}\,.$$

Beispiel 3.1 (Augensumme beim zweifachen Würfelwurf)**.** Wir modellieren den zweifachen Würfelwurf mit dem LAPLACE-Raum

$$\Omega = \Omega_{\mathrm{I}}^{2,6} = \{\omega = (\omega_1, \omega_2) \in \mathbb{N}^2 \mid 1 \leq \omega_1, \omega_2 \leq 6\} = \{1, \ldots, 6\}^2.$$

Dann lässt sich die Augensumme darstellen mit

$$X : \Omega \to \mathbb{R}, \quad X((\omega_1, \omega_2)) = \omega_1 + \omega_2.$$

Hier gilt z. B.

$$\begin{aligned}
\{X = 2\} &= \{(1,1)\}\\
\{X = 3\} &= \{(1,2),(2,1)\}\\
\{X = 4\} &= \{(1,3),(2,2),(3,1)\}\\
\{X = 5\} &= \{(1,4),(2,3),(3,2),(4,1)\}\\
\{X = 6\} &= \{(1,5),(2,4),(3,3),(4,2),(5,1)\}\\
\{X \leq 4\} &= \{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)\}\,.
\end{aligned}$$

Bemerkung (Kardinalität des Bildbereichs). Ist $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum und $X : \Omega \to X(\Omega)$ eine Abbildung, dann ist der Bildbereich

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$$

diskret, da $\operatorname{card} X(\Omega) \leq \operatorname{card} \Omega$ gelten muss.

Es ist naheliegend, dass wir Ereignissen dass $X(\omega)$ Werte in einer bestimmten Teilmenge von $X(\Omega)$ annimmt, Wahrscheinlichkeiten zuordnen wollen. So möchte der Lotto-Spieler wissen, mit welcher Wahrscheinlichkeit er welchen Gewinn erzielt. Da wir in Wahrscheinlichkeitsräumen nur messbaren Teilmengen von Ω Wahrscheinlichkeiten zuordnen, benötigen wir den folgenden Begriff *messbarer Abbildungen*.

Ist Ω eine nicht-leere Menge und $\mathcal{A}$ eine σ-Algebra über Ω, also $(\Omega, \mathcal{A})$ ein Messraum. $(Y, \mathcal{B})$ sei nachfolgend ebenfalls ein Messraum, also mit irgendeiner σ-Algebra $\mathcal{B}$ über Y.

Definition 3.2. *Eine Abbildung $f : \Omega \to Y$ heißt $(\mathcal{A} - \mathcal{B})$-**messbar** (kurz **messbar**), falls $f^{-1}(B) \in \mathcal{A}$ für alle $B \in \mathcal{B}$, also $f^{-1}(\mathcal{B}) \subseteq \mathcal{A}$.*

Bemerkung. Für jede Abbildung $f : \Omega \to Y$ vertauscht f^{-1} als Mengenabbildung mit den üblichen Mengenoperationen, also $f^{-1}(Y) = \Omega$, $f^{-1}(B^{\complement}) = \left(f^{-1}(B)\right)^{\complement}$,

$$f^{-1}\Big(\bigcup_n B_n\Big) = \Big(\bigcup_n f^{-1}(B_n)\Big), \qquad f^{-1}\Big(\bigcap_n B_n\Big) = \Big(\bigcap_n f^{-1}(B_n)\Big),$$

wobei $B, B_n \subseteq Y$. Zum Beispiel ist

$$f^{-1}(B^{\complement}) = \{\omega \in \Omega : \ f(\omega) \in B^{\complement}\} = \Omega \setminus \{\omega \in \Omega : \ f(\omega) \in B\} = \left(f^{-1}(B)\right)^{\complement},$$

da entweder $f(\omega) \in B$ oder $f(\omega) \in B^{\complement}$.

Bemerkung. Es folgt direkt:

1. Ist $\mathcal{B}$ eine σ-Algebra über Y, so ist $f^{-1}(\mathcal{B})$ eine σ-Algebra über Ω.
2. Ist $\mathcal{A}$ eine σ-Algebra über Ω, so ist $\{B \subseteq Y : \ f^{-1}(B) \in \mathcal{A}\}$ eine σ-Algebra über Y.

Nun könnte es manchmal schwierig sein, die Bedingung $f^{-1}(\mathcal{B}) \subseteq \mathcal{A}$ zu überprüfen. Dafür sind Erzeuger nützlich.

Satz 3.3. *Ist $f : \Omega \to Y$ eine Abbildung und $\mathcal{E}$ ein Erzeuger von $\mathcal{B}$, so ist $f^{-1}(\mathcal{E})$ ein Erzeuger von $f^{-1}(\mathcal{B})$.*

Beweis. Offenbar ist $f^{-1}(\mathcal{E}) \subseteq f^{-1}(\mathcal{B})$, und da $f^{-1}(\mathcal{B})$ σ-Algebra ist, folgt $\sigma\left(f^{-1}(\mathcal{E})\right) \subseteq f^{-1}(\mathcal{B})$. Umgekehrt ist

$$\mathcal{C} = \left\{B \subseteq Y : \ f^{-1}(B) \in \sigma\left(f^{-1}(\mathcal{E})\right)\right\}$$

eine σ-Algebra über Y, die $\mathcal{E}$ enthält, also $\mathcal{B} \subseteq \mathcal{C}$ und nach Def. von $\mathcal{C}$ folgt $f^{-1}(\mathcal{B}) \subseteq \sigma\left(f^{-1}(\mathcal{E})\right)$. ∎

Korollar 3.4. Ist $f : \Omega \to Y$ und $\mathcal{E}$ ein Erzeuger von $\mathcal{B}$, so ist f genau dann $(\mathcal{A} - \mathcal{B})$-messbar, falls $f^{-1}(\mathcal{E}) \subseteq \mathcal{A}$.

Beweis. Ist $f^{-1}(\mathcal{E}) \subseteq \mathcal{A}$, so auch $f^{-1}(\mathcal{B}) = \sigma\left(f^{-1}(\mathcal{E})\right) \subseteq \mathcal{A}$, da $\mathcal{A}$ σ-Algebra. ∎

Beispiel 3.5. Sei $\mathcal{B}$ die Borel-σ-Algebra über $\mathbb{R}$. Für $f : \Omega \to \mathbb{R}$ sind äquivalent:

1. f ist $(\mathcal{A} - \mathcal{B})$-messbar (dies heißt auch *Borel-messbar*).
2. $f^{-1}([a, \infty)) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.
3. $f^{-1}((a, \infty)) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.
4. $f^{-1}((-\infty, a]) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.
5. $f^{-1}((-\infty, a)) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.

Dies ist klar, da jedes der angegebenen Mengensysteme ein Erzeuger für $\mathcal{B}$ ist.

Korollar 3.6. Seien Ω und Y topologische Räume mit Borel-σ-Algebren $\mathcal{B}(\Omega)$ und $\mathcal{B}(Y)$, und ist $f : \Omega \to Y$ stetig, so ist f $(\mathcal{B}(\Omega) - \mathcal{B}(Y))$-messbar.

Beweis. Stetigkeit wurde in Analysis II so definiert, dass für $U \subseteq Y$ offen, auch $f^{-1}(U)$ offen ist. Die offenen Teilmengen von Y sind ein Erzeuger von $\mathcal{B}(Y)$, und ist $U \subseteq Y$ offen, so ist $f^{-1}(U)$ offen, also in $\mathcal{B}(\Omega)$. ∎

Korollar 3.7. Ist $D \subseteq \Omega$ beliebig und $\mathcal{A}(D) = \{A \cap D, \quad A \in \mathcal{A}\}$ die Spur σ-Algebra (vgl. Beispiel 2.4), und ist $\mathcal{E}$ ein Erzeuger von $\mathcal{A}$, so ist $\mathcal{E}(D) = \{E \cap D,\ E \in \mathcal{E}\}$ ein Erzeuger von $\mathcal{A}(D)$.

Beweis. Wende Satz 3.3 an auf die Inklusionsabbildung $i : D \to \Omega$: Es ist $i^{-1}(\mathcal{A}) = \mathcal{A}(D)$ und $i^{-1}(\mathcal{E}) = \mathcal{E}(D)$. ∎

Satz 3.8. *Seien $(\Omega, \mathcal{A})$, $(Y, \mathcal{B})$ und $(Z, \mathcal{C})$ Messräume und ist $f : \Omega \to Y$ $(\mathcal{A} - \mathcal{B})$-messbar und $g : Y \to Z$ $(\mathcal{B} - \mathcal{C})$-messbar, so ist die Verknüpfung $g \circ f : \Omega \to Z$, $(\mathcal{A} - \mathcal{C})$-messbar.*

Beweis. Ist $C \in \mathcal{C}$, so ist $g^{-1}(C) \in \mathcal{B}$ und daher $(g \circ f)^{-1}(C) = f^{-1}\big(g^{-1}(C)\big) \in \mathcal{A}$. ∎

Definition 3.9. *Seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(Y, \mathcal{B})$ ein Messraum. Eine Abbildung $X : \Omega \to Y$ heißt (Y-wertige)* ***Zufallsvariable****, falls die Abbildung messbar ist.*

Bemerkung. Manche Bücher verwenden den Begriff *Zufallsvariable* nur im Standardfall, dass $Y \subseteq \mathbb{R}$. Allgemeiner wird vereinzelt dann der Begriff Zufallselement eingeführt. Wir verwenden den Begriff Zufallsvariable allgemein und spezifizieren ggf. *reellwertige Zufallsvariablen*. Für den Fall $Y \subseteq \mathbb{R}^p$, $p > 1$, wird später der Begriff *Zufallsvektor* eingeführt.

Wesentlich für eine Zufallsvariable ist nun, mit welchen Wahrscheinlichkeiten sie bestimmte Werte annimmt. Betrachte dazu allgemein einen Maßraum $(\Omega, \mathcal{A}, \mu)$, sowie eine $(\mathcal{A} - \mathcal{B})$-messbare Abbildung $f : \Omega \to Y$. Die Abbildung f induziert auf natürliche Weise ein Maß auf dem Messraum $(Y, \mathcal{B})$.

Satz 3.10. *Ist $f : \Omega \to Y$ eine $(\mathcal{A} - \mathcal{B})$-messbare Abbildung und μ ein Maß auf dem Messraum $(\Omega, \mathcal{A})$. Dann wird durch*

$$\nu(B) = \mu(f^{-1}(B)),\, B \in \mathcal{B}\,, \tag{3.1}$$

ein Maß ν auf $(Y, \mathcal{B})$ definiert.

Beweis. Da $f^{-1}(B) \in \mathcal{A}$ ist $\mu(f^{-1}(B))$ wohldefiniert. Offensichtlich gilt $\nu(\emptyset) = 0$ und Nicht-Negativität, so dass die σ-Additivität zu zeigen bleibt. Bezeichne hierzu (B_n) eine Folge paarweiser disjunkter Mengen aus $\mathcal{B}$. Dann erhält man für die Folge paarweiser disjunkter Mengen $(f^{-1}(B_n))$ aus $\mathcal{A}$:

$$\begin{aligned} \nu\Big(\bigcup_n B_n\Big) &= \mu\Big(f^{-1}\Big(\bigcup_n B_n\Big)\Big) = \mu\Big(\bigcup_n f^{-1}(B_n)\Big) \\ &= \sum_n \mu(f^{-1}(B_n)) = \sum_n \nu(B_n)\,. \end{aligned}$$

∎

Definition 3.11. *Das im vorangehenden Satz definierte Maß ν heißt* ***Bildmaß*** *bezüglich f und man schreibt $\nu = \mu \circ f^{-1}$.*

Definition 3.12. *Auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ sei eine reellwertige Zufallsvariable $X : \Omega \to \mathbb{R}$, also $(\mathcal{A} - \mathcal{B}^1)$-messbar, gegeben. Dann wird durch $\mathbb{P} \circ X^{-1}$ ein Wahrscheinlichkeitsmaß auf dem Messraum $(\mathbb{R}, \mathcal{B}^1)$ induziert. Dieses nennt man die* ***Verteilung*** *der reellwertigen Zufallsvariablen X. Wir schreiben $\mathbb{P}_X$ für die Verteilung einer reellwertigen Zufallsvariablen X.*

Bemerkung. Wir schreiben statt $\mathbb{P}(\{X \in A\})$ in Zukunft meist nur $\mathbb{P}(X \in A)$ und entsprechend $\mathbb{P}(X = x)$ für $\mathbb{P}(\{X = x\})$.

Bemerkung. Viele Charakteristika einer Zufallsvariablen X werden nur von der Verteilung $\mathbb{P}_X$ von X abhängen. Dann spezifizieren wir später $X \sim \mathbb{P}_X$, kurz für X ist verteilt nach $\mathbb{P}_X$, ohne den Urbildraum Ω noch zu nennen.

Beispiel 3.13 (Augensumme beim zweifachen Würfelwurf)**.** Hier gilt

$$X(\Omega) = \{2, 3, 4, \dots, 12\}.$$

Ist $x \in X(\Omega)$, so ist $\mathrm{p}_X(x) = \mathbb{P}(X = x) = \frac{\operatorname{card}\{X=x\}}{36}$. Die einzelnen Werte sind in Tabelle 3.1 dargestellt.

Tabelle 3.1: Wahrscheinlichkeiten der Augensummen bei zweifachem Würfelwurf

x	2	3	4	5	6	7	8	9	10	11	12
$\mathrm{p}_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

3.2 Diskrete Zufallsvariablen und Verteilungen

Auf einem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(X), \mathbb{P})$ ist jede Abbildung automatisch messbar.

Lemma 3.14. *Ist $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable auf dem diskreten Wahrscheinlichkeitsraum $(\Omega, \mathcal{P}(X), \mathbb{P})$, so wird auf dem Bildbereich $X(\Omega)$ eine Wahrscheinlichkeitsfunktion*

$$\mathrm{p}_X : X(\Omega) \to [0,1]$$

definiert durch

$$\mathrm{p}_X(x) = \mathbb{P}(\{X = x\}), \qquad x \in X(\Omega),$$

mit zugehörigem Wahrscheinlichkeitsmaß

$$\mathbb{P}_X(A) = \mathbb{P}(\{X \in A\}), \qquad A \subseteq X(\Omega) \subseteq \mathbb{R}.$$

Beweis. Es ist

$$\Omega = \bigcup_{x \in X(\Omega)} \{X = x\}$$

eine disjunkte Zerlegung von Ω, daher ist

$$\sum_{x \in X(\Omega)} \mathrm{p}_X(x) = \sum_{x \in X(\Omega)} \mathbb{P}(\{X = x\}) = \mathbb{P}\Big(\bigcup_{x \in X(\Omega)} \{X = x\}\Big) = \mathbb{P}(\Omega) = 1,$$

also ist p_X eine Wahrscheinlichkeitsfunktion. Da für $A \subseteq X(\Omega)$

$$\{X \in A\} = \bigcup_{x \in A} \{X = x\}$$

eine disjunkte Zerlegung ist, zeigt eine analoge Rechnung die Gleichheit

$$\mathbb{P}(\{X \in A\}) = \sum_{x \in A} \mathrm{p}_X(x).$$

■

Wir nennen in einem solchen Fall X eine *diskrete Zufallsvariable*. Eine *diskrete Verteilung* einer diskreten Zufallsvariable lässt sich also eindeutig durch eine Zähldichte (Wahrscheinlichkeitsfunktion) charakterisieren. Wir führen in diesem Kapitel die wichtigsten Beispiele diskreter Verteilungen ein.

3.2.1 Indikatorzufallsvariablen, BERNOULLI-Verteilung und Binomialverteilung

Wir betrachten eine Klasse von besonders einfachen Zufallsvariablen, den *Indikatorzufallsvariablen* oder auch *Indikatorfunktionen*. Diese sind besonders nützlich um zu zählen, wie oft Ereignisse auftreten.

Definition 3.15. *Ist $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $A \subseteq \Omega$, so heißt*

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

***Indikatorfunktion** von A. Eine Verteilung $\mathbb{P}_X$ auf $\{0,1\}$ mit $\mathbb{P}_X(1) = \mathrm{p}_X(1) = p \in [0,1]$ heißt **BERNOULLI-Verteilung** $\mathrm{Ber}(p)$.*

Bemerkung (Verteilung von $\mathbb{1}_A$). Es gilt

$$\{\mathbb{1}_A = 1\} = \{\omega \in \Omega \mid \mathbb{1}_A(\omega) = 1\} = A, \quad \{\mathbb{1}_A = 0\} = \{\omega \in \Omega \mid \mathbb{1}_A(\omega) = 0\} = A^{\complement},$$

somit für $A \in \mathcal{A}$, dass $\mathrm{p}_{\mathbb{1}_A}(1) = \mathbb{P}(A)$ und $\mathrm{p}_{\mathbb{1}_A}(0) = \mathbb{P}(A^{\complement}) = 1 - \mathbb{P}(A)$. Für $A \in \mathcal{A}$ ist eine Indikatorfunktion $\mathbb{1}_A$ also BERNOULLI-verteilt mit Erfolgswahrscheinlichkeit $p = \mathbb{P}(A)$.

Bemerkung (Rechenregeln für Indikatorfunktionen). Für alle $A, B \subseteq \Omega$ gelten die Rechenregeln

- $\mathbb{1}_{A^{\complement}}(\omega) = 1 - \mathbb{1}_A(\omega)$,
- $\mathbb{1}_{A \cap B}(\omega) = \mathbb{1}_A(\omega) \cdot \mathbb{1}_B(\omega)$.

Daraus folgt für $A_1, \ldots, A_n \subseteq \Omega$

$$\begin{aligned} \mathbb{1}_{A_1 \cup \cdots \cup A_n}(\omega) &= 1 - \mathbb{1}_{(A_1 \cup \cdots \cup A_n)^{\complement}}(\omega) = 1 - \mathbb{1}_{A_1^{\complement} \cap \cdots \cap A_n^{\complement}}(\omega) = 1 - \prod_{k=1}^{n} \mathbb{1}_{A_k^{\complement}}(\omega) \\ &= 1 - \prod_{k=1}^{n} (1 - \mathbb{1}_{A_k}(\omega)) = 1 - \sum_{j=0}^{n} (-1)^j \sum_{1 \le i_1 < \ldots < i_j \le n} \mathbb{1}_{A_{i_1}}(\omega) \cdot \cdots \cdot \mathbb{1}_{A_{i_j}}(\omega) \\ &= \sum_{j=1}^{n} (-1)^{j-1} \sum_{1 \le i_1 < \ldots < i_j \le n} \mathbb{1}_{A_{i_1} \cap \ldots \cap A_{i_j}}(\omega). \end{aligned} \tag{3.2}$$

Durch Anwendung des Erwartungswerts, der später eingeführt wird, erhalten wir aus (3.2) unmittelbar wieder die Einschluss-Ausschluss-Formel.

Bemerkung (Summe von Indikatorfunktionen). Sind $A_1, \ldots, A_n \subseteq \Omega$, so zählt

$$X = \mathbb{1}_{A_1} + \cdots + \mathbb{1}_{A_n},$$

in wie vielen A_i ein ω liegt, also gilt $X(\omega) = k$, falls $\omega \in A_i$ für genau k Indizes $i = 1, \ldots, n$.

Betrachten wir $(\Omega_j, \mathbb{P}_j)$, $j = 1, \ldots, n$, diskrete Wahrscheinlichkeitsräume, die unabhängig voneinander durchgeführte Teilexperimente eines gesamten Experimentes modellieren. Dazu können wir auf $\Omega_1 \times \cdots \times \Omega_n$ die Wahrscheinlichkeitsfunktion

$$\mathrm{p}_1 \otimes \ldots \otimes \mathrm{p}_n : \Omega_1 \times \cdots \times \Omega_n \to [0,1], \qquad \mathrm{p}_1 \otimes \ldots \otimes \mathrm{p}_n(\omega) = \prod_{i=1}^{n} \mathrm{p}_i(\omega_i) \tag{3.3}$$

betrachten, wobei $\omega = (\omega_1, \dots, \omega_n) \in \Omega_1 \times \dots \times \Omega_n$ und p_i die zugehörigen Wahrscheinlichkeitsfunktionen der $\mathbb{P}_i$ seien. Dass es sich tatsächlich um eine Wahrscheinlichkeitsfunktion handelt, folgt später als Spezialfall der allgemeineren *Produkträume*. Dies kann hier aber auch leicht nachgeprüft werden. Wenden wir dies mit $(\Omega_i, \mathbb{P}_i) = (\Omega, \mathbb{P})$, $i = 1, \dots, n$ an, so erhalten wir ein *n-faches Produktexperiment* $(\Omega^n, \mathbb{P}^{\otimes n})$ zu $(\Omega, \mathbb{P})$, wobei $\mathbb{P}^{\otimes n} = \mathbb{P} \otimes \ldots \otimes \mathbb{P}$. Ist $\mathrm{p}(\omega) = \mathbb{P}(\{\omega\})$ die Wahrscheinlichkeitsfunktion von $\mathbb{P}$, so schreiben wir $\mathrm{p}^{\otimes n}$ für die Wahrscheinlichkeitsfunktion von $\mathbb{P}^{\otimes n}$. Für $\Omega = \{0, 1\}$ mit einer BERNOULLI-Verteilung $\mathrm{p}(1) = p$, $\mathrm{p}(0) = 1 - p$, $p \in [0, 1]$, erhalten wir entsprechend das n-fache BERNOULLI-Experiment. Für $\omega = (\omega_1, \dots, \omega_n) \in \{0, 1\}^n$ ist dann

$$\mathrm{p}^{\otimes n}(\omega) = p^{\sum_{j=1}^n \omega_j} \cdot (1-p)^{n - \sum_{j=1}^n \omega_j}.$$

Dabei ist $\sum_{j=1}^n \omega_j$ die Anzahl der Einsen und $n - \sum_{j=1}^n \omega_j$ die Anzahl der Nullen. Wir definieren eine Zufallsvariable

$$X : \{0, 1\}^n \to \{0, \dots, n\}, \quad X(\omega) = \sum_{j=1}^n \omega_j,$$

welche die Anzahl der Erfolge in n unabhängigen BERNOULLI-Experimenten mit jeweils gleicher Erfolgswahrscheinlichkeit darstellt. Dann ist

$$\{X = k\} = \Big\{\omega = (\omega_1, \dots, \omega_n) \Big| \sum_{j=1}^n \omega_j = k\Big\}, \quad \operatorname{card}\{X = k\} = \binom{n}{k},$$

da wir genau k Einsen auf n mögliche Plätze verteilen. Dabei hat jedes $\omega \in \{X = k\}$ die Wahrscheinlichkeit $p^k (1-p)^{n-k}$, also gilt

$$\mathbb{P}^{\otimes n}(\{X = k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Definition 3.16. *Seien $n \in \mathbb{N}$ und $p \in [0, 1]$. Dann heißt die zur Wahrscheinlichkeitsfunktion p auf $\{0, \dots, n\}$ mit*

$$\mathrm{p}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

*gehörige Verteilung die **Binomialverteilung** mit Parametern n und p. Ist eine Zufallsvariable X verteilt nach $X \sim \mathrm{p}$, so schreiben wir $X \sim \mathrm{Bin}(n, p)$.*

3.2.2 Hypergeometrische Verteilung

Problem 3.17. *In einer Urne befinden sich die Kugeln $1, \dots, N$, davon sind die Kugeln $1, \dots, R$ rot und die Kugeln $R+1, \dots, N$ schwarz. Wir ziehen nun $n \le N$ Kugeln ohne Zurücklegen. Wie wahrscheinlich ist es, dass davon genau r Kugeln rot sind?*

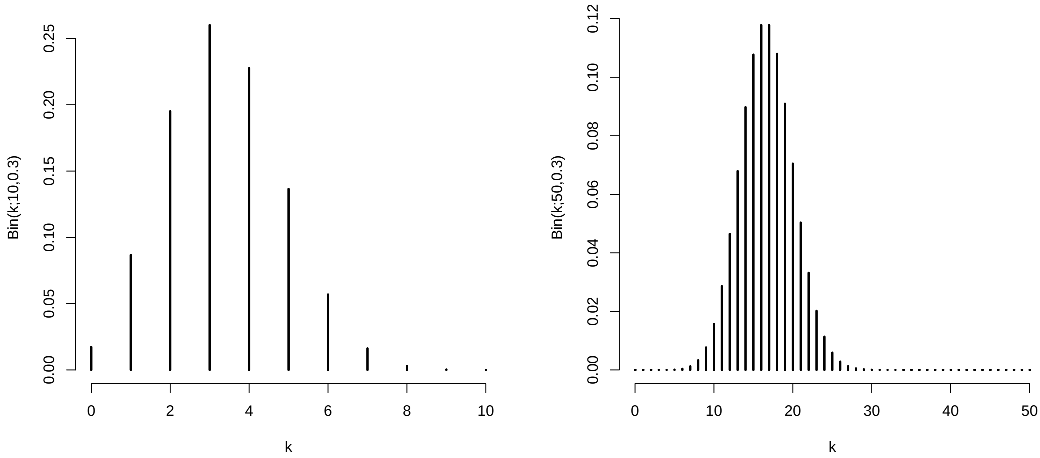
Modell. Da wir n aus N Kugeln ohne Reihenfolge und ohne Zurücklegen ziehen, modellieren wir das Zufallsexperiment durch den LAPLACE-Raum

$$\Omega = \Omega_{\mathrm{III}}^{n,N} = \{A \subseteq \{1, \dots, N\} \mid \operatorname{card} A = n\}, \quad \operatorname{card} \Omega = \binom{N}{n}.$$

$\Omega_{\mathrm{III}}^{n,N}$ ist die Bezeichnung des entsprechenden Urnenmodells. Das gesuchte Ereignis ist nun

$$E_r = \{A \in \Omega_{\mathrm{III}}^{n,N} \mid \operatorname{card}(A \cap \{1, \dots, R\}) = r\} = \{A \in \Omega_{\mathrm{III}}^{n,N} \mid \operatorname{card}(A \cap \{R+1, \dots, N\}) = n - r\}.$$

Für die Wahl von $A \cap \{1, \dots, R\}$ gibt es also $\binom{R}{r}$ Möglichkeiten, falls $0 \le r \le R$ (sonst nicht möglich), und für die Wahl von $A \cap \{R+1, \dots, N\}$ gibt es $\binom{N-R}{n-r}$ Möglichkeiten, falls $\max(n+R-N, 0) \le r \le n$



(a) $n = 10$, $p = 1/3$ (b) $n = 50$, $p = 1/3$

Abbildung 3.1: Wahrscheinlichkeiten $\mathrm{Bin}(k;n,p)$ der Binomialverteilung.

(sonst nicht möglich). Somit gilt

$$\operatorname{card} E_r = \binom{R}{r}\binom{N-R}{n-r} \Leftrightarrow \mathbb{P}(E_r) = \frac{\operatorname{card} E_r}{\operatorname{card}\Omega} = \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}, \quad r = \max(n+R-N,0),\ldots,\min(n,R).$$

Definition 3.18. *Ist $n,R,N \in \mathbb{N}$ mit $n \leq N$, so heißt* $\mathrm{Hyp}(\cdot;n,R,N)$ *mit der Wahrscheinlichkeitsfunktion*

$$\mathrm{Hyp}(r;n,R,N) := \frac{\binom{R}{r}\binom{N-R}{n-r}}{\binom{N}{n}}, \quad r = \max(n+R-N,0),\ldots,\min(n,R)$$

*die **hypergeometrische Verteilung** mit Parametern n,R,N.*

Bemerkung. Es gilt

$$\mathrm{Hyp}(r;n,R,N) = \mathrm{Hyp}(r;R,n,N), \quad r = \max(n+R-N,0),\ldots,\min(n,R),$$

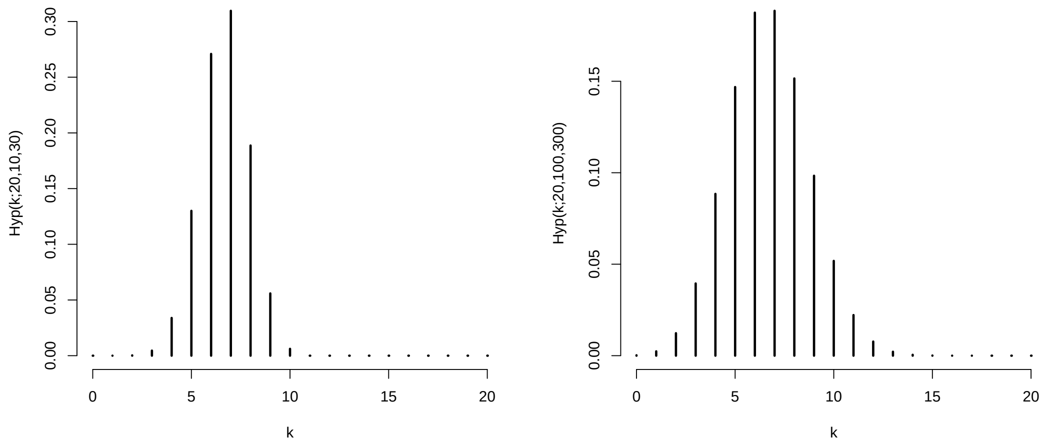
wie sich einfach nachrechnen lässt. Das ist auch argumentativ plausibel, indem wir die R Kugeln, die rot gefärbt werden, zufällig vor jedem Experiment bestimmen. Das verändert im LAPLACE-Raum die Wahrscheinlichkeiten nicht. r ist dann aber die Kardinalität der Schnittmenge zwischen einer zufälligen R-elementigen und einer zufälligen n-elementigen Teilmenge von $\{1,\ldots,N\}$. Da die Bildung der Schnittmenge kommutativ ist, können n und R beliebig vertauscht werden.

Sei $\mathcal{R}$ die Menge der roten Kugeln. Wir definieren nun für $i = 1,\ldots,n$ Ereignisse

$$A_i = \{\omega \in \Omega \mid \omega_i \in \mathcal{R}\},$$

die anzeigen, dass die Kugel i rot ist. Dann ist $X = \sum_{i=1}^{n} 1_{A_i}$ die Anzahl der gezogenen roten Kugeln, also gilt $X \sim \mathrm{Hyp}(\cdot;n,R,N)$.

Beispiel 3.19. 1. Lotto 6 aus 49. Hier werden vom Spieler 6 Ziffern als korrekt getippt. Dann werden zufällig 6 Ziffern gezogen, die der Spieler mit seinem Tipp vergleicht. Es gilt also $N = 49$ und



(a) $N = 30$, $R = 10$, $n = 20$ (b) $N = 300$, $R = 100$, $n = 20$

Abbildung 3.2: Wahrscheinlichkeiten $\mathrm{Hyp}(k;n,R,N)$ der hypergeometrischen Verteilung.

$R = n = 6$ und somit

$$\mathbb{P}(r \text{ richtige getippt}) = \frac{\binom{6}{r}\binom{43}{6-r}}{\binom{49}{6}}, \quad r = 0,\dots,6.$$

Insbesondere

$$\mathbb{P}(6 \text{ richtige getippt}) = \frac{\binom{6}{6}\binom{43}{0}}{\binom{49}{6}} = \frac{1}{\binom{49}{6}} = \frac{1}{13983816}.$$

2. Skat. Gesucht ist die Wahrscheinlichkeit für mindestens einen Buben im Skat. Dabei wird aus $N = 32$ Karten mit $R = 4$ Buben ein Skat aus $n = 2$ Karten gezogen. Somit gilt

$$\mathbb{P}(r \text{ Buben im Skat}) = \frac{\binom{4}{r}\binom{28}{2-r}}{\binom{32}{2}}, \quad r = 0,1,2.$$

Daraus folgt

$$\begin{aligned}\mathbb{P}(\text{mindestens ein Bube im Skat}) &= 1 - \mathbb{P}(0 \text{ Buben im Skat}) \\ &= 1 - \frac{\binom{28}{2}}{\binom{32}{2}} = 1 - \frac{28 \cdot 27}{32 \cdot 31} \approx 23.79\%.\end{aligned}$$

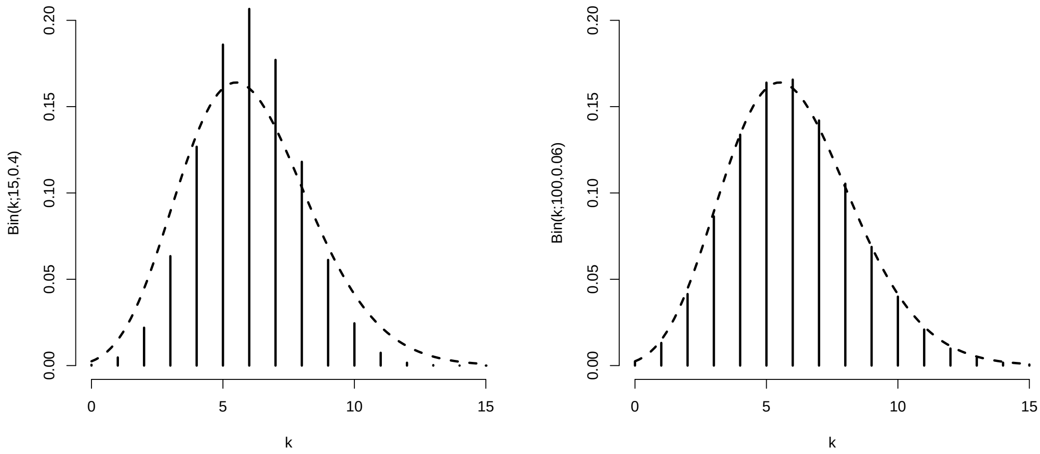
3.2.3 Geometrische Verteilung

Definition 3.20. *Ist $p \in (0,1)$, so ist $\mathrm{p} : \mathbb{N} \to [0,1]$ mit*

$$\mathrm{p}(k) = (1-p)^{k-1} \cdot p$$

eine Wahrscheinlichkeitsfunktion. Die zugehörige Verteilung heißt die ***geometrische Verteilung*** *mit Parameter p. Ist eine Zufallsvariable X verteilt nach p, so schreiben wir $X \sim \mathrm{Geom}(p)$.*

Bemerkung. Eine Zufallsvariable $X \sim \mathrm{Geom}(p)$ stellt die Wartezeit bis zum ersten Erfolg in einer un-



(a) $n = 15$, $p = 0.4$

(b) $n = 100$, $p = 0.06$

Abbildung 3.3: Wahrscheinlichkeiten $\mathrm{Bin}(k;n,p)$ der Binomialverteilung zusammen mit der Grenzfunktion $x \mapsto e^{-\lambda}\lambda^x/x!$ der POISSON-Approximation (gestrichelt) für $\lambda = 6$.

endlichen Folge von unabhängigen BERNOULLI-Experimenten mit jeweils gleicher Erfolgswahrscheinlichkeit p dar.

Es gilt

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

also ist $\mathbb{P}$ eine Verteilung.

Bemerkung. Es liegt nahe, eine Zufallsvariable $X \sim \mathrm{Geom}(p)$ mit $X : \{0,1\}^{\mathbb{N}} \to \mathbb{N}$ zu definieren. Das ist aber im Rahmen der Theorie diskreter Wahrscheinlichkeitsräume so nicht möglich, da die Menge der $\{0,1\}$-Folgen überabzählbar ist, also kein diskreter Wahrscheinlichkeitsraum vorliegt.

3.2.4 POISSON-Verteilung

Satz 3.21 (POISSON-Approximation der Binomialverteilung)**.** *Sei* $(p_n) \in (0,1)^{\mathbb{N}}$*, sodass*

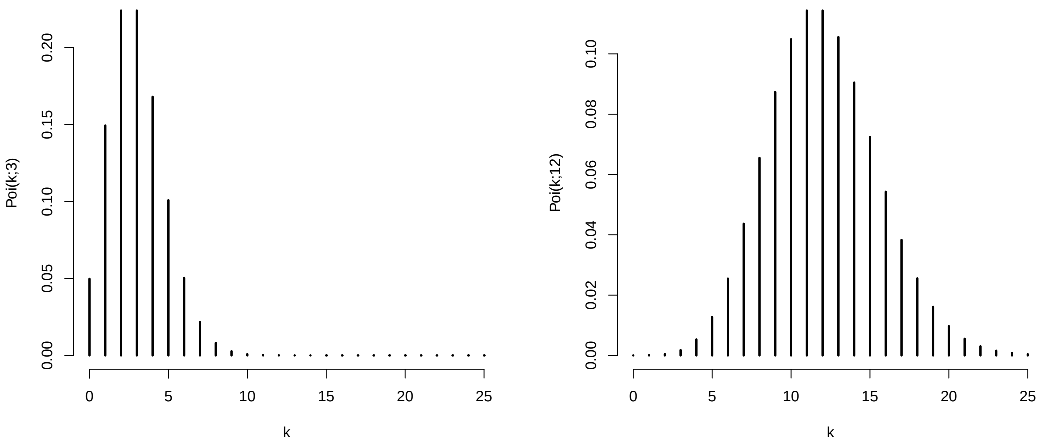
$$\lim_{n\to\infty} p_n = 0, \quad \lim_{n\to\infty} (n \cdot p_n) = \lambda \in (0,\infty).$$

Ist $X_n \sim \mathrm{Bin}(n, p_n)$*, so gilt für* $k \in \mathbb{N}_0$

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!}.$$

Beweis. Für n groß ist $n \geq k$, da k fest ist. Also gilt

$$\mathbb{P}(X_n = k) = \binom{n}{k} \cdot p_n^k \cdot (1-p_n)^{n-k} = \frac{(n)_k}{n^k} \cdot (np_n)^k \cdot \frac{1}{k!} \cdot \left(1 - \frac{np_n}{n}\right)^n \cdot (1-p_n)^{-k}.$$



(a) $\lambda = 3$ (b) $\lambda = 12$

Abbildung 3.4: Wahrscheinlichkeiten $\mathrm{Poi}(k;\lambda)$ der POISSON-Verteilung.

Für die einzelnen Faktoren gilt

$$\lim_{n\to\infty} \frac{(n)_k}{n^k} = 1, \quad \lim_{n\to\infty} (1-p_n)^k = (1-0)^k = 1, \quad \lim_{n\to\infty} (np_n)^k = \lambda^k.$$

Weiterhin gilt

$$\lim_{n\to\infty} x_n = x \Rightarrow \lim_{n\to\infty} \left(1 - \frac{x_n}{n}\right)^n = \mathrm{e}^{-x},$$

also

$$\lim_{n\to\infty} \left(1 - \frac{np_n}{n}\right)^n = \mathrm{e}^{-\lambda}.$$

Es folgt also

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = 1 \cdot \lambda^k \cdot \frac{1}{k!} \cdot \mathrm{e}^{-\lambda} \cdot 1 = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!}. \qquad \blacksquare$$

Definition 3.22. *Ist* $\lambda > 0$*, so ist* $\mathrm{p} : \mathbb{N}_0 \to [0,1]$ *mit*

$$\mathrm{p}(k) = \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!}$$

eine Wahrscheinlichkeitsfunktion. Die zugehörige Verteilung heißt die **POISSON-*Verteilung*** *mit Parameter* λ*. Ist eine Zufallsvariable* X *verteilt nach* $X \sim \mathrm{p}$*, so schreiben wir* $X \sim \mathrm{Poi}(\lambda)$.

Es gilt

$$\sum_{k=0}^{\infty} \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \cdot \mathrm{e}^{\lambda} = 1,$$

also ist $\mathrm{p}(k)$ eine Wahrscheinlichkeitsfunktion.

Bemerkung. Die POISSON-Verteilung $\mathrm{Poi}(n \cdot p_n)$ ist eine gute Approximation der Binomialverteilung $\mathrm{Bin}(n, p_n)$ für große n und kleine p_n.

Bemerkung. POISSON-Verteilungen sind ein Modell für die Anzahl der Ereignisse bei vielen möglichen, aber jeweils unwahrscheinlichen Ereignissen. Diese müssen nicht notwendigerweise unabhängig sein.

Ein Beispiel ist das Rencontre-Problem. Hier wird zufällig eine Permutation $\pi \in \mathfrak{S}_n$ gewählt. Gesucht ist nun die asymptotische Verteilung der Zufallsvariablen X der Anzahl der Fixpunkte. An jeder Stelle ist die Wahrscheinlichkeit für einen Fixpunkt $\frac{1}{n}$, diese Ereignisse sind aber nicht unabhängig. Wir hatten aber gesehen, dass

$$\lim_{n\to\infty} \mathbb{P}(X=k) = \lim_{n\to\infty} \left(\frac{1}{k!} \cdot \left(\frac{1}{2!} - \frac{1}{3!} \pm \cdots + (-1)^{n-k} \frac{1}{(n-k)!} \right) \right) = \frac{1}{k!} \mathrm{e}^{-1},$$

also ist X asymptotisch nach $\mathrm{Poi}(1)$ verteilt.

3.3 Messbare numerische Funktionen

Die erweiterte Zahlengerade $\overline{\mathbb{R}}$

Als Vorbereitung für die Integrationstheorie betrachten wir reellwertige Funktionen, die zusätzlich auch Werte $+\infty$ und $-\infty$ annehmen können. Setze dazu $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$.

Topologie und Anordnung: $\overline{\mathbb{R}}$ ist ein kompakter metrischer Raum bzgl. der Metrik $d(x,y) = |\arctan x - \arctan y|$. Die Topologie auf der Teilmenge $\mathbb{R}$ bleibt dabei erhalten. Für alle $x \in \mathbb{R}$ ist $-\infty < x < \infty$, $|\infty| = |-\infty| = \infty$, und die Betragsfunktion $|\cdot| : \overline{\mathbb{R}} \to [0,\infty]$ ist stetig.

Rechenregeln:

$$\begin{aligned} x + (\pm\infty) &= (\pm\infty) + x = \pm\infty \,, \forall x \in \mathbb{R}\,, \\ x \cdot (\pm\infty) &= (\pm\infty) \cdot x = \begin{cases} \pm\infty \,, x > 0 \\ \mp\infty \,, x < 0 \end{cases} , \\ 0 \cdot (\pm\infty) &= (\pm\infty) \cdot 0 = 0\,. \end{aligned}$$

Nicht definiert wird hingegen $(\pm\infty) + (\pm\infty)$.
In $\overline{\mathbb{R}}$ existieren stets Supremum, Infimum, und somit auch limsup, liminf.

Messbare Mengen: Die Borel-σ-Algebra des topologischen Raums $\overline{\mathbb{R}}$ ist

$$\mathcal{B}(\overline{\mathbb{R}}) = \overline{\mathcal{B}} := \{A \subseteq \overline{\mathbb{R}} :\ A \cap \mathbb{R} \in \mathcal{B}(\mathbb{R})\} = \{B \cup E;\ B \in \mathcal{B}(\mathbb{R}),\ E \subseteq \{-\infty,\infty\}\}.$$

Dabei ist $\mathcal{B}^1 \supseteq \overline{\mathcal{B}}$ klar. Weiter ist $\overline{\mathcal{B}}$ eine σ-Algebra, und enthält die offenen Teilmengen von $\overline{\mathbb{R}}$, also ist $\mathcal{B}^1 \subseteq \overline{\mathcal{B}}$
Offenbar ist die Spur-σ-Algebra von $\overline{\mathcal{B}}$ in $\mathbb{R}$ gerade $\mathcal{B}^1$. Die Mengensysteme $\{[a,\infty], a \in \mathbb{R}\}$, $\{(a,\infty], a \in \mathbb{R}\}$ usw. sind Erzeuger von $\overline{\mathcal{B}}$.

Messbare numerische Funktionen

Ist $(\Omega, \mathcal{A})$ ein fixierter Messraum und $f : \Omega \to \overline{\mathbb{R}}$ $(\mathcal{A} - \overline{\mathcal{B}})$-messbar, so heißt f eine *messbare numerische Funktion*. Wir bezeichnen mit $\mathcal{M}$ die Menge der messbaren numerischen Funktionen und mit

$$\mathcal{M}^+ := \{f : \Omega \to [0,\infty] \text{ messbar}\}$$

die Menge der nicht-negativen messbaren numerischen Funktion.

Satz 3.23. *Für $f : \Omega \to \overline{\mathbb{R}}$ sind äquivalent:*
1. f messbar.
2. $f^{-1}([a,\infty]) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.
3. $f^{-1}((a,\infty]) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.
4. $f^{-1}([-\infty,a]) \in \mathcal{A}$ für alle $a \in \mathbb{R}$.
5. $f^{-1}([-\infty,a)) \in \mathcal{A}$ für alle $a \in \mathbb{R}$. ▪

Dies ist klar mit Korollar 3.4, da jedes der angegebenen Mengensysteme ein Erzeuger für $\overline{\mathcal{B}}$ ist. Für

$f, g : \Omega \to \mathbb{R}$, so setze

$$(f \vee g)(x) = \max\big(f(x), g(x)\big), \qquad (f \wedge g)(x) = \min\big(f(x), g(x)\big),$$

und

$$f^+(x) = \big(f \vee 0\big)(x), \qquad f^-(x) = -\big(f \wedge 0\big)(x).$$

Satz 3.24. *Für $f, g : \Omega \to \overline{\mathbb{R}}$ messbar sind $f \vee g$, $f \wedge g$, f^+, f^- und $|f|$ messbar.*

Beweis. Es ist

$$\{(f \vee g) < a\} = \{f < a\} \cap \{g < a\}.$$

Weiter ist die konstante Funktion 0 messbar, und $x \mapsto |x|$ stetig, also messbar. ∎

Satz 3.25. *Ist $f_n : \Omega \to \overline{\mathbb{R}}$ eine Folge messbarer Funktionen, so sind auch $\sup_n f_n$, $\inf_n f_n$, $\liminf_n f_n$ und $\limsup_n f_n$ messbar. Existiert insbesondere $\lim_n f_n$ (punktweise), so ist die Grenzfunktion messbar.*

Beweis. Für $a \in \mathbb{R}$ ist

$$\{\sup_n f_n > a\} = \bigcup_n \{f_n > a\},$$

und

$$\{\limsup_n f_n > a\} = \{\omega \in \Omega : f_n(\omega) > a \text{ für } \infty\text{-viele } n\} = \bigcap_n \bigcup_{k \geq n} \{f_k > a\},$$

analog für das Infimum. Falls der Limes existiert, ist dieser $= \limsup$, also messbar. ∎

Satz 3.26. *Eine Abbildung $f = (f_1, \dots, f_p)^T : \Omega \to \mathbb{R}^p$ ist genau dann messbar, falls die Koordinatenfunktionen $f_j : \Omega \to \mathbb{R}$ messbar sind, $j = 1, \dots, p$.*

Beweis. $\Rightarrow$ ist klar, da die Projektionen auf Koordinaten stetig sind.
$\Leftarrow$: Für $(\mathbf{a}, \mathbf{b}] \in \mathfrak{J}^p$ ist

$$f^{-1}((\mathbf{a}, \mathbf{b}]) = \bigcap_{j=1}^{p} f_j^{-1}((a_j, b_j]).$$

∎

Satz 3.27. *Sind $f, g : \Omega \to \mathbb{R}$ messbar, so sind $f + g$, $f \cdot g$, af für $a \in \mathbb{R}$ messbar.*

Beweis. Die Abbildung $(f, g) : \Omega \to \mathbb{R}^2$, $\omega \mapsto \big(f(\omega), g(\omega)\big)$ ist messbar, und $\mathbb{R}^2 \to \mathbb{R}$, $(x, y) \mapsto x + y$ oder $x \cdot y$ sind stetig, also messbar. Für af setze $g = a$. ∎

Insbesondere sind dann auch die Mengen

$$\{f \leq g\} = \{f - g \leq 0\},$$

sowie $\{f = g\}$ messbar, und f ist genau dann messbar, falls f^+ und f^- messbar sind (da $f = f^+ - f^-$). Satz 3.27 gilt auch für $\overline{\mathbb{R}}$-wertige Funktionen, falls deren Summe definiert ist, siehe Elstrodt (1996) (Seite 106). Die Resultate dieses Teilkapitels können insbesondere für reellwertige, Borel-messbare Zufallsvariablen verwendet werden.

3.4 Verteilungsfunktionen

3.4.1 Maß-Eindeutigkeit

Wir betrachten zunächst zwei endliche Maße $\mu, \nu : \mathcal{A} \to [0,\infty)$ über einem Messraum $(\Omega, \mathcal{A})$, mit $\mu(\Omega) = \nu(\Omega)$, sowie die Menge

$$\mathcal{D}_{\mu,\nu} = \{A \in \mathcal{A} : \ \mu(A) = \nu(A)\}. \tag{3.4}$$

Angenommen, $\mathcal{D}_{\mu,\nu}$ wäre eine σ-Algebra, und die Maße μ und ν stimmten auf einem Erzeuger $\mathcal{E}$ von $\mathcal{A}$ überein. Dann folgte $\mathcal{E} \subseteq \mathcal{D}_{\mu,\nu}$ und somit $\mathcal{A} \subseteq \mathcal{D}_{\mu,\nu}$, also $\mu = \nu$.
Nun ist $\mathcal{D}_{\mu,\nu}$ jedoch im Allgemeinen keine σ-Algebra, da $\mathcal{D}_{\mu,\nu}$ nicht abgeschlossen unter beliebigen Vereinigungen sein muss. Wegen der σ-Additivität von μ und ν ist $\mathcal{D}_{\mu,\nu}$ jedoch abgeschlossen unter disjunkten Vereinigungen. Dies führt auf den folgenden Begriff.

Definition 3.28. *Ein Mengensystem $\mathcal{D} \subseteq \mathcal{P}(\Omega)$ heißt **Dynkin-System**, falls*
1. $\Omega \in \mathcal{D}$,
2. $A \in \mathcal{D} \ \Rightarrow A^{\complement} \in \mathcal{D}$,
3. Für disjunkte $(A_n)_{n\geq 1} \subseteq \mathcal{D}$ ist $\bigcup_n A_n \in \mathcal{D}$.

Lemma 3.29. *Seien $\mu, \nu : \mathcal{A} \to [0,\infty)$ endliche Maße über Ω mit $\mu(\Omega) = \nu(\Omega)$. Dann ist $\mathcal{D}_{\mu,\nu}$ in (3.4) ein Dynkin-System.*

Beweis. $\Omega \in \mathcal{D}_{\mu,\nu}$ nach Voraussetzung, $A^{\complement} = \Omega \setminus A \in \mathcal{D}_{\mu,\nu}$ wegen der Subtraktivität, und disjunkte Vereinigungen $\bigcup_n A_n \in \mathcal{D}_{\mu,\nu}$ wegen der σ-Additivität. ∎

Wir werden untersuchen, in welcher Situation ein (erzeugtes) Dynkin-System einer (erzeugten) σ-Algebra entspricht.

Satz 3.30. *Ein Dynkin-System $\mathcal{D}$ ist genau dann eine σ-Algebra, falls $\mathcal{D}$ $\cap$-stabil ist.* ▪

Beweis. Da $\mathcal{D}$ $\cap$-stabil und abgeschlossen unter Komplementen, liegen endliche (nicht notwendigerweise disjunkte) Vereinigungen in $\mathcal{D}$. Sind nun $A_n \in \mathcal{D}$, so sind

$$\tilde{A}_n = A_n \setminus (A_1 \cup \ldots \cup A_{n-1}) = A_n \cap (A_1 \cup \ldots \cup A_{n-1})^{\complement} \in \mathcal{D},$$

$(n \geq 2, \tilde{A}_1 = A_1)$ disjunkt, und somit

$$\bigcup_n A_n = \bigcup_n \tilde{A}_n \in \mathcal{D}.$$

∎

Lemma 3.31. *Sei $\mathcal{D}$ ein Dynkin-System. Dann gilt*
4. $A, B \in \mathcal{D}$, $A \subseteq B \Rightarrow B \setminus A \in \mathcal{D}$.
5. $(A_n)_{n\geq 1} \subseteq \mathcal{D}$, $A_n \subseteq A_{n+1} \Rightarrow \bigcup_n A_n \in \mathcal{D}$.

Beweis. 4.: Wegen $\emptyset = \Omega^{\complement} \in \mathcal{D}$ sind endliche disjunkte Vereinigungen in $\mathcal{D}$. Somit ist

$$B \setminus A = B \cap A^{\complement} = (B^{\complement} \cup A)^{\complement} \in \mathcal{D},$$

da $B^{\complement} \cap A = \emptyset$, und somit $B^{\complement} \cup A$ disjunkte Vereinigung ist.
5.: $\tilde{A}_n = A_n \setminus A_{n-1} \in \mathcal{D}$ nach 4. ($\tilde{A}_1 := A_1$), disjunkt, $\bigcup_n A_n = \bigcup_n \tilde{A}_n$. ∎

Ähnlich wie für σ-Algebren definieren wir das von einem Mengensystem $\mathcal{E}$ erzeugte Dynkin-System $\mathcal{D}(\mathcal{E})$ als Schnitt über alle Dynkin-Systeme, die $\mathcal{E}$ enthalten. $\mathcal{D}(\mathcal{E})$ ist dann ein Dynkin-System, und ist in jedem Dynkin-System enthalten, das $\mathcal{E}$ enthält.

Satz 3.32. *Sei $\mathcal{E} \subseteq \mathcal{P}(\Omega)$ ein $\cap$-stabiles Mengensystem. Dann ist das von $\mathcal{E}$ erzeugte Dynkin-System gleich der von $\mathcal{E}$ erzeugten σ-Algebra: $\sigma(\mathcal{E}) = \mathcal{D}(\mathcal{E})$.*

Beweis. Für $A \in \mathcal{D}(\mathcal{E})$ setze

$$\mathcal{D}_A = \big\{Q \in \mathcal{P}(\Omega) : \ A \cap Q \in \mathcal{D}(\mathcal{E})\big\}.$$

Wir wollen zeigen, dass

$$\forall\, A \in \mathcal{D}(\mathcal{E}) \text{ ist } \mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_A. \tag{3.5}$$

Dies bedeutet, dass für $A, B \in \mathcal{D}(\mathcal{E})$ auch $A \cap B \in \mathcal{D}(\mathcal{E})$, also nach Satz 3.30 $\mathcal{D}(\mathcal{E})$ eine σ-Algebra ist, welche $\mathcal{E}$ enthält, somit $\sigma(\mathcal{E}) \subseteq \mathcal{D}(\mathcal{E})$. Die umgekehrte Inklusion ist trivial.

Es bleibt, (3.5) zu zeigen. Zunächst bemerken wir, dass $\mathcal{D}_A$ ein Dynkin-System ist. Offenbar ist $\Omega \in \mathcal{D}_A$, da $A \in \mathcal{D}(\mathcal{E})$. Ist $B \in \mathcal{D}_A$, also $A \cap B \in \mathcal{D}(\mathcal{E})$, so nach Lemma 3.31 4. auch $A \cap B^{\complement} = A \setminus (A \cap B) \in \mathcal{D}(\mathcal{E})$. Sind schließlich $B_n \in \mathcal{D}_A$ disjunkt, so sind $A \cap B_n \in \mathcal{D}(\mathcal{E})$ disjunkt, und daher

$$\bigcup_n (A \cap B_n) = A \cap \Big(\bigcup_n B_n\Big) \in \mathcal{D}(\mathcal{E}),$$

also $\bigcup_n B_n \in \mathcal{D}_A$.

Bisher wurde nicht genutzt, dass $\mathcal{E}$ $\cap$-stabil ist. Ist nun $E \in \mathcal{E}$, so ist $\mathcal{E} \subseteq \mathcal{D}_E$, da $\mathcal{E}$ $\cap$-stabil ist. Da $\mathcal{D}_E$ ein Dynkin-System ist, folgt $\mathcal{D}(\mathcal{E}) \subseteq \mathcal{D}_E$ für alle $E \in \mathcal{E}$, also

$$E \cap A \in \mathcal{D}(\mathcal{E}) \quad \forall\, E \in \mathcal{E}, A \in \mathcal{D}(\mathcal{E}). \tag{3.6}$$

Ist nun $A \in \mathcal{D}(\mathcal{E})$, so ist wegen (3.6) $\mathcal{E} \subseteq \mathcal{D}_A$, und da $\mathcal{D}_A$ ein Dynkin-System ist, folgt (3.5). ∎

Satz 3.33 (Maß-Eindeutigkeitssatz). *Seien $\mu, \nu : \mathcal{A} \to [0, \infty)$ endliche Maße auf einem Messraum $(\Omega, \mathcal{A})$ mit $\mu(\Omega) = \nu(\Omega)$. Ist $\mathcal{E}$ ein $\cap$-stabiler Erzeuger von $\mathcal{A}$ und gilt $\mu(E) = \nu(E)$ für alle $E \in \mathcal{E}$, so ist $\mu = \nu$.*

Beweis. Das Mengensystem $\mathcal{D}_{\mu,\nu}$ in (3.4) ist nach Lemma 3.29 ein Dynkin-System, und enthält nach Voraussetzung $\mathcal{E}$. Somit enthält es das von $\mathcal{E}$ erzeugte Dynkin-System und nach Satz 3.32 die von $\mathcal{E}$ erzeugte σ-Algebra $\mathcal{A}$, daher $\mathcal{A} = \mathcal{D}_{\mu,\nu}$. ∎

Natürlich lässt sich der Satz auf zwei Wahrscheinlichkeitsmaße anwenden, woran wir hier vor allem interessiert sind.

3.4.2 Charakterisierung von Verteilungen über Verteilungsfunktionen

Das Resultat über Maßeindeutigkeit führt dazu, dass wir Wahrscheinlichkeitsmaße auf dem Messraum $(\mathbb{R}, \mathcal{B})$ geschickt durch Funktionen charakterisieren können.

Definition 3.34. *Eine Funktion $F : \mathbb{R} \to [0,1]$ heißt **Verteilungsfunktion**, falls sie*
1. monoton wachsend ist,
2. rechtsseitig stetig ist, d.h. für alle x und $x_n \downarrow x$ gilt $F(x_n) \to F(x)$, $n \to \infty$, sowie
3. $\lim\limits_{n\to\infty} F(n) = 1$, $\lim\limits_{n\to\infty} F(-n) = 0$.

Satz 3.35 (Stetigkeit von unten und oben)**.** *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Sind $A_1, A_2, \dots$ Mengen, $A_i \in \mathcal{A}$ für alle $i \in \mathbb{N}$, mit*

1. *$A_i \subseteq A_{i+1}$ für $i \in \mathbb{N}$ und $A := \bigcup_{n=1}^{\infty} A_n$ oder*
2. *$A_i \supseteq A_{i+1}$ für $i \in \mathbb{N}$ und $A := \bigcap_{n=1}^{\infty} A_n$,*

so ist $A \in \mathcal{A}$ und $\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$. ▪

Beweis. 1. Wir definieren $B_1 = A_1$ und $B_n = A_n \setminus A_{n-1}$ für $n \geq 2$. Dann sind für $n \in \mathbb{N}$ die B_n paarweise disjunkt und

$$A_n = \bigcup_{i=1}^{n} B_i, \quad A = \bigcup_{i=1}^{\infty} B_i.$$

Daraus folgt unmittelbar, dass

$$\mathbb{P}(A) = \mathbb{P}\big(\bigcup_{i=1}^{\infty} B_i\big) = \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) = \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n).$$

2. Es gilt für $n \in \mathbb{N}$ stets $A_n^{\complement} \subseteq A_{n+1}^{\complement}$ und

$$A^{\complement} = \left(\bigcap_{i=1}^{\infty} A_i\right)^{\complement} = \bigcup_{i=1}^{\infty} A_i^{\complement}.$$

Wenden wir nun 1. auf die Komplemente an, so gilt

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^{\complement}) = 1 - \lim_{n\to\infty} \mathbb{P}(A_n^{\complement}) = \lim_{n\to\infty} (1 - \mathbb{P}(A_n^{\complement})) = \lim_{n\to\infty} \mathbb{P}(A_n).$$ ∎

Satz 3.36. *a. Für jedes Wahrscheinlichkeitsmaß μ über $(\mathbb{R}, \mathcal{B})$ ist*

$$F_\mu(x) = \mu\big((-\infty, x]\big), \qquad x \in \mathbb{R}, \tag{3.7}$$

eine Verteilungsfunktion.
b. Umgekehrt ist für jede Verteilungsfunktion $F : \mathbb{R} \to \mathbb{R}$ das durch (3.7) definierte Maß μ eine Verteilung. ▪

Wir beweisen nur a. Für Teil b. würde an dieser Stelle ein Maßfortsetzungssatz benötigt, zum Beispiel der bekannte Satz von Carathéodory. Siehe Satz 1.42 in Doering (2022) für einen vollständigen Beweis.

Beweis. a. F_μ ist wachsend (Monotonie) und rechtsstetig (Stetigkeit von oben in $\emptyset$). Die Limeseigenschaften sind klar mit der Stetigkeit von oben und von unten. ∎

Für $a < b$ gilt mit Subtraktivität, dass

$$F_\mu(b) - F_\mu(a) = \mu\big((-\infty, b]\big) - \mu\big((-\infty, a]\big) = \mu\big((a, b]\big).$$

Ist $X : \Omega \to \mathbb{R}$ eine reellwertige Zufallsvariable, so ist die Funktion $F_X(x) = \mathbb{P}(X \leq x)$ eine Verteilungsfunktion, wir nennen diese auch die Verteilungsfunktion von X.

Korollar 3.37. Eine Verteilungsfunktion legt die Verteilung eindeutig fest.

Dies folgt direkt aus Satz 3.33 mit Satz 2.32.

Notation. Hat eine reellwertige Zufallsvariable X die Verteilung μ, gilt also $\mathbb{P}_X = \mu$, so schreiben wir $X \sim \mu$. Hat X bzw. μ die Verteilungsfunktion F, so schreiben wir auch $X \sim F$.

3.5 Absolutstetige Verteilungen

Eine diskrete Verteilung einer diskreten Zufallsvariable X wird durch $\mathbb{P}(X = x)$, für x aus der diskreten Bildmenge $X(\Omega)$, charakterisiert. Für die Verteilung eines stetigen Merkmals, zum Beispiel die Lebensdauer einer Glühbirne oder die Körpergröße einer zufällig ausgewählten Person, erwarten wir dagegen, dass die Wahrscheinlichkeiten, dass Beobachtungen in gewisse Intervalle fallen, positiv sind, und kleiner werden wenn das Intervall verkleinert wird. Im Grenzwert sollte dann jedem individuellen Ergebnis $x \in \mathbb{R}$ eine Wahrscheinlichkeit 0 zugeordnet werden. Nach dem bereits erwähnten Satz von Banach und Kuratowski gibt es kein solches Wahrscheinlichkeitsmaß auf $\mathcal{P}(\mathbb{R})$. Wir betrachten als Messraum nachfolgend $(\mathbb{R}, \mathcal{B})$, mit der Borel-$\sigma$-Algbera $\mathcal{B}$, oder Teilmengen davon. Dazu können wir uns die Verteilungsfunktionen aus dem letzten Kapitel noch einmal genauer anschauen.

Satz 3.38. *Ist $F : \mathbb{R} \to [0,1]$ eine Verteilungsfunktion, so existiert in jedem $x \in \mathbb{R}$ der linksseitige Limes*

$$\lim_{y \uparrow x} F(y) =: F(x-),$$

und die Menge der Unstetigkeitsstellen D_F von F ist gegeben durch $D_F = \{x \in \mathbb{R} : F(x) > F(x-)\}$. D_F ist höchstens abzählbar.

Beweis. Wegen der Monotonie von F ist

$$\lim_{y \uparrow x} F(y) =: \sup_{y<x} F(y),$$

somit ist die Existenz des Limes klar. Wegen der rechtsseitigen Stetigkeit und Monotonie ist F daher nur dann unstetig in $x \in \mathbb{R}$, falls $F(x) > F(x-)$. Um zu zeigen, dass D_F höchstens abzählbar unendlich ist, konstruieren wir (falls $D_F \neq \emptyset$) eine injektive Abbildung

$$\pi : D_F \to \mathbb{Q},$$

die Behauptung folgt dann aus der Abzählbarkeit von $\mathbb{Q}$. Für $x \in D_F$ wählen wir ein $q_x \in I_x := \big(F(x-), F(x)\big)$ und setzen $\pi(x) = q_x$. Sind $x, y \in D_F$, $x \neq y$, so gilt $I_x \cap I_y = \emptyset$, also sind $q_x \neq q_y$, was die Injektivität von π zeigt. ∎

Satz 3.39. *Ist μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ mit Verteilungsfunktion F_μ, so gilt*

$$\mu(\{x\}) = F_\mu(x) - F_\mu(x-), \qquad x \in \mathbb{R}.$$

Insbesondere ist $\mu(\{x\}) = 0$ für $x \in \mathbb{R} \setminus D_{F_\mu}$.
Ist X eine Zufallsvariable mit Verteilungsfunktion F_X, so ist

$$\mathbb{P}(X = x) = \mathbb{P}_X(\{x\}) = F_X(x) - F_X(x-).$$

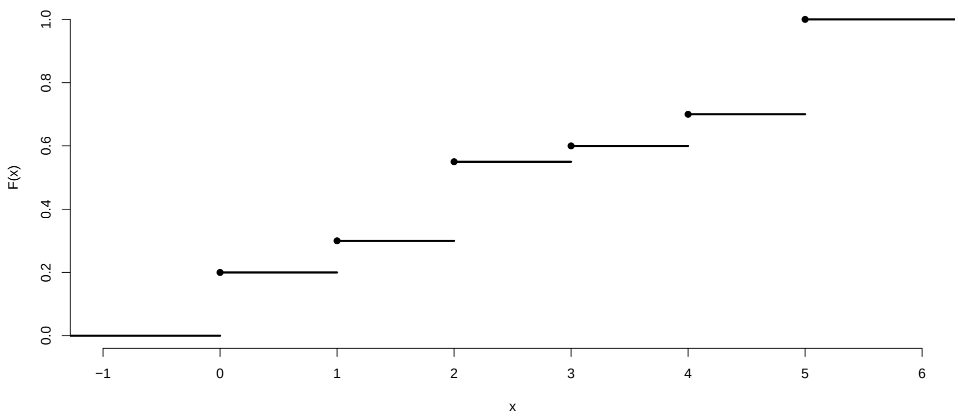


Abbildung 3.5: Graph einer diskreten Verteilungsfunktion $F(x)$.

Beweis. Ist $x_n = x - \frac{1}{n}$ für $n \in \mathbb{N}$, so gilt $(x_n, x] \supset (x_{n+1}, x]$ und $\bigcap_{i=1}^{\infty}(x_n, x] = \{x\}$, also

$$\mu(\{x\}) = \lim_{n\to\infty} \mu((x_n, x]) = \lim_{n\to\infty} \left(F_\mu(x) - F_\mu\left(x - \frac{1}{n}\right)\right) = F_\mu(x) - F_\mu(x-).$$ ∎

Für eine Verteilungsfunktion F eines diskreten Wahrscheinlichkeitsmaßes μ gilt

$$\sum_{x \in D_F} \big(F(x) - F(x-)\big) = 1.$$

Also gilt

$$\mu(D_{F_\mu}) = 1.$$

Beispiel 3.40. Die Unstetigkeitsstellen einer Verteilungsfunktion können dicht in $\mathbb{R}$ liegen. Als Beispiel sei $(q_n)_{n\geq 1}$ eine Abzählung von $\mathbb{Q}$. Dann ist

$$F(x) = \sum_n \frac{1}{2^n} 1_{[q_n,\infty)}(x)$$

eine Verteilungsfunktion mit $D_F = \mathbb{Q}$.

Definition 3.41. *Eine Verteilungsfunktion F, und die zugehörige Verteilung auf $(\mathbb{R}, \mathcal{B})$, heißen **(absolut)stetig**, falls eine integrierbare Funktion $f : \mathbb{R} \to [0,\infty)$ mit*

$$\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1, \quad \forall t \in \mathbb{R} : F(t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x \tag{3.8}$$

*existiert. f heißt **Wahrscheinlichkeitsdichte**, oder kurz **Dichte**. Wir nennen ein Wahrscheinlichkeitsmaß μ absolutstetig, falls F_μ absolutstetig ist, und eine Zufallsvariable X absolutstetig verteilt, falls F_X absolutstetig ist. Für die Dichte von X schreiben wir dann auch f_X, und $X \sim f_X$.*

Bemerkung. 1. Wir können hier davon ausgehen, dass f auf Intervallen RIEMANN-integrierbar ist und verstehen (3.8) als uneigentliches RIEMANN-Integral im Sinne der Analysis I. Tatsächlich kann die Definition auf Basis der Integrationstheorie später noch allgemeiner erfolgen.

2. Es gilt für $t \in \mathbb{R}$, $\Delta > 0$ und für stetiges f
$$\mathbb{P}(t < X \le t+\Delta) = \int_t^{t+\Delta} f(x)\,\mathrm{d}x \approx f(t)\cdot\Delta,$$
falls Δ klein. f bestimmt also die lokale Konzentration der Wahrscheinlichkeit bei t.
3. Gilt (3.8), so ist F stetig und an jeder Stetigkeitsstelle x von f differenzierbar mit der Ableitung $F'(x) = f(x)$. Insbesondere gilt für stetig verteilte X
$$\mathbb{P}_X(\{x\}) = \mathbb{P}(X = x) = F_X(x) - F_X(x-) = 0.$$

Bemerkung (Mischformen aus diskret und stetig). Es gibt Verteilungen, die weder diskret noch stetig sind. Ist beispielsweise $f : \mathbb{R} \to [0,\infty)$ eine Wahrscheinlichkeitsdichte, also
$$\int_{-\infty}^{\infty} f(t)\,\mathrm{d}t = 1,$$
und $p \in (0,1)$, dann ist eine Verteilungsfunktion, die nicht diskret und nicht absolutstetig ist, gegeben durch
$$F(x) = \begin{cases} p + (1-p)\int_{-\infty}^{x} f(t)\,\mathrm{d}t, & \text{falls } x \ge 0, \\ 0 & \text{falls } x < 0. \end{cases}$$
Als Modelle sind diese Formen von Verteilungen jedoch eher ungewöhnlich.

Wir führen nun die wichtigsten Familien von Dichtefunktionen und absolutstetigen Verteilungen ein. Im Folgenden sei jeweils $f : \mathbb{R} \to [0,\infty)$ eine Wahrscheinlichkeitsdichte, also $\int_{\mathbb{R}} f(x)\,\mathrm{d}x = 1$ und $X \sim f$ eine Zufallsvariable mit der Verteilungsfunktion $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$.

1. Ist $\mu \in \mathbb{R}$ und $Y = X + \mu$, so hat Y die Dichte $f_Y(x) = f(x-\mu)$ und die Verteilungsfunktion $F_Y(x) = F(x-\mu)$, denn
$$F_Y(t) = \mathbb{P}(Y \le t) = \mathbb{P}(X \le t-\mu) = F(t-\mu) = \int_{-\infty}^{t-\mu} f(x)\,\mathrm{d}x = \int_{-\infty}^{t} f(y-\mu)\,\mathrm{d}y.$$

Definition 3.42. *Eine Familie $(f(x-\mu))_{\mu\in\mathbb{R}}$ von Dichtefunktionen heißt* ***Lokationsfamilie****.*

2. Wir bezeichnen mit $\mathbb{R}^+ = (0,\infty)$ die positiven reellen Zahlen. Ist $\sigma \in \mathbb{R}^+$ und $Y = \sigma \cdot X$, so hat Y die Wahrscheinlichkeitsdichte $f_Y(x) = \frac{1}{\sigma} f(\frac{x}{\sigma})$ und die Verteilungsfunktion $F_Y(x) = F(\frac{x}{\sigma})$, denn
$$F_Y(t) = \mathbb{P}(Y \le t) = \mathbb{P}\Big(X \le \frac{t}{\sigma}\Big) = F\Big(\frac{t}{\sigma}\Big) = \int_{-\infty}^{\frac{t}{\sigma}} f(x)\,\mathrm{d}x = \int_{-\infty}^{t} \frac{1}{\sigma} f\Big(\frac{y}{\sigma}\Big)\,\mathrm{d}y.$$

Definition 3.43. *Eine Familie $(\frac{1}{\sigma} f(\frac{x}{\sigma}))_{\sigma\in\mathbb{R}^+}$ von Dichtefunktionen heißt* ***Skalenfamilie****.*

3. Ist $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^+$ und $Y = \sigma \cdot X + \mu$, so hat Y die Wahrscheinlichkeitsdichte $f_Y(x) = \frac{1}{\sigma} f(\frac{x-\mu}{\sigma})$ und die Verteilungsfunktion $F(\frac{x-\mu}{\sigma})$.

Definition 3.44. *Eine Familie $(\frac{1}{\sigma} f(\frac{x-\mu}{\sigma}))_{\mu\in\mathbb{R},\sigma\in\mathbb{R}^+}$ heißt* ***Lokations-Skalenfamilie****.*

Beispiel 3.45 (Uniforme Verteilung)**.** Die uniforme Verteilung auf $[-1,1]$ hat die Dichte- und Verteilungsfunktionen
$$f(x) = \frac{1}{2} 1_{[-1,1]}(x), \quad F(x) = \frac{x+1}{2} 1_{[-1,1]}(x) + 1_{(1,\infty)}(x).$$

Damit definieren wir eine Lokations-Skalenfamilie

$$f(x;\mu,\sigma) = \frac{1}{2\sigma}\,1_{[-1,1]}\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{2\sigma}\,1_{[\mu-\sigma,\mu+\sigma]}(x).$$

Eine Funktion aus dieser Familie ist Dichte einer uniformen Verteilung auf $[\mu-\sigma,\mu+\sigma]$, da

$$-1 \le \frac{x-\mu}{\sigma} \le 1 \Leftrightarrow \mu-\sigma \le x \le \mu+\sigma.$$

Die uniforme Verteilung auf einem Intervall $[a,b]$ hat damit die Dichte

$$f\left(x;\frac{a+b}{2},\frac{b-a}{2}\right) = \frac{1}{b-a}\,1_{[a,b]}(x).$$

Definition 3.46. *Für $a<b\in\mathbb{R}$ heißt die Verteilung mit der Dichtefunktion*

$$f(x) = \frac{1}{b-a}\,1_{[a,b]}(x)$$

***uniforme Verteilung** auf $[a,b]$. Ist eine Zufallsvariable X auf $[a,b]$ uniform verteilt, dann schreiben wir $X\sim\mathcal{U}[a,b]$.*

Beispiel 3.47 (Exponentialverteilung)**.** Die *Standard-Exponentialverteilung* hat die Dichte- und Verteilungsfunktionen

$$f(x) = \mathrm{e}^{-x}\,1_{[0,\infty)}(x), \quad F(x) = 1_{[0,\infty)}(x)\int_0^x \mathrm{e}^{-t}\,\mathrm{d}t = (1-\mathrm{e}^{-x})\,1_{[0,\infty)}(x).$$

Die Familie der Exponentialverteilungen ist nun die Skalenfamilie zu dieser Verteilung. Als Parameter verwendet man aber meist $\lambda=\frac{1}{\sigma}$. Damit ist die Exponential-λ-Verteilung definiert über

$$f(x;\lambda) = \lambda\mathrm{e}^{-\lambda x}\cdot 1_{[0,\infty)}(x), \quad F(x;\lambda) = (1-\mathrm{e}^{-\lambda x})\cdot 1_{[0,\infty)}(x).$$

Definition 3.48. *Für $\lambda\in\mathbb{R}^+$ heißt die Verteilung mit der Dichtefunktion*

$$f(x) = \lambda\mathrm{e}^{-\lambda x}\cdot 1_{[0,\infty)}(x)$$

***Exponentialverteilung** mit Parameter λ. Für exponentialverteilte Zufallsvariablen X schreibt man $X\sim\mathrm{Exp}(\lambda)$.*

Bemerkung. Analog zur geometrischen Verteilung gilt für $X\sim\mathrm{Exp}(\lambda)$ und $t,s\in\mathbb{R}^+$

$$\begin{aligned}\mathbb{P}(X>t+s\mid X>s) &= \frac{\mathbb{P}(X>t+s)}{\mathbb{P}(X>s)} = \frac{1-F(t+s;\lambda)}{1-F(s;\lambda)} = \frac{\mathrm{e}^{-\lambda(t+s)}}{\mathrm{e}^{-\lambda s}}\\ &= \mathrm{e}^{-\lambda t} = 1-F(t;\lambda) = \mathbb{P}(X>t).\end{aligned}$$

Eine solche Verteilung heißt *gedächtnislos*.

Beispiel 3.49 (Normalverteilung)**.** Die *Standard-Normalverteilung* hat die Dichte und die Verteilungsfunktion

$$\varphi(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right), \quad \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right)\mathrm{d}t.$$

Die Verteilungsfunktion $\Phi(x)$ kann nicht in geschlossener Form angegeben werden.

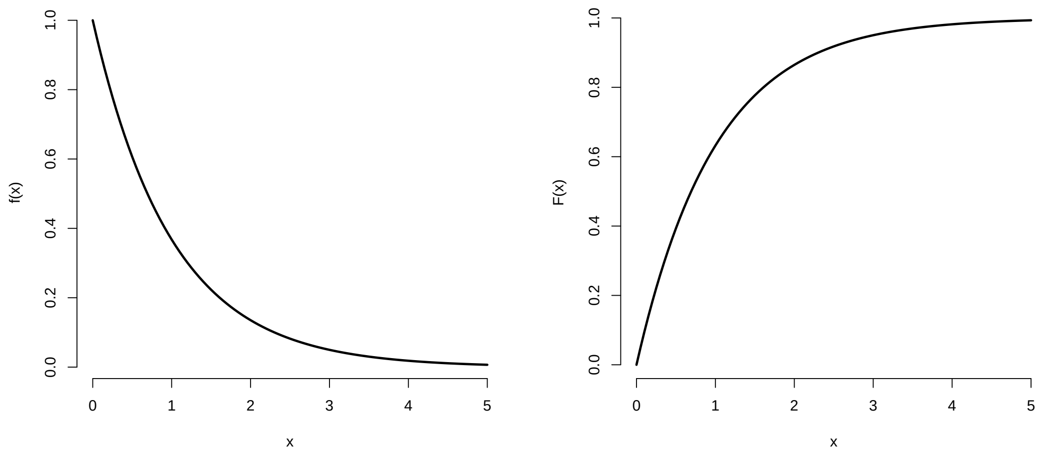


Abbildung 3.6: Dichte $f(x)$ und Verteilungsfunktion $F(x)$ der Standard-Exponentialverteilung.

Die *Familie der Normalverteilungen* ist nun die zugehörige Lokations-Skalenfamilie mit

$$f(x;\mu,\sigma)=\frac{1}{\sigma\sqrt{2\pi}}\exp\Big(-\frac{(x-\mu)^2}{2\sigma^2}\Big).$$

Definition 3.50. *Für $\mu\in\mathbb{R}$ und $\sigma\in\mathbb{R}^+$ heißt die Verteilung mit der Dichtefunktion*

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\Big(\frac{x-\mu}{\sigma}\Big)^2\right)$$

***Normalverteilung** mit Parametern μ und σ^2. Hat eine Zufallsvariable X diese Verteilung, so schreiben wir $X\sim\mathcal{N}(\mu,\sigma^2)$.*

Definition 3.51. *Eine Funktion $u:\mathbb{R}\to\mathbb{R}$ heißt **(Borel)-messbar**, falls sie $(\mathcal{B}-\mathcal{B})$-messbar ist, also falls $\forall B\in\mathcal{B}: u^{-1}(B)\in\mathcal{B}$.*

Bemerkung. Sei $X:\Omega\to\mathbb{R}$ eine reellwertige Zufallsvariable, also $\forall B\in\mathcal{B}:\{X\in B\}=X^{-1}(B)\in\mathcal{A}$. Ist die Funktion $u:\mathbb{R}\to\mathbb{R}$ Borel-messbar, so gilt auch

$$\{u(X)\in B\}=X^{-1}(\underbrace{u^{-1}(B)}_{\in\mathcal{B}})\in\mathcal{A},$$

also ist auch $u(X)$ eine reellwertige Zufallsvariable.

Satz 3.52 (Dichtetransformation)**.** *Ist $X\sim f$ absolutstetig, $u:\mathbb{R}\to\mathbb{R}$ stetig differenzierbar mit $u'(x)\neq 0$ für alle x, mit der Inversen $u^{-1}:u(\mathbb{R})\to\mathbb{R}$ und $Y=u(X)$, so hat Y die Dichte*

$$g(y)=\begin{cases} f(u^{-1}(y))\left|\frac{du^{-1}(y)}{dy}\right|=\frac{f(u^{-1}(y))}{|u'(u^{-1}(y))|}, & \text{falls } y\in u(\mathbb{R}),\\ 0, & \text{sonst.}\end{cases}$$

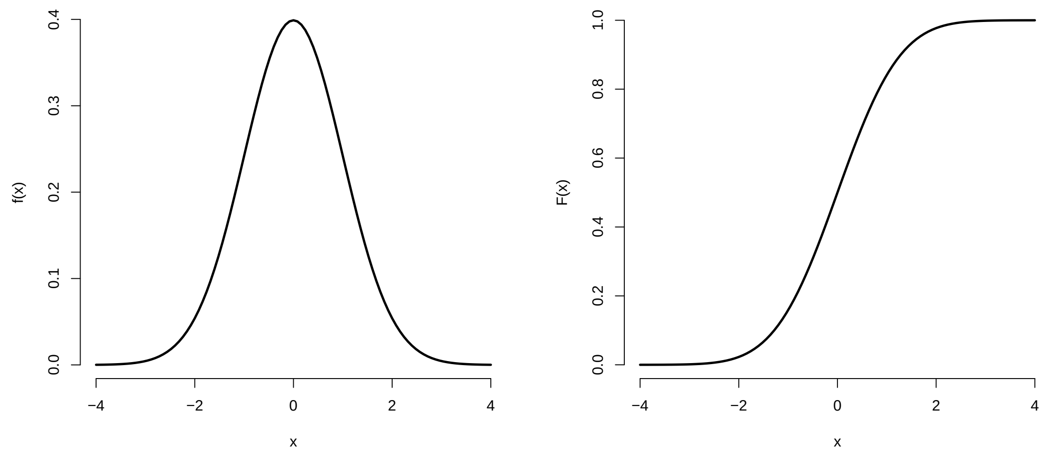


Abbildung 3.7: Dichte $\varphi(x)$ und Verteilungsfunktion $\Phi(x)$ der Standard-Normalverteilung.

Beweis. Unter obigen Voraussetzungen ist u streng monoton wachsend oder fallend und dann auch Borel-messbar, somit $u(X)$ eine Zufallsvariable. Sei u zunächst streng monoton wachsend. Da u stetig ist, ist $u(\mathbb{R})$ ein offenes Intervall. Da u streng monoton wächst, ist für $y \in u(\mathbb{R})$

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(u(X) \leq y) = \mathbb{P}(X \leq u^{-1}(y)) = F_X(u^{-1}(y)),$$

also für allgemeine $y \in \mathbb{R}$

$$F_Y(y) = \begin{cases} F_X(u^{-1}(y)), & \text{falls } y \in u(\mathbb{R}), \\ 0, & \text{falls } y \leq \inf u(\mathbb{R}), \\ 1, & \text{falls } y \geq \sup u(\mathbb{R}). \end{cases}$$

Leiten wir das ab, so erhalten wir für $y \in u(\mathbb{R})$

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_X(u^{-1}(y)) = \frac{\mathrm{d}}{\mathrm{d}u^{-1}(y)} F_X(u^{-1}(y)) \frac{\mathrm{d}}{\mathrm{d}y} u^{-1}(y) = f_X(u^{-1}(y)) \frac{\mathrm{d}}{\mathrm{d}y} u^{-1}(y),$$

also für allgemeine $y \in \mathbb{R}$

$$f_Y(y) = \begin{cases} f(u^{-1}(y)) \frac{\mathrm{d}}{\mathrm{d}y} u^{-1}(y), & \text{falls } y \in u(\mathbb{R}), \\ 0, & \text{sonst.} \end{cases}$$

Für u monoton fallend ist auch u^{-1} monoton fallend und mit der Ersetzung

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(u(X) \leq y) = \mathbb{P}(X \geq (u^{-1}(y))) = 1 - F_X(u^{-1}(y))$$

schließt man analog, wobei sich das Vorzeichen wechselt.
Die zweite Formel im Satz ergibt sich aus der Ableitung der Umkehrfunktion. ■

Beispiel 3.53. Ist $X \sim \mathcal{N}(\mu, \sigma^2)$, so heißt die Zufallsvariable $Y = \exp(X)$ *logarithmisch normalverteilt*. Wir schreiben dann $Y \sim \mathcal{LN}(\mu, \sigma^2)$. Die Dichte berechnet sich zu

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma y} \exp\left(-\frac{(\log(y) - \mu)^2}{2\sigma^2}\right), \qquad y > 0.$$

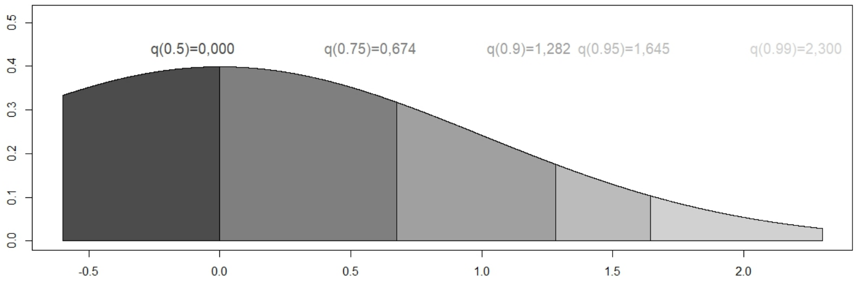


Abbildung 3.8: Quantile der Standard-Normalverteilung

Beispiel 3.54 (Dichten bei nicht-streng-monotoner Transformation)**.** Ist $X \sim f$ stetig, so ist für $t > 0$

$$F_{X^2}(t) = \mathbb{P}(X^2 \leq t) = \mathbb{P}(-\sqrt{t} \leq X \leq \sqrt{t}) = F_X(\sqrt{t}) - F_X(-\sqrt{t}).$$

Leiten wir das ab, so erhalten wir die Dichte

$$f_{X^2}(t) = \frac{\mathrm{d}}{\mathrm{d}t} F_{X^2}(t) = \frac{\mathrm{d}}{\mathrm{d}t}(F_X(\sqrt{t}) - F_X(-\sqrt{t})) = \frac{f_X(\sqrt{t}) + f_X(-\sqrt{t})}{2\sqrt{t}}.$$

Für $t \leq 0$ ist klar $F_{X^2}(t) = f_{X^2}(t) = 0$.

3.6 Quantile

Folgende Fragestellungen sind oft von Interesse: Gegeben $p \in (0,1)$, was ist der minimale Wert x, den eine Zufallsvariable X mit Wahrscheinlichkeit $\geq p$ nicht überschreitet?

Definition 3.55. *Ist X eine Zufallsvariable mit Verteilungsfunktion F und $p \in (0,1)$, so heißt*

$$Q_p(F) := F^{-1}(p) := \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$$

*das p-**Quantil** von X bzw. F. Die verallgemeinerte Inverse (Pseudoinverse) F^{-1} heißt die **Quantilfunktion**. Insbesondere nennt man im Fall $p = 1/2$ das zugehörige Quantil den **Median** der Verteilung. Im Falle $p = k/4, k = 1,3$, spricht man von den **Quartilen**.*

Bemerkung.

1. Ist F stetig und streng monoton, so gilt mit der existierenden Umkehrfunktion $F^{-1} : (0,1) \to \mathbb{R}$ für das p-Quantil $Q_p(F) = F^{-1}(p)$.
2. Hat F eine Sprungstelle, sodass kein x mit $F(x) = p$ existiert, so ist $Q_p(F)$ die Stelle, an der F den Wert p "überspringt".
3. Ist F konstant auf einem Intervall, sodass mehrere x mit $F(x) = p$ existieren, so ist $Q_p(F)$ die kleinste Stelle, an der dies gilt.

Abbildung 3.8 zeigt Quantile $q(p)$ einer Standard-Normalverteilung für Werte p, die häufig in der Statistik benötigt werden.

Satz 3.56. *Seien $x \in \mathbb{R}$ und $p \in (0,1)$. Dann gilt*

$$x \geq Q_p(F) \Leftrightarrow F(x) \geq p.$$

Insbesondere folgt (für $x = Q_p(F)$)

$$F(Q_p(F)) \geq p.$$

Beweis. 1. Wenn $F(x) \geq p$ gilt, dann ist wegen $x \in \{y \mid F(y) \geq p\}$

$$x \geq \inf\{y \mid F(y) \geq p\} = Q_p(F).$$

2. Wir zeigen erst $F(Q_p(F)) \geq p$. Dies folgt aus der rechtsseitigen Stetigkeit von F: Ist

$$t_n \in \{y \mid F(y) \geq p\}, \quad \lim_{n\to\infty} t_n = Q_p(F),$$

also insbesondere $F(t_n) \geq p$, dann folgt

$$F(Q_p(F)) = \lim_{n\to\infty} F(t_n) \geq p.$$

Ist nun $x \geq Q_p(F)$, so folgt aus der Monotonie von F, dass

$$F(x) \geq F(Q_p(F)) \geq p.$$

■

Beispiel 3.57. Quantile spielen im Risikocontrolling eine große Rolle: Der *Value at risk (VaR)* eines Finanz-Portfolios zum Zeitpunkt t_0 über einen Zeitraum T (meist ein oder 10 Tage) zum Niveau $\alpha \in (0,1)$ (meist nahe bei 1) ist der minimale Wert, den der Verlust des Portfolios in $t_0 + T$ mit Wahrscheinlichkeit α nicht überschreitet. Formal ist der VaR also das α-Quantil der Verlustverteilung zum Zeitpunkt $T + t_0$.

Satz 3.58 (Quantile unter linear-affiner Transformation)**.** *Sei X eine Zufallsvariable mit Verteilungsfunktion F_X, $Y = \sigma X + \mu$ für $\sigma > 0$, $\mu \in \mathbb{R}$. Dann gilt für $p \in (0,1)$*

$$Q_p(F_Y) = \sigma Q_p(F_X) + \mu.$$

Beweis. Es ist $F_Y(y) = F_X\big((y-\mu)/\sigma\big)$, $y \in \mathbb{R}$, bzw. $F_Y(\sigma x + \mu) = F_X\big(x\big)$, $x \in \mathbb{R}$. Daher ist

$$\begin{aligned}\big\{y \in \mathbb{R} \mid F_Y(y) \geq p\big\} &= \big\{\sigma x + \mu; x \in \mathbb{R} \mid F_Y(\sigma x + \mu) \geq p\big\} \\ &= \sigma\big\{x \in \mathbb{R} \mid F_X(x) \geq p\big\} + \mu,\end{aligned}$$

wobei für $A \subseteq \mathbb{R}$ gesetzt wird $\sigma A + \mu = \{\sigma a + \mu;\ a \in A\}$. Anwendung des Infimums liefert die Behauptung. ■

Die Quantiltransformation wird zur Generierung von Zufallszahlen eingesetzt. Dies erfolgt in zwei Schritten: Zunächst werden Pseudo-Zufallszahlen erzeugt, die auf dem Intervall $(0,1)$ uniform verteilt sind. Danach werden diese mit Hilfe des folgenden Satzes transformiert, um Zufallszahlen mit einer gewünschten Verteilungsfunktion F zu erhalten.

Satz 3.59. *Ist F eine Verteilungsfunktion und $U \sim \mathcal{U}(0,1)$ uniform verteilt auf $(0,1)$, so ist $F^{-1}(U) \sim F$, d.h. $F^{-1}(U)$ hat Verteilungsfunktion F.*

Beweis. Aus Satz 3.56 folgt

$$F^{-1}(U) \leq x \quad \Leftrightarrow \quad U \leq F(x),$$

also $\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x)$. ■

Für stetige Verteilungsfunktionen gilt eine Umkehrung.

Satz 3.60. *Ist X eine Zufallsvariable mit stetiger Verteilungsfunktion F, so ist $F(X) \sim \mathcal{U}[0,1]$ uniform verteilt auf $(0,1)$.*

Beweis. Wir zeigen zunächst, dass für nicht notwendig stetiges F und $p \in (0,1)$ gilt

$$p \in F(\mathbb{R}) = \{F(x) | \, x \in \mathbb{R}\} \quad \Leftrightarrow \quad F\big(F^{-1}(p)\big) = p. \tag{3.9}$$

Dabei ist "$\Leftarrow$" klar, da für $x = F^{-1}(p)$ gilt dass $F(x) = p$. Für die Umkehrung sei $x \in \mathbb{R}$ mit $F(x) = p$, also $x \in \{y \mid F(y) \geq p\}$, daher mit Satz 3.56

$$x \geq F^{-1}(p) \quad \Rightarrow \quad p = F(x) \geq F\big(F^{-1}(p)\big) \geq p.$$

Daher stehen überall Gleichheitszeichen und (3.9) folgt. Wir folgern nun die Aussage des Satzes: Da die Verteilung von $F(X)$ nur von der Verteilung von X und der Transformation F abhängt, aber nicht von der Definition von X als Abbildung, vgl. zu der Bemerkung unter Definition 3.51, können wir nach Satz 3.59 wählen $X = F^{-1}(U)$, $U \sim \mathcal{U}(0,1)$. Da F stetig ist, folgt mit den Grenzwerteigenschaften einer Verteilungsfunktion und dem Zwischenwertsatz $(0,1) \subseteq F(\mathbb{R})$, also nimmt U stets Werte in $F(\mathbb{R})$ an, und nach (3.9) folgt $F\big(F^{-1}(U)\big) = U$. ■

Satz 3.61. *Sei F eine Verteilungsfunktion und $F^{-1} : (0,1) \to \mathbb{R}$ die zugehörige Quantilfunktion. Es gelten:*

1. *F^{-1} ist monoton wachsend*
2. *F^{-1} ist linksseitig stetig*
3. *F ist stetig $\quad \Leftrightarrow \quad$ F^{-1} ist streng monoton wachsend.*

Beweis. Zu 1.: Seien $p, q \in (0,1)$, $p < q$, dann gilt

$$\{x \mid F(x) \geq p\} \supseteq \{x \mid F(x) \geq q\},$$

daher

$$F^{-1}(p) = \inf\{x \mid F(x) \geq p\} \leq \inf\{x \mid F(x) \geq q\} = F^{-1}(q).$$

Zu 2.: Seien $p_n, p \in (0,1)$, $p_n < p$ und $p_n \uparrow p$, $n \to \infty$. Wegen der Monotonie von F^{-1} gilt für ein $x \in \mathbb{R}$

$$F^{-1}(p_n) \uparrow x \leq F^{-1}(p).$$

Nach Satz 3.56 ist

$$\forall\, n: \quad p_n \leq F\big(F^{-1}(p_n)\big) \leq F(x),$$

im Grenzwert für $n \to \infty$ also auch $p \leq F(x)$, daher mit Satz 3.56 auch $x \geq F^{-1}(p)$, insgesamt $x = F^{-1}(p)$, also die linksseitige Stetigkeit von F in p.

Zu 3.: "$\Rightarrow$" Ist F stetig, also $(0,1) \subseteq F(\mathbb{R})$, so ist nach (3.9) für $p, q \in (0,1)$ mit $p < q$ auch $p = F\big(F^{-1}(p)\big) < F\big(F^{-1}(q)\big) = q$, daher muss $F^{-1}(p) < F^{-1}(q)$ sein.

"$\Leftarrow$" Angenommen, F ist nicht stetig in x, also $F(x-) < F(x)$, dann gilt für alle $p \in \big(F(x-), F(x)\big]$, dass $F^{-1}(p) = x$ und daher ist F^{-1} nicht streng monoton wachsend.

■

4 Maßintegral und Erwartungswert

4.1 Diskreter Erwartungswert und Maßintegral von Treppenfunktionen

Beispiel 4.1. Wir setzen beim Roulette einen Euro auf eine gerade Zahl. Was ist unser mittlerer oder zu erwartender Gewinn?

Zunächst ergeben sich folgende Gewinnmöglichkeiten:

- Falls eine positive gerade Zahl fällt, bekommen wir den gesetzten Euro zurück und einen weiteren Euro Gewinn.
- Sonst bekommen wir nichts zurück und haben einen Euro verloren.

Beim Roulette ist der endliche Wahrscheinlichkeitsraum der LAPLACE-Raum mit Ergebnismenge

$$\Omega = \{0, 1, \ldots, 36\}, \quad \operatorname{card}\Omega = 37.$$

In diesem Szenario stellt die Zufallsvariable

$$X : \Omega \to \{-1, 1\}, \quad X(\omega) = \begin{cases} 1, & \text{falls } \omega \in \{2, 4, \ldots, 36\}, \\ -1, & \text{sonst,} \end{cases}$$

den Gewinn dar. Unser zu erwartender Gewinn ist nun

$$\begin{aligned} \mathbb{E}[X] &= \sum_{\omega=0}^{36} X(\omega) \cdot \mathrm{p}(\omega) \\ &= \sum_{j=1}^{18} X(2j) \cdot \mathrm{p}(2j) + \sum_{j=1}^{18} X(2j-1) \cdot \mathrm{p}(2j-1) + X(0) \cdot \mathrm{p}(0) \\ &= \sum_{j=1}^{18} \mathrm{p}(2j) - \sum_{j=1}^{18} \mathrm{p}(2j-1) - \mathrm{p}(0) = \frac{18}{37} - \frac{18}{37} - \frac{1}{37} = -\frac{1}{37} \end{aligned}$$

oder über die Verteilung von X

$$\mathbb{E}[X] = \sum_{x \in X(\Omega)} x \cdot \mathbb{P}(X = x) = 1 \cdot \mathbb{P}(X = 1) + (-1) \cdot \mathbb{P}(X = -1) = \frac{18}{37} - \frac{19}{37} = -\frac{1}{37}.$$

Das im Beispiel berechnete nach Wahrscheinlichkeit gewichtete Mittel der Werte, die eine Zufallsvariable annehmen kann, wird als *Erwartungswert* der Zufallsvariable bezeichnet. Wir können die Definition auf diskrete Zufallsvariablen direkt erweitern:

Ist $(\Omega, \mathbb{P})$ ein diskreter Wahrscheinlichkeitsraum mit Wahrscheinlichkeitsfunktion p und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable, dann gilt

1. Ist X nicht-negativ, also $X(\omega) \geq 0$ für alle $\omega \in \Omega$, so ist der *Erwartungswert von X* gegeben durch

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega) \cdot \mathrm{p}(\omega), \ \mathbb{E}[X] \in [0, \infty]. \tag{4.1}$$

   Ist dabei $\mathbb{E}[X] < \infty$, so sagen wir "X hat endlichen Erwartungswert".

2. Ist X nicht notwendigerweise nicht-negativ, aber $\mathbb{E}[|X|] < \infty$, setzen wir

$$\mathbb{E}[X] := \sum_{\omega \in \Omega} X(\omega) \cdot \mathrm{p}(\omega), \ \mathbb{E}[X] \in \mathbb{R}, \tag{4.2}$$

und sagen, dass "der Erwartungswert von X existiert".

In positiven Reihen kann umgeordnet werden, sodass (4.1) wohldefiniert ist. Ist $\mathbb{E}[|X|] < \infty$, so ist (4.2) absolut konvergent, da gilt

$$\mathbb{E}[|X|] = \sum_{\omega\in\Omega} |X(\omega)| \cdot \mathrm{p}(\omega) = \sum_{\omega\in\Omega} |X(\omega)| \cdot \mathrm{p}(\omega) < \infty.$$

Damit kann wieder umgeordnet werden, der Wert ist also wohldefiniert.

Als Spezialfall der später entwickelten Transformationsformel erhalten wir für X eine diskrete, reellwertige Zufallsvariable mit der diskreten Verteilung $\mathbb{P}_X$, mit zugehöriger Wahrscheinlichkeitsfunktion p_X, und $x_1, x_2, \dots$ eine Abzählung des Wertebereichs $X(\Omega)$, dass $\mathbb{E}[|X|] < \infty$ genau dann, wenn

$$\sum_{i=1}^{\infty} |x_i| \cdot \mathrm{p}_X(x_i) < \infty. \tag{4.3}$$

Falls $\mathbb{E}[|X|] < \infty$ oder $X \geq 0$, so gilt

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} x_i \cdot \mathrm{p}_X(x_i). \tag{4.4}$$

Dies kann hier durch Umordnen in den Reihen elementar nachgerechnet werden.

Beispiel 4.2. 1. Ist $X \sim \mathrm{Poi}(\lambda)$, so gilt

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k \cdot \mathrm{e}^{-\lambda} \cdot \frac{\lambda^k}{k!} = \mathrm{e}^{-\lambda} \cdot \lambda \cdot \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \mathrm{e}^{-\lambda} \cdot \lambda \cdot \mathrm{e}^{\lambda} = \lambda.$$

2. Ist $X \sim \mathrm{Geom}(p)$, so gilt

$$\mathbb{E}[X] = \sum_{k=1}^{\infty} k \cdot \mathbb{P}(X = k) = \sum_{k=1}^{\infty} k \cdot p(1-p)^{k-1}.$$

Es gilt weiterhin für $|x| < 1$

$$\frac{1}{1-x} = \sum_{k=0}^{\infty} x^k \Rightarrow \frac{\mathrm{d}}{\mathrm{d}x} \frac{1}{1-x} = \frac{1}{(1-x)^2} = \sum_{k=0}^{\infty} k \cdot x^{k-1} = \frac{\mathrm{d}}{\mathrm{d}x} \sum_{k=0}^{\infty} x^k.$$

Daraus folgt

$$\mathbb{E}[X] = p \cdot \sum_{k=0}^{\infty} k(1-p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

Wir entwickeln in diesem Kapitel parallel die Begriffe *Maßintegral* und Erwartungswert wobei wir feststellen werden, dass letzterer ein spezielles Maßintegral ist. Maßintegral bedeutet, dass wir für einen Maßraum $(\Omega, \mathcal{A}, \mu)$ ein Integral $\int_\Omega f \, d\mu$ für bestimmte Integranden f konstruieren wollen. Wichtige Spezialfälle sind das Lebesgueintegral, wenn μ das Lebesguemaß ist, und eben Erwartungswerte. Typischerweise beginnt so eine Konstruktion mit möglichst einfachen Integranden f.

Definition 4.3. *Eine messbare Funktion $f : \Omega \to \mathbb{R}$, die nur endlich viele Werte annimmt, heißt nachfolgend* ***Treppenfunktion****. Die Menge der Treppenfunktionen wird mit $\mathcal{T}$ bezeichnet, die Menge der nicht-negativen Treppenfunktionen mit $\mathcal{T}^+$.*

1. $\mathcal{T}$ ist ein Vektorraum über $\mathbb{R}$, für $f, g \in \mathcal{T}$ sind $f \cdot g \in \mathcal{T}$, $\max(f,g), \min(f,g) \in \mathcal{T}$ und $|f| \in \mathcal{T}$.
2. Für $f, g \in \mathcal{T}^+$, $\alpha \geq 0$ sind $\alpha f \in \mathcal{T}^+$, $f \cdot g \in \mathcal{T}^+$, $\max(f,g), \min(f,g) \in \mathcal{T}^+$, $f + g \in \mathcal{T}^+$. Sind

$\{\alpha_1,\dots,\alpha_m\} \subset \mathbb{R}$ die verschiedenen Werte von f und ist $A_j = f^{-1}(\alpha_j)$, so gilt

$$f = \sum_{j=1}^{m} \alpha_j 1_{A_j}.$$

Dies wird *Standarddarstellung einer Treppenfunktion* genannt. Allgemeiner betrachten wir Darstellungen $f = \sum_{i=1}^{n} \beta_i 1_{B_i}$, mit disjunkten Mengen $B_i \in \mathcal{A}$, für die $\bigcup_i B_i = \Omega$, also bezüglich disjunkten Zerlegungen von Ω. Ist $g = \sum_{j=1}^{k} \gamma_j 1_{C_j}$ eine Darstellung einer weiteren Treppenfunktion g, so können wir zur *gemeinsamen Verfeinerung* $B_i \cap C_j$ der Zerlegungen übergehen, und erhalten Darstellungen

$$f = \sum_{i,j} \beta_i 1_{B_i \cap C_j}, \qquad g = \sum_{i,j} \gamma_j 1_{B_i \cap C_j},$$

bezüglich einer gemeinsamen Zerlegung.

Definition 4.4. *Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $f \in \mathcal{T}^+$, etwa $f = \sum_{i=1}^{n} \alpha_i 1_{A_i}$ mit disjunkten Mengen $A_i \in \mathcal{A}$ für die $\bigcup_{i=1}^{\infty} A_i = \Omega$, also einer disjunkten Zerlegung von Ω, und $\alpha_i \geq 0$, so setzten wir*

$$\int_\Omega f\,d\mu = \sum_{i=1}^{n} \alpha_i \mu(A_i), \ \int_\Omega f\,d\mu \in [0,\infty]. \tag{4.5}$$

Als Kurzschreibweise für $\int_\Omega f\,d\mu$ verwenden wir auch $\int f\,d\mu$ oder $\int f$, falls der Maßraum klar ist.

Wohldefiniertheit: Zunächst ist die Summe in (4.5) immer definiert (auch wenn $\mu(A_i) = \infty$ für einige i), da alle $\alpha_i \geq 0$. Es bleibt zu zeigen, dass (4.5) nicht von der Darstellung von f als Treppenfunktion abhängt, also von der Wahl der Zerlegung. Sei dazu eine weitere Darstellung von f als Treppenfunktion gegeben, etwa $f = \sum_{j=1}^{m} \beta_j 1_{B_j}$, wobei $\beta_j \geq 0$, $B_j \in \mathcal{A}$ disjunkt und $\Omega = \bigcup_{j=1}^{\infty} B_j$. Dann gilt auch

$$f = \sum_{i,j} \alpha_i 1_{A_i \cap B_j} = \sum_{i,j} \beta_j 1_{A_i \cap B_j},$$

und daher $\alpha_i = \beta_j$ falls $A_i \cap B_j \neq \emptyset$, insbesondere also falls $\mu(A_i \cap B_j) > 0$. Wir erhalten

$$\begin{aligned} \sum_{i=1}^{n} \alpha_i \mu(A_i) &= \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{m} \mu(A_i \cap B_j) = \sum_{i,j} \alpha_i \mu(A_i \cap B_j) \\ &= \sum_{i,j} \beta_j \mu(A_i \cap B_j) = \sum_{j=1}^{m} \beta_j \mu(B_j). \end{aligned}$$

Beispiel 4.5. Wir werfen einen fairen Würfel zweimal und X bezeichne wie in Beispiel 3.1 und 3.13 die Augensumme $X(\omega) = \omega_1 + \omega_2$. Die Zufallsvariable X erfüllt $X \in \mathcal{T}^+$ und wir sehen, dass

$$\mathbb{E}[X] = \int_\Omega X\,d\mathbb{P} = \frac{2\cdot1+3\cdot2+4\cdot3+5\cdot4+6\cdot5+7\cdot6+8\cdot5+9\cdot4+10\cdot3+11\cdot2+12\cdot1}{36} = 7.$$

Satz 4.6. *1. Für $A \in \mathcal{A}$ ist $\int 1_A\,d\mu = \mu(A)$.*

*2. **Linearität**: Für $f,g \in \mathcal{T}^+$, $a,b \geq 0$ gilt*

$$\int (af+bg)\,d\mu = a\int f d\mu + b\int g\,d\mu.$$

*3. **Monotonie**: Für $f,g \in \mathcal{T}^+$, $f \leq g$ gilt $\int f\,d\mu \leq \int g\,d\mu$.* ∎

Beweis. 1. klar nach Def., 2. mit Übergang zu gemeinsamer Darstellung von f und g als Treppenfunktion.

3. $g = f + (g-f)$ mit $g - f \in \mathcal{T}^+$, dann klar nach 2. und $\int (g-f)\,d\mu \geq 0$. ■

Beispiel 4.7. 1. Betrachten wir eine BERNOULLI-p-verteilte Zufallsvariable Y_1. Da der Erwartungswert über die Verteilung bestimmt werden kann, kann Y_1 auf einem beliebigen Wahrscheinlichkeitsraum definiert sein. Es gilt, dass

$$\mathbb{E}[Y_1] = 1 \cdot p + 0 \cdot (1-p) = \int_{\mathbb{R}} Y_1 \, d\mathbb{P}_{Y_1} = p\,.$$

Sei $X \sim \mathrm{Bin}(n,p)$. Wir können X darstellen als

$$X = \sum_{i=1}^{n} Y_i$$

mit BERNOULLI-p-verteilten Zufallsvariablen Y_i, die den Wert 1 annehmen, falls im i-ten Experiment ein Erfolg auftritt. Es gilt mit Linearität also

$$\mathbb{E}[X] = \mathbb{E}\Big[\sum_{i=1}^{n} Y_i\Big] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} p = n \cdot p\,.$$

2. Wir bestimmen den Erwartungswert von $\mathrm{Hyp}(\cdot; n, R, N)$. Es kann im LAPLACE-Raum

$$\Omega = \{A \subseteq \{1, \dots, N\} \mid \operatorname{card} A = n\}$$

eine Zufallsvariable $Y : \Omega \to \{\max(0, n-N+R), \dots, \min(n,R)\}$ durch

$$Y = \sum_{j=1}^{R} 1_{\{j \in A\}}, \quad \mathbb{P}(j \in A) = \frac{n}{N}$$

definiert werden mit $Y \sim \mathrm{Hyp}(\cdot; n, R, N)$. Es folgt mit Linearität, dass

$$\mathbb{E}[Y] = \mathbb{E}\Big[\sum_{j=1}^{R} 1_{\{j \in A\}}\Big] = \sum_{j=1}^{R} \mathbb{E}[1_{\{j \in A\}}] = \sum_{j=1}^{R} \mathbb{P}(j \in A) = \sum_{j=1}^{R} \frac{n}{N} = \frac{R \cdot n}{N}.$$

4.2 Fortsetzung des Maßintegrals und Erwartungswert

Wir konstruieren aufbauend auf dem Maßintegral für nicht-negative Treppenfunktionen zunächst das Maßintegral zur Integration nicht-negativer Funktionen. Sei weiterhin $(\Omega, \mathcal{A}, \mu)$ ein Maßraum. Wir bezeichnen nachfolgend mit $\mathcal{M}$ die Menge der messbaren numerischen Funktionen und mit

$$\mathcal{M}^+ := \{f : \Omega \to [0, \infty] \text{ messbar}\}$$

die Menge der nicht-negativen messbaren numerischen Funktion. Zu $f \in \mathcal{M}^+$ existiert stets eine monoton wachsende Folge von nicht-negativen Treppenfunktionen $(\varphi_n) \subseteq \mathcal{T}^+$ mit $\varphi_n \nearrow f$.

Satz 4.8. *Ist $f \in \mathcal{M}^+$, so existiert eine monoton wachsende Folge $(f_n) \subseteq \mathcal{T}^+$ von nicht-negativen Treppenfunktionen, die punktweise gegen f konvergiert: $f_n \nearrow f$.*

Beweis. Setze für $n \in \mathbb{N}$

$$A_{j,n} = \begin{cases} \{\frac{j}{2^n} \leq f < \frac{j+1}{2^n}\}, & j = 0, 1, \dots, n2^n - 1, \\ \{f \geq n\}, & j = n2^n \end{cases}.$$

Für festes n sind die Mengen $A_{j,n}$, $j = 0, 1, \dots n2^n$, eine disjunkte Zerlegung von Ω, und

$$f_n = \sum_{j=0}^{n2^n} \frac{j}{2^n} 1_{A_{j,n}} \in \mathcal{T}^+ .$$

a. (f_n) ist wachsend: $A_{j,n}$ ist disjunkte Vereinigung von $A_{2j,n+1}$ und $A_{2j+1,n+1}$ für $j = 0, 1, \dots, n2^n - 1$, und $A_{n2^n,n}$ ist disjunkte Vereinigung von $A_{j,n+1}$ für $j = n2^{n+1}, \dots, (n+1)2^{n+1}$. Somit $f_{n+1} \geq f_n$.
b. $f_n \to f$: Ist $x \in \Omega$ mit $f(x) < \infty$ und $n \geq f(x)$, so ist $f_n(x) \leq f(x) \leq f_n(x) + 2^{-n}$. Ist dagegen $f(x) = \infty$, so ist $f_n(x) = n \to \infty$. ■

Die Folge der Integrale $\int \varphi_n \, d\mu$ ist dann monoton wachsend, daher existiert $\lim_{n\to\infty} \int \varphi_n \, d\mu \in [0, \infty]$.

Definition 4.9. *Wir setzen*

$$\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega \varphi_n \, d\mu \quad (\in [0, \infty]). \tag{4.6}$$

Für Wohldefiniertheit bleibt zu zeigen, dass (4.6) unabhängig von der Wahl der approximierenden Folge von Treppenfunktionen ist.

Satz 4.10. *Sind $(u_n), (v_n) \subseteq \mathcal{T}^+$ wachsende Folgen von Treppenfunktionen mit $\lim_{n\to\infty} u_n = \lim_{n\to\infty} v_n$, so ist $\lim_{n\to\infty} \int u_n \, d\mu = \lim_{n\to\infty} \int v_n \, d\mu$. Insbesondere ist das Integral in (4.6) wohldefiniert.* ▪

Der Satz folgt aus der einseitigen Version in folgendem Lemma.

Lemma 4.11. *Ist $(u_n) \subseteq \mathcal{T}^+$ eine wachsende Folge von Treppenfunktionen, $v \in \mathcal{T}^+$, mit $v \leq \lim_{n\to\infty} u_n$, so gilt dass $\int v \, d\mu \leq \lim_{n\to\infty} \int u_n \, d\mu$.*

Beweis. Für $\beta > 1$ setze

$$B_n = \{\beta u_n \geq v\}.$$

Dann gelten (Beweis folgt)

$$B_n \subseteq B_{n+1}, \qquad \bigcup_n B_n = \Omega, \qquad \beta u_n \geq v 1_{B_n}. \tag{4.7}$$

Die Inklusion $B_n \subseteq B_{n+1}$ ist klar, da u_n wachsend. Ist $x \in \Omega$ mit $v(x) = 0$, so ist $x \in B_n$ für alle n. Ist dagegen $v(x) > 0$, so ist $\beta \lim_{n\to\infty} u_n(x) = \lim_{n\to\infty} \big(\beta u_n(x)\big) > v(x)$, also $x \in B_n$ für alle $n \geq n(x)$. Der letzte Teil ist klar nach Def. von B_n.

Schreibe $v = \sum_{j=1}^m \alpha_j 1_{A_j}$, mit $\alpha_1, \dots, \alpha_m \geq 0$, und $A_1, \dots, A_m \in \mathcal{A}$ disjunkte Zerlegung von Ω. Sei $\beta > 1$ fest. Mit (4.7), der Stetigkeit von μ von unten und der Monotonie des Integrals für Treppenfunktionen folgt

$$\begin{aligned}
\int_\Omega v \, d\mu &= \sum_{j=1}^m \alpha_j \mu(A_j) = \sum_{j=1}^m \alpha_j \lim_{n\to\infty} \mu(A_j \cap B_n) && \text{(Stetigkeit von unten)} \\
&= \lim_{n\to\infty} \sum_{j=1}^m \alpha_j \mu(A_j \cap B_n) = \lim_{n\to\infty} \int_\Omega v 1_{B_n} \, d\mu \\
&\leq \lim_{n\to\infty} \int_\Omega \beta \, u_n \, d\mu = \beta \lim_{n\to\infty} \int_\Omega u_n \, d\mu && \text{(Monotonie)} .
\end{aligned}$$

Da $\beta > 1$ beliebig, folgt die Behauptung. ■

Beweis von Satz 4.10. Da für jedes k, $v_k \leq \lim_{n\to\infty} u_n$ gilt, ist nach dem Lemma $\int v_k\,d\mu \leq \lim_{n\to\infty}\int u_n\,d\mu$, also

$$\lim_{k\to\infty}\int v_k\,d\mu \leq \lim_{n\to\infty}\int u_n\,d\mu\,,$$

und die Gleichheit folgt durch Vertauschung der Rollen von v_n und u_n. ■

Wir halten einige elementare Eigenschaften des Integrals fest.

Satz 4.12. *1.* ***Linearität*** *Für $f,g \in \mathcal{M}^+$, $a,b \in [0,\infty]$ gilt*

$$\int (af+bg)\,d\mu = a\int f\,d\mu + b\int g\,d\mu.$$

2. ***Monotonie****: Für $f,g \in \mathcal{M}^+$, $f \leq g$ gilt $\int f\,d\mu \leq \int g\,d\mu$.*

Beweis. 1. Für $a,b \in [0,\infty)$ folgt dies unmittelbar aus der Linearität des Integrals für Treppenfunktionen und der Limesbildung in (4.6). Dies lässt sich elementar verallgemeinern. Eine ausführliche Behandlung des allgemeinen Falls findet sich in Elstrodt (1996), Seiten 122–123.
2. Sind $(\varphi_n),(\psi_n) \subseteq \mathcal{T}^+$, $\varphi_n \nearrow f$, $\psi_n \nearrow g$, so ist auch $\gamma_n = \max(\varphi_n,\psi_n) \in \mathcal{T}^+$, $\gamma_n \nearrow g$, und $\int \varphi_n\,d\mu \leq \int \gamma_n\,d\mu$ wegen der Monotonie für Treppenfunktionen. Somit folgt 2. durch Monotonie für den Limes (4.6). ■

Korollar 4.13. Für $f \in \mathcal{M}^+$ gilt

$$\int_\Omega f\,d\mu = \sup\Big\{\int_\Omega \varphi\,d\mu :\ \varphi \in \mathcal{T}^+, \varphi \leq f\Big\}. \tag{4.8}$$

Beweis. "$\geq$" ist klar wegen Monotonie des Integrals, und "$=$" folgt aus (4.6). ■

Bemerkung: (4.8) kann auch zur Integraldefinition benutzt werden. Ein Vorteil davon wäre, dass dieses unmittelbar wohldefiniert ist. Ein Nachteil ist jedoch, dass die Konstruktion wenig explizit ist und Eigenschaften wie Linearität schwieriger zu zeigen sind.

Satz 4.14 (Monotone Konvergenz, Beppo-Levi). *Sei $(f_n) \subseteq \mathcal{M}^+$, und (f_n) sei punktweise wachsend. Dann ist $\lim_{n\to\infty} f_n \in \mathcal{M}^+$ und*

$$\lim_{n\to\infty}\int_\Omega f_n\,d\mu = \int_\Omega \big(\lim_{n\to\infty} f_n\big)\,d\mu\,. \tag{4.9}$$

Beweis. 1. $f = \lim_{n\to\infty} f_n$ ist messbar nach Satz 3.25, und da $f \geq f_n$ für alle n, folgt mit Monotonie $\int f\,d\mu \geq \int f_n\,d\mu$, also $\leq$ in (4.9).
2. Für jedes n sei

$$(\varphi_{n,k})_{k\geq 1} \subseteq \mathcal{T}^+, \quad \varphi_{n,k} \leq \varphi_{n,k+1}, \quad \varphi_{n,k} \nearrow f_n,\ k\to\infty.$$

Setze

$$\varphi_n := \max_{1\leq j\leq n}\varphi_{j,n}\ (\in \mathcal{T}^+).$$

Dann gelten (Beweis folgt)

$$\varphi_n \leq \varphi_{n+1}, \quad \varphi_n \leq f_n, \quad \varphi_n \nearrow f. \tag{4.10}$$

Dabei ist $\varphi_n \leq \varphi_{n+1}$ klar, da $\varphi_{j,n} \leq \varphi_{j,n+1}$ für alle j,n. Da $\varphi_{j,n} \leq f_j$, ist $\varphi_n = \max_{1\leq j\leq n}\varphi_{j,n} \leq \max_{1\leq j\leq n} f_j = f_n$, da die f_j wachsend sind.
Ist $f(x) < \infty$ und $\varepsilon > 0$, so existiert $N = N(x)$, so dass $f(x) - f_N(x) \leq \varepsilon/2$. Für $f_N(x)$ existiert K, so dass

$0 \leq f_N(x) - \varphi_{N,K}(x) \leq \varepsilon/2$. Dann ist $0 \leq f(x) - \varphi_n(x) \leq \varepsilon$, $n \geq \max(N,K)$, da $f(x) \geq \varphi_n(x) \geq \varphi_{N,K}(x)$ für diese n. Für $f(x) = \infty$ argumentiert man analog.

Mit (4.10) und der Integraldefinition (4.6) folgt

$$\int_\Omega f \, d\mu = \lim_{n\to\infty} \int_\Omega \varphi_n \, d\mu \leq \lim_{n\to\infty} \int_\Omega f_n \, d\mu .$$

Zusammen folgt (4.9). ∎

Wir erweitern das Maßintegral ein weiteres Mal auf integrierbare Funktionen als mögliche Integranden, welche nicht notwendigerweise nicht-negativ sind. Für jede messbare Funktion $f : \Omega \to \overline{\mathbb{R}}$ sind $f^+, f^- \in \mathcal{M}^+$ messbar.

Definition 4.15. *Eine Funktion $f \in \mathcal{M}$ heißt **quasiintegrierbar**, falls mindestens eines der Integrale $\int f^+ \, d\mu$ und $\int f^- \, d\mu$ reell, also $\in \mathbb{R}$, ist. Dann setze*

$$\int_\Omega f \, d\mu = \int_\Omega f^+ \, d\mu - \int_\Omega f^- \, d\mu . \tag{4.11}$$

*Es ist dann $\int_\Omega f \, d\mu \in \overline{\mathbb{R}}$. Die Funktion f heißt **integrierbar**, falls $\int f^+, \int f^- \in \mathbb{R}$, also auch $\int f \in \mathbb{R}$. Die Menge der reellwertigen, integrierbaren Funktionen wird mit $\mathcal{L}_1(\Omega, \mathcal{A}, \mu)$ bezeichnet.*

Satz 4.16. *Für $f \in \mathcal{M}$ sind äquivalent:*

1. f ist integrierbar.
2. f^+ und f^- sind integrierbar.
3. $|f|$ ist integrierbar.
4. $|f| \leq g$ für $g \in \mathcal{M}$, und g ist integrierbar.

Beweis. 1. $\Leftrightarrow$ 2.: klar nach Def.
2. $\Rightarrow$ 3.: $|f| = f^+ + f^-$, und das Integral ist additiv auf $\mathcal{M}^+$ (vgl. Satz 4.12).
3. $\Rightarrow$ 2.: $0 \leq f^+ \leq |f|$, $0 \leq f^- \leq |f|$, und das Integral ist monoton auf $\mathcal{M}^+$ (vgl. Satz 4.12)
3. $\Rightarrow$ 4.: klar mit $g = |f|$.
4. $\Rightarrow$ 3.: $\int |f| \, d\mu \leq \int g \, d\mu < \infty$ wegen Monotonie des Integrals auf $\mathcal{M}^+$. ∎

Beispiel 4.17. Sei $(\Omega, \mathcal{A}, \mu) = (\mathbb{N}, \mathcal{P}(\mathbb{N}), \nu)$ mit dem Zählmaß ν. Ist $f : \mathbb{N} \to \overline{\mathbb{R}}$, so ist f genau dann integrierbar, wenn $\int_\mathbb{N} |f| \, d\nu = \sum_{n=1}^\infty |f(n)| < \infty$, falls die zugehörige Reihe also absolut konvergiert. Wir erkennen an dieser Stelle, dass für $X \sim \mathrm{Geom}(p)$ gerade

$$\mathbb{E}[X] = \int_\mathbb{N} X \, d\mathbb{P}_X .$$

Eine analoge Darstellung gilt für die Poisson-Verteilung auf $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$.

Satz 4.18. *Ist $f \in \mathcal{M}$ integrierbar, so gilt $\mu\{|f| = \infty\} = 0$.*

Beweis. Für $\varphi = \infty \cdot 1_{\{|f|=\infty\}} \in \mathcal{M}^+$ gilt $\varphi \leq |f|$, somit $\infty \cdot \mu\big(\{|f| = \infty\}\big) = \int \varphi \, d\mu \leq \int |f| \, d\mu < \infty$, und $\mu\big(\{|f| = \infty\}\big) = 0$. ∎

Wir erhalten die bereits vertrauten einfachen Eigenschaften des Integrals nun auch für unser allgemeines Maßintegral.

Satz 4.19. *Seien $f,g \in \mathcal{L}_1$ integrierbar.*
1. ***Linearität.*** *Sind $a,b \in \mathbb{R}$, so ist $af+bg$ integrierbar und es gilt*

$$\int (af+bg)\,d\mu = a\int f\,d\mu + b\int g\,d\mu.$$

2. ***Monotonie.*** *Für $f \leq g$ gilt $\int f\,d\mu \leq \int g\,d\mu$.*
3. $|\int f\,d\mu| \leq \int |f|\,d\mu$.

Beweis. 1. Für $a \geq 0$ ist $(af)^+ = af^+$, $(af)^- = af^-$, für $a<0$ ist $(af)^+ = (-a)f^-$, $(af)^- = (-a)f^+$. $\int(af) = a\int f$ und somit die Integrierbarkeit von af folgen aus Satz 4.12 und der Definition (4.11).
Da $|f+g| \leq |f|+|g|$, folgt die Integrierbarkeit von $h=f+g$ aus Satz 4.16 und Satz 4.12. Dann ist

$$h = h^+ - h^- = f^+ - f^- + g^+ - g^-, \qquad \text{also} \qquad h^+ + f^- + g^- = g^+ + h^- + f^+.$$

Integration und Satz 4.12 liefern

$$\int h^+ + \int f^- + \int g^- = \int g^+ + \int h^- + \int f^+,$$

und da alle Integrale endlich sind, liefert Subtraktion von $\int h^-, \int f^-, \int g^-$ die Linearität.
2. Ist $f \leq g$, so ist $f^+ \leq g^+$, also $\int f^+ \leq \int g^+$ nach Satz 4.12, sowie $f^- \geq g^-$, also $\int f^- \geq \int g^-$, und somit folgt nach Definition (4.11) die Behauptung.
3. Mit Linearität und der gewöhnlichen Dreiecksungleichung für reelle Zahlen folgt:

$$\left|\int f\,d\mu\right| = \left|\int f^+\,d\mu - \int f^-\,d\mu\right| \leq \int f^+\,d\mu + \int f^-\,d\mu = \int |f|\,d\mu\,.$$

Man nennt 3. daher auch *Dreiecksungleichung für Integrale*. ■

Wir nutzen nun die allgemeine Definition des Maßintegrals einer messbaren numerischen Funktion auf einem Maßraum, um den Erwartungswert einer numerischen Zufallsvariable $X : \Omega \to \overline{\mathbb{R}}$ zu definieren.

Definition 4.20. *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und sei $X : \Omega \to \overline{\mathbb{R}}$ eine numerische Zufallsvariable. Ist X quasiintegrierbar bzgl. $\mathbb{P}$, d.h. $\min(\int X^+\,d\mathbb{P}, \int X^-\,d\mathbb{P}) < \infty$, so setze*

$$\mathbb{E}[X] := \int_\Omega X\,d\mathbb{P},$$

*$\mathbb{E}[X]$ ist der **Erwartungswert** von X. Soll das zugrunde liegende Maß betont werden, schreiben wir auch genauer $\mathbb{E}_{\mathbb{P}}[X]$. Ist X integrierbar, also $\mathbb{E}[|X|] < \infty$, so hat X einen **endlichen Erwartungswert**.*

Rekapitulation der Integralkonstruktion für Erwartungswerte:

Die Konstruktion des allgemeinen Maßintegrals verlief in drei Schritten. Analog überträgt sich dies auf Erwartungswerte.
1. Für nicht-negative einfache Zufallsvariablen der Form $X = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$, mit $\alpha_k \geq 0, A_1, \ldots, A_n, A_i \in \mathcal{A}$, eine disjunkte Zerlegung von Ω, ist

$$\mathbb{E}[X] = \int_\Omega X\,d\mathbb{P} = \sum_{k=1}^n \alpha_k\,\mathbb{P}(A_k)\,.$$

2. Für nicht-negative numerische Zufallsvariablen $X : \Omega \to [0,\infty]$ existiert eine monoton wachsende Folge (X_n) von nicht-negativen Treppenfunktionen (einfachen Zufallsvariablen) mit $X_n \nearrow X$ punktweise.

Setze dann

$$\mathbb{E}[X] = \lim_{n\to\infty} \mathbb{E}[X_n].$$

3. Für allgemeines X, setze $\mathbb{E}[X] = \mathbb{E}[X^+] - \mathbb{E}[X^-]$.

Heuristisch ist der Erwartungswert der mittlere Wert einer Zufallsvariablen, also das gemäß eines Wahrscheinlichkeitsmaßes gewichtete Mittel. Wir sehen im nächsten Kapitel warum wir so definierte Erwartungswerte üblicherweise über die Verteilung bestimmen werden und dies dürfen.

4.3 Transformationsformel und Varianz

Sei $(\Omega, \mathcal{A}, \mu)$ ein Maßraum und $(Y, \mathcal{C})$ ein Messaum. Ist $T : \Omega \to Y$ eine messbare Abbildung, so ist

$$(\mu \circ T^{-1})(C) = \mu\big(T^{-1}(C)\big), \qquad C \in \mathcal{C},$$

ein Maß auf $(Y, \mathcal{C})$, das Bildmaß aus Definition 3.11 von μ unter T. Für reellwertige Zufallsvariablen ist dieses gerade ihre Verteilung auf $(\mathbb{R}, \mathcal{B})$. Für die Integration bezüglich des Bildmaßes gilt folgende grundlegende *Transformationsformel*.

Satz 4.21 (Transformationsformel). *Sei $f : Y \to \overline{\mathbb{R}}$ messbar. Dann sind äquivalent:*
i) f ist integrierbar über $(Y, \mathcal{C}, \mu \circ T^{-1})$.
ii) $f \circ T$ ist integrierbar über $(\Omega, \mathcal{A}, \mu)$.
Gilt i) oder ii), so ist

$$\int_\Omega f \circ T \, d\mu = \int_Y f \, d(\mu \circ T^{-1}). \tag{4.12}$$

Beweis. Der Beweis folgt einer “maßtheoretischen Induktion”. Ist $f = \sum_j \beta_j 1_{B_j} \in \mathcal{T}^+(Y, \mathcal{C})$, mit B_j eine Zerlegung von Y und $\beta_j \geq 0$, so ist

$$f \circ T = \sum_j \beta_j 1_{B_j} \circ T = \sum_j \beta_j 1_{T^{-1}(B_j)},$$

und (4.12) ergibt sich direkt nach Definition von $\mu \circ T^{-1}$. Allgemeines $f \in \mathcal{M}^+$ wird monoton durch nicht-negative Treppenfunktion approximiert und es folgt (4.12) mit monotoner Konvergenz. Integrierbare Funktionen zerlegt man in Positiv- und Negativteil und beachtet, dass $(f \circ T)^+ = f^+ \circ T$. ∎

Sei nun also $X : \Omega \to \Omega'$ eine $(\Omega', \mathcal{A}')$-wertige Zufallsvariable und $h : \Omega' \to \mathbb{R}$ Borel-messbar. Dann ist h eine reellwertige Zufallsvariable auf $(\Omega', \mathcal{A}', \mathbb{P}_X)$, und $h(X)$ eine reellwertige Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$. Wir formulieren die Transformationsformel für diesen für die Stochastik wichtigen Spezialfall.

Korollar 4.22 (Transformationsformel für Erwartungswert). $h(X)$ ist genau dann quasiintegrierbar (über $(\Omega, \mathcal{A}, \mathbb{P})$), falls h quasiintegrierbar über $(\Omega', \mathcal{A}', \mathbb{P}_X)$ ist, und dann gilt

$$\mathbb{E}[h(X)] = \int_{\Omega'} h \, d\mathbb{P}_X. \tag{4.13}$$

Für den Erwartungswert einer reellwertigen Zufallsvariablen erhalten wir also insbesondere $\mathbb{E}[X] = \int_\mathbb{R} X \, d\mathbb{P}_X = \int_\mathbb{R} x \, d\mathbb{P}_X(x)$. Die Berechnung des Erwartungswertes über die Verteilung ist die erste wichtige Implikation der Transformationsformel. Weiter können wir $\mathbb{E}[h(X)]$, für Transformationen h, über die Verteilung von X berechnen und müssen nicht zunächst die Verteilung von $h(X)$ herleiten. Wichtige Spezialfälle sind a. für $h(x) = x^k$: *k-tes Moment*.

b. für $h(x) = (x - \mathbb{E}[X])^k$: *$k$-tes zentriertes Moment*, falls $\mathbb{E}[X]$ existiert.

Nach (4.13) hängen die Momente einer Zufallsvariable, also insbesondere der Erwartungswert, nur von der Verteilung der Zufallsvariable ab, wir sprechen daher auch von Momenten der Verteilung. Der Erwartungswert ist ein Maß für den mittleren Wert einer Verteilung, sagt aber nichts über die Streuung der Werte aus. Er kann einen Wert haben, der nicht im Wertebereich der Zufallsvariable liegt, wie etwa $\mathbb{E}[W] = 3{,}5$ zeigt für die Augenzahl W eines fairen Würfels.

Beispiel 4.23. X habe eine diskrete uniforme Verteilung auf $\{-n, \dots, n\}$. Dann ist $\forall n \in \mathbb{N} : \mathbb{E}[X] = 0$, aber die Streuung hängt offensichtlich von n ab.

Lemma 4.24. *Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathbb{R}$ eine Zufallsvariable. Gilt $\mathbb{E}[X^2] < \infty$, so*

1. *existiert $\mathbb{E}[X]$,*
2. *ist $\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] < \infty$.*

Beweis.

1. Es gilt
$$|X| \le \max(X^2, 1) \le X^2 + 1 \Rightarrow \mathbb{E}[|X|] \le \mathbb{E}[X^2 + 1] < \infty,$$
somit existiert $\mathbb{E}[X]$.
2. Es gilt
$$(X - \mathbb{E}[X])^2 = X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2 \le X^2 + 2|X| \cdot |\mathbb{E}[X]| + (\mathbb{E}[X])^2,$$
also ist
$$\mathbb{E}\left[(X - \mathbb{E}[X])^2\right] \le \mathbb{E}[X^2] + 2|\mathbb{E}[X]| \cdot \mathbb{E}[|X|] + (\mathbb{E}[X])^2 < \infty.$$ ∎

Definition 4.25. *Ist $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$, so heißen*

$$\operatorname{Var}(X) := \mathbb{E}\left[(X - \mathbb{E}X)^2\right]$$

*die **Varianz** von X und*

$$\sigma(X) := \sqrt{\operatorname{Var}(X)}$$

*die **Standardabweichung** von X.*

Satz 4.26 (Eigenschaften der Varianz)**.** *Ist $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $\mathbb{E}[X^2] < \infty$ und $a, b \in \mathbb{R}$, so ist*

1. $\operatorname{Var}(aX + b) = a^2 \cdot \operatorname{Var}(X)$ *und* $\sigma(aX + b) = |a| \cdot \sigma(X)$,
2. $\operatorname{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$,
3. $\operatorname{Var}(X) = \min\limits_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2]$ *und* $\mathbb{E}[X] = \arg\min\limits_{a \in \mathbb{R}} \mathbb{E}[(X - a)^2]$,
4. $\operatorname{Var}(X) \ge 0$ *und* $\operatorname{Var}(X) = 0 \Leftrightarrow \exists a \in \mathbb{R} : \mathbb{P}(X = a) = 1$.

Beweis.

1. Sei $Z = aX + b$, dann ist $\mathbb{E}[Z] = a\mathbb{E}[X] + b$, somit $Z - \mathbb{E}[Z] = a \cdot (X - \mathbb{E}[X])$ und
$$\operatorname{Var}(Z) = \mathbb{E}[(Z - \mathbb{E}Z)^2] = \mathbb{E}[a^2 \cdot (X - \mathbb{E}[X])^2] = a^2 \cdot \operatorname{Var}(X).$$
2. Es gilt
$$\operatorname{Var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - 2 \cdot \mathbb{E}\left[X\mathbb{E}[X]\right] + \mathbb{E}\left[(\mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2.$$

3. Nach Punkt 2. gilt
$$\mathbb{E}[(X-a)^2] = \mathrm{Var}(X) + \underbrace{(\mathbb{E}[X]-a)^2}_{\geq 0,\ =0 \text{ für } a=\mathbb{E}[X]}.$$
Damit ist $\min\limits_{a\in\mathbb{R}} \mathbb{E}[(X-a)^2] = \mathrm{Var}(X)$ sowie $\mathbb{E}[X] = \arg\min\limits_{a\in\mathbb{R}} \mathbb{E}[(X-a)^2]$.

4. Zu zeigen: Ist $Z:\Omega\to[0,\infty)$, so gilt $\mathbb{E}[Z]=0 \Leftrightarrow \mathbb{P}(Z=0)=1$.

   $\Leftarrow$ Sei $\mathbb{P}(Z=0)=1$. Dann muss für $\varepsilon>0$ stets $\mathbb{P}(Z\geq\varepsilon)=0$ gelten. Dann ist $\mathbb{E}[Z]=0$.

   $\Rightarrow$ Sei $\mathbb{E}[Z]=0$. Ist nun $\mathbb{P}(Z\geq\varepsilon)>0$ für ein $\varepsilon>0$, so wäre
$$\mathbb{E}[Z]\geq\varepsilon\cdot\mathbb{P}(Z\geq\varepsilon)>0.$$
   Also ist $\mathbb{P}(Z\geq\varepsilon)=0$ für alle $\varepsilon>0$, es muss also $\mathbb{P}(Z=0)=1$ sein.

   Wenden wir das auf $Z=(X-\mathbb{E}[X])^2$ an, so erhalten wir
$$\mathrm{Var}(X)=\mathbb{E}[Z]=0 \Leftrightarrow \mathbb{P}(X=\mathbb{E}[X])=\mathbb{P}(Z=0)=1.$$
   Wegen $Z\geq 0$ gilt auch $\mathrm{Var}(X)=\mathbb{E}[Z]\geq 0$. ■

Beispiel 4.27. 1. Ist $X\sim\mathrm{Poi}(\lambda)$, so ist $\mathrm{Var}(X)=\lambda$.

Beweis. Sei $X\sim\mathrm{Poi}(\lambda)$ und $h(x)=x(x-1)$. Dann gilt, dass
$$\mathbb{E}[h(X)]=\mathbb{E}[X(X-1)]=\sum_{k=0}^{\infty}k\cdot(k-1)\cdot\mathrm{e}^{-\lambda}\cdot\frac{\lambda^k}{k!}=\mathrm{e}^{-\lambda}\cdot\lambda^2\cdot\sum_{k=2}^{\infty}\frac{\lambda^{k-2}}{(k-2)!}=\mathrm{e}^{-\lambda}\cdot\lambda^2\cdot\mathrm{e}^{\lambda}=\lambda^2.$$
Die Varianz folgt aus
$$\mathrm{Var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=\mathbb{E}[X(X-1)]+\mathbb{E}[X]-(\mathbb{E}[X])^2=\lambda^2+\lambda-\lambda^2=\lambda.$$ ■

2. Sei $X\sim\mathrm{Bin}(n,p)$, so ist $\mathrm{Var}(X)=np(1-p)$.

Beweis. Es reicht aus, für eine spezielle Zufallsvariable mit dieser Verteilung die Varianz zu bestimmen.

Auf $\Omega=\{0,1\}^n$ seien
$$X_i:\Omega\to\Omega_i,\quad X_i(\omega)=\omega_i,\quad X_i\sim\mathrm{Ber}(p),\quad A_i=\{X_i=1\},\quad i=1,\dots,n.$$
Dann sind $A_1,\dots,A_n$ unabhängig und es gilt $\mathbb{P}(A_i)=p$. Wir definieren nun
$$X=\sum_{i=1}^n 1_{A_i}.$$
Dann gilt $X\sim\mathrm{Bin}(n,p)$. Daraus folgt nun
$$X^2=\Big(\sum_{i=1}^n 1_{A_i}\Big)^2=\sum_{j=1}^n\sum_{k=1}^n 1_{A_j}1_{A_k}=\sum_{j=1}^n 1_{A_j}+\sum_{j=1}^n\sum_{\substack{k=1\\ j\neq k}}^n 1_{A_j\cap A_k}.$$
Für $j\neq k$ gilt wegen der Unabhängigkeit $\mathbb{P}(A_j\cap A_k)=p^2$, also erhalten wir
$$\mathrm{Var}(X)=\mathbb{E}[X^2]-(\mathbb{E}[X])^2=n\cdot p+n(n-1)\cdot p^2-(n\cdot p)^2=np-np^2=np(1-p).$$ ■

4.4 Absolutstetigkeit von Maßen und der Satz von Radon-Nikodym

Wir führen zunächst Integration über Teilmengen $A \in \mathcal{A}$, $A \subset \Omega$ ein. Ist $A \in \mathcal{A}$, $A \neq \emptyset$, so ist die Spur-σ-Algebra gegeben durch

$$\mathcal{A}(A) = \{B \cap A,\ B \in \mathcal{A}\} = \{C \subseteq A,\ C \in \mathcal{A}\}.$$

Dann ist $(A, \mathcal{A}(A), \mu_A)$ ein Maßraum, die *Einschränkung von μ auf A*.

Ist nun $f \in \mathcal{L}_1$ integrierbar, so wollen wir das Integral von f über A definieren. Nach Satz 4.16 ist wegen $|f 1_A| \leq |f|$ auch $f 1_A$ integrierbar, und wir setzen

$$\int_A f\, d\mu := \int_\Omega f 1_A\, d\mu.$$

Andererseits hätten wir auch f auf A einschränken können und das Integral mit dem Maßraum $(A, \mathcal{A}(A), \mu_A)$ bilden können. Der folgende Satz zeigt, dass dies zum gleichen Ergebnis führt.

Satz 4.28. *Ist $f : \Omega \to \overline{\mathbb{R}}$ integrierbar und $A \in \mathcal{A}$, so ist die Einschränkung f_A von f auf A integrierbar über $(A, \mathcal{A}(A), \mu_A)$, und es gilt*

$$\int_A f_A\, d\mu_A = \int_\Omega f 1_A\, d\mu. \tag{4.14}$$

Beweis. a. Angenommen, $f = \sum_i \alpha_i 1_{A_i} \in \mathcal{T}^+$ ist eine Treppenfunktion. Dann ist

$$f_A = \sum_i \alpha_i 1_{A_i \cap A} = f 1_A$$

ebenfalls eine Treppenfunktion, und (4.14) ist klar.

b. Ist $f \in \mathcal{M}^+$ und $\phi_n \in \mathcal{T}^+$ mit $\phi_n \nearrow f$, so gilt auch $\phi_n 1_A = \phi_n|A \nearrow f|A = f 1_A$, und (4.14) überträgt sich von Treppenfunktionen auf den Limes. Schließlich zerlegt man $f \in \mathcal{L}_1$ in Positiv- und Negativteil und nutzt die Linearität des Integrals. ∎

Lemma 4.29. *Ist $f \in \mathcal{M}^+$ oder $f \in \mathcal{M}$ integrierbar und $A \in \mathcal{A}$ mit $\mu(A) = 0$, so folgt $\int_A f\, d\mu = 0$.*

Beweis. Sei $f \in \mathcal{M}^+$, dann ist $|f| 1_A \leq \infty 1_A$, und daher $\int_A f \leq \int_A \infty = \infty \cdot \mu(A) = 0$. Für $f \in \mathcal{M}$ integrierbar wenden wir den ersten Teil auf $|f|$ an. ∎

Wir sagen, eine Eigenschaft einer Familie von Funktionen $(f_i)_{i \in I}$ gilt *μ-fast überall*, schreibe kurz μ-f.ü., falls eine Menge $A \in \mathcal{A}$ mit $\mu(A) = 0$ existiert (Nullmenge), so dass diese Eigenschaft für alle $x \in A^\complement$ gilt.

Sind etwa $f, g \in \mathcal{M}$, so gilt $f \leq g$ μ-f.ü., falls $f(x) \leq g(x)$ für alle $x \in A^\complement$ für eine Menge $A \in \mathcal{A}$ mit $\mu(A) = 0$. Für $\mu = \mathbb{P}$ ein Wahrscheinlichkeitsmaß bezeichnen wir dies als fast sicher.

Satz 4.30. *Seien $f, g \in \mathcal{M}$ integrierbar. Dann sind äquivalent:*

1. Es gilt

$$\int_A f\, d\mu \overset{(=)}{\leq} \int_A g\, d\mu \quad \text{für alle } A \in \mathcal{A}, \tag{4.15}$$

2. $f \overset{(=)}{\leq} g$ μ-f.ü.

Beweis. Wir betrachten die Äquivalenz im Falle $\leq$. Sei $M := \{f > g\} \in \mathcal{A}$.
1. $\Rightarrow$ 2.: Setze $M_n := \{f \geq g + 1/n\}$. Mit (4.15) folgt

$$\int_{M_n} f\,d\mu \geq \int_{M_n} (g+1/n)\,d\mu = \int_{M_n} g\,d\mu + \mu(M_n)/n \geq \int_{M_n} f\,d\mu + \mu(M_n)/n,$$

also $\mu(M_n) = 0$ für alle n, und da $M = \bigcup_{n=1}^{\infty} M_n$, folgt auch $\mu(M) = 0$.
2. $\Rightarrow$ 1.: Für $A \in \mathcal{A}$ ist nach obigem Lemma und mit Monotonie

$$\int_A f\,d\mu = \int_{A\cap M} f\,d\mu + \int_{A\setminus M} f\,d\mu \leq \int_{A\setminus M} g\,d\mu \leq \int_A g\,d\mu\,.$$

■

Definition 4.31. *Auf einem Messraum $(\Omega, \mathcal{A})$ seien zwei Maße μ und ν gegeben. Dann ist ν **absolutstetig** bezüglich μ, $\nu \ll \mu$, falls jede μ-Nullmenge auch eine ν-Nullmenge ist, $A \in \mathcal{A}$ mit $\mu(A) = 0$ $\Rightarrow \nu(A) = 0$.*

Ein erste Intuition der Verwandtschaft des Begriffes zur „Stetigkeit" ergibt sich aus folgendem Satz.

Satz 4.32. *Auf $(\Omega, \mathcal{A})$ seien μ, ν Maße und ν endlich. Dann sind äquivalent:*

1. *$\nu \ll \mu$;*
2. *Für jedes $\varepsilon > 0\ \exists\, \delta > 0$, so dass $\forall A \in \mathcal{A}$ mit $\mu(A) < \delta \Rightarrow \nu(A) < \varepsilon$.*

Beweis. Wir beginnen mit "1. $\Rightarrow$ 2.". Angenommen die Implikation wäre falsch. Also existierten Mengen $A_n \in \mathcal{A}$ und $\varepsilon > 0$, so dass

$$\mu(A_n) < 2^{-n} \quad \text{und} \quad \nu(A_n) \geq \varepsilon\,.$$

Für $A = \limsup_{n\to\infty} A_n = \bigcap_{k=1}^{\infty} \bigcup_{n\geq k} A_n$ folgt:

$$\mu(A) \leq \mu\Big(\bigcup_{n\geq k} A_n\Big) \leq \sum_{n\geq k} \mu(A_n) < \sum_{n\geq k} 2^{-n}\,.$$

Die rechte Summe konvergiert für $k \to \infty$ gegen Null (wegen Konvergenz der gesamten Reihe), also $\mu(A) = 0$. Da ν endlich ist

$$\nu(A) = \nu\big(\limsup_{n\to\infty} A_n\big) \geq \limsup_{n\to\infty} \nu(A_n) \geq \varepsilon\,,$$

was zum Widerspruch führt.

Es bleibt "1. $\Leftarrow$ 2." zu zeigen. Aus $\mu(A) = 0$ folgt nach Voraussetzung $\nu(A) < \varepsilon$ für jedes beliebig kleine $\varepsilon > 0$ womit direkt $\nu(A) = 0$ und damit $\nu \ll \mu$ folgt. ■

Beispiel 4.33. 1. Sei f eine nicht-negative messbare numerische Funktion auf $(\Omega, \mathcal{A}, \mu)$. Dann wird durch $\nu(A) = \int_A f\,d\mu$ ein Maß definiert. Dieses erfüllt $\nu \ll \mu$.

2. Sei $(\Omega, \mathcal{A})$ ein Messraum und $\mathcal{A}$ enthalte alle einpunktigen Mengen. Dann sind für $x \neq y$ die Dirac-Maße δ_x und δ_y nicht absolutstetig zueinander.

Satz 4.34 (Radon-Nikodym). *Auf $(\Omega, \mathcal{A})$ seien μ, ν σ-endliche Maße. Dann ist $\nu \ll \mu$ genau dann* *wenn eine nicht-negative messbare numerische Funktion f auf $(\Omega, \mathcal{A})$ existiert mit*

$$\nu(A) = \int_A f \, d\mu \quad \forall A \in \mathcal{A} \, .$$

Die Funktion f ist μ-f.ü. eindeutig bestimmt. f bezeichnet man als ***(Radon-Nikodym) Dichte*** *oder auch $f = d\nu/d\mu$ als* ***Radon-Nikodym-Ableitung****.* ▪

Beweis. Wir beweisen den Satz für endliche Maße μ, ν. Für die Erweiterung zum allgemeinen Fall sei auf Elstrodt, Kapitel VII Satz 2.3., verwiesen.

(a) Betrachte die Menge

$$\mathcal{G} = \left\{ g : g \geq 0 \text{ messbar} \, , \int_A g \, d\mu \leq \nu(A) \; \forall A \in \mathcal{A} \right\} .$$

Da zumindest $g \equiv 0 \ \in \mathcal{G}$, ist $\mathcal{G} \neq \emptyset$. Des Weiteren ist für $g_1, g_2 \in \mathcal{G}$ auch deren Maximum $(g_1 \vee g_2) \in \mathcal{G}$, da

$$\begin{aligned}
\int_A (g_1 \vee g_2) \, d\mu &= \int_{A \cap \{g_1 \leq g_2\}} (g_1 \vee g_2) \, d\mu + \int_{A \cap \{g_1 > g_2\}} (g_1 \vee g_2) \, d\mu \\
&= \int_{A \cap \{g_1 \leq g_2\}} g_2 \, d\mu + \int_{A \cap \{g_1 > g_2\}} g_1 \, d\mu \\
&\leq \nu(A \cap \{g_1 \leq g_2\}) + \nu(A \cap \{g_1 > g_2\}) = \nu(A) \, .
\end{aligned}$$

(b) Es existiert eine Funktion $f \in \mathcal{G}$, für welche gilt:

$$\int f \, d\mu = \sup_{g \in \mathcal{G}} \{ \textstyle\int g \, d\mu \} \, .$$

Um dies zu zeigen, wähle man eine Folge (g_n) mit $g_n \in \mathcal{G}$ und $\int g_n \, d\mu \to \sup_{g \in \mathcal{G}} \{ \int g \, d\mu \}$. Nach **(a)** ist $\tilde{g}_n = \max(g_1, \dots, g_n) \in \mathcal{G}$ und für die Folge $(\tilde{g}_n)$ gilt $\tilde{g}_n \nearrow f$ mit einer messbaren nicht-negativen numerischen Funktion f. Mit Satz 4.14 von der monotonen Konvergenz folgern wir:

$$\int f \, d\mu = \lim_{n \to \infty} \int \tilde{g}_n \, d\mu \geq \lim_{n \to \infty} \int g_n \, d\mu = \sup_{g \in \mathcal{G}} \{ \textstyle\int g \, d\mu \} \, .$$

Für beliebiges $A \in \mathcal{A}$ folgt andererseits mit $\mathbf{1}_A \tilde{g}_n \nearrow \mathbf{1}_A f$, dass

$$\int f \, d\mu = \lim_{n \to \infty} \int_A \tilde{g}_n \, d\mu \leq \nu(A) \, .$$

Damit ist $f \in \mathcal{G}$ und somit $\int f \, d\mu = \sup_{g \in \mathcal{G}} \{ \int g \, d\mu \}$.

(c) Als nächstes zeigen wir, dass $\nu(A) = \int_A f \, d\mu$ für alle $A \in \mathcal{A}$.
Setze $\tilde{\nu}(A) = \nu(A) - \int_A f \, d\mu \geq 0$. Falls $\tilde{\nu}(\Omega) > 0$, existiert ein $\varepsilon > 0$ so dass $\tilde{\nu}(\Omega) > \varepsilon \mu(\Omega)$. Bezeichne nun $\Omega_+ \subseteq \Omega$ eine Teilmenge auf der $\tilde{\nu} - \varepsilon \mu$ ein Maß, also nicht-negativ, ist. [1]

$$\tilde{\nu}(A) \geq \tilde{\nu}(A \cap \Omega_+) \geq \varepsilon \, \mu(A \cap \Omega_+) \, ,$$

und hiermit folgt

$$\nu(A) = \int_A f \, d\mu + \tilde{\nu}(A) \geq \int_A f \, d\mu + \varepsilon \int_A \mathbf{1}_{\Omega_+} \, d\mu \, ,$$

[1] $\mu : \mathcal{A} \to \overline{\mathbb{R}}$ mit $\mu(\emptyset) = 0$ und σ-additiv heißt *signiertes Maß*. Nach dem *Hahnschen Zerlegungssatz* existiert stets eine disjunkte Zerlegung $\Omega_+, \Omega_- \in \mathcal{A}$ von Ω, so dass $\mu(A) \leq 0$ für alle $A \subseteq \Omega_-$, und $\mu(A) \geq 0$ für alle $A \subseteq \Omega_+$. Siehe Elstrodt (1996) §1, Kapitel VII.

woraus nun $f+\varepsilon\mathbf{1}_{\Omega_+}\in\mathcal{G}$ folgt. Dann ist aber $\mu(\Omega_+)=0$, wegen Absolutstetigkeit $\nu(\Omega_+)=0$ und somit $\tilde\nu(\Omega_+)=0$. Aus

$$(\tilde\nu-\varepsilon\mu)(\Omega)=(\tilde\nu-\varepsilon\mu)(\Omega_-)\le 0$$

ergibt sich aber ein Widerspruch. Also, $\tilde\nu(\Omega)=0$.

(d) Als letztes bleibt die Eindeutigkeit zu zeigen. Hinreichend ist die Aussage wenn f und g nichtnegative messbare numerische Funktionen seien mit $\int_A f\,d\mu=\int_A g\,d\mu<\infty$ für alle $A\in\mathcal{A}$, dass dann $f=g$ μ-f.ü. folgt. Dies wurde in Satz 4.30 gezeigt.

■

Eine „Rechenregel mit Dichten“ ist in nachfolgender Proposition festgehalten.

Proposition 4.35. *Gegeben zwei σ-endliche Maße μ,ν auf $(\Omega,\mathcal{A})$ mit $\nu\ll\mu$. Dann gilt für alle messbaren numerischen Funktionen f auf $(\Omega,\mathcal{A})$:*

$$\int f\,d\nu=\int f\frac{d\nu}{d\mu}\,d\mu\,.$$

Im Falle der Existenz existieren immer beide Integrale und stimmen überein.

4.5 Erwartungswert und Varianz absolutstetiger Verteilungen

Für die in Definition 3.41 eingeführten, absolutstetigen Verteilungen gilt gerade, dass sie absolutstetig bezüglich dem Lebesguemaß über $\mathbb{R}$ sind und die Dichten sind formal Radon-Nikodym-Dichten bezüglich diesem Lebesguemaß. Daher können wir Maßintegrale bezüglich solchen Verteilungen mit den Dichten berechnen. Ist $X\sim f$ eine absolutstetig verteilte Zufallsvariable, dann gilt

1. Ist $X\ge 0$ (also $f(x)=0$ für $x<0$), so ist

$$\mathbb{E}[X]=\int_0^\infty xf(x)\,\mathrm{d}x\in[0,\infty]$$

der Erwartungswert von X. Es sei hierzu bemerkt, dass für Integrale $\int g\,d\lambda$, wenn λ das Lebesguemaß bezeichnet, meist $\int g(x)\,\mathrm{d}x$ geschrieben wird und dies für Riemann-integrierbare Funktionen g mit dem Riemann-Integral übereinstimmt.

2. Ist im allgemeinen Fall $\int_{\mathbb{R}}|x|f(x)\,\mathrm{d}x<\infty$, so ist also

$$\mathbb{E}[X]=\int_{\mathbb{R}}xf(x)\,\mathrm{d}x$$

der Erwartungswert von X. Es gilt hier also wie in Proposition 4.35:

$$\mathbb{E}[X]=\int_{\mathbb{R}}x\,d\mathbb{P}_X(x)=\int_{\mathbb{R}}x\frac{d\mathbb{P}_X}{d\lambda}(x)\,d\lambda(x)=\int_{\mathbb{R}}xf(x)\,\mathrm{d}x\,.$$

Bemerkung. Wir geben $\mathbb{E}[X]$ hier direkt über die Verteilung von X an. Für Zufallsvariablen mit diskreter Verteilung galt

$$\int_\Omega X(\omega)\,d\mathbb{P}(\omega)=\sum_{x\in X(\Omega)}x\mathbb{P}_X(\{x\}),$$

für Zufallsvariablen mit absolutstetiger Verteilung

$$\int_\Omega X(\omega)\,d\mathbb{P}(\omega)=\int_{\mathbb{R}}xf(x)\,\mathrm{d}x.$$

Beispiel 4.36. Ist $X \sim f_X$ mit $\mathbb{E}[X] = a$, so hat $Y = \sigma X + \mu$ (Element der Lokations-Skalenfamilie) den Erwartungswert $\mathbb{E}[Y] = \sigma \cdot a + \mu$. Dies können wir auch durch eine direkte Rechnung mit der Dichte von Y verifizieren:

$$\mathbb{E}[Y] = \int_{\mathbb{R}} y \frac{1}{\sigma} f_X\big((y-\mu)/\sigma\big)\,\mathrm{d}y = \int_{\mathbb{R}} (\sigma x + \mu)\, f_X(x)\,\mathrm{d}x = \sigma \mathbb{E}[X] + \mu\,.$$

1. Für $X \sim \mathcal{U}[-1,1]$ ist $\mathbb{E}[X] = 0$, also ist für $X \sim \mathcal{U}[a,b]$ der Erwartungswert $\mathbb{E}[X] = \frac{a+b}{2}$.
2. Für $X \sim \operatorname{Exp}(1)$ ist

$$\mathbb{E}[X] = \int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x = \underbrace{[-x\mathrm{e}^{-x}]_0^\infty}_{=0} + \int_0^\infty \mathrm{e}^{-x}\,\mathrm{d}x = [-\mathrm{e}^{-x}]_0^\infty = 1,$$

   also ist für $X \sim \operatorname{Exp}(\lambda)$ der Erwartungswert $\mathbb{E}[X] = \frac{1}{\lambda}$.
3. Für $X \sim \mathcal{N}(0,1)$ ist

$$\begin{aligned}\mathbb{E}[X] &= \frac{1}{\sqrt{2\pi}}\Big(\int_0^\infty x\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x + \int_{-\infty}^0 x\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\Big)\\ &= \frac{1}{\sqrt{2\pi}}\Big(\int_0^\infty x\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x + \int_0^\infty -x\mathrm{e}^{-\frac{x^2}{2}}\,\mathrm{d}x\Big) = \frac{1}{\sqrt{2\pi}}\int_0^\infty 0\,\mathrm{d}x = 0,\end{aligned}$$

   also ist für $X \sim \mathcal{N}(\mu,\sigma^2)$ der Erwartungswert $\mathbb{E}[X] = \mu$.

Die Transformationsformel für Erwartungswerte lautet im Spezialfall absolutstetiger Verteilungen: Ist $X : \Omega \to \mathbb{R}$ eine Zufallsvariable mit $X \sim f$ und $u : \mathbb{R} \to \mathbb{R}$ Borel-messbar, und ist

$$\int_{\mathbb{R}} |u(x)| f(x)\,\mathrm{d}x < \infty,$$

so gilt

$$\mathbb{E}[u(X)] = \int_{\mathbb{R}} u(x) f(x)\,\mathrm{d}x.$$

Bemerkung. Für X^2 folgt

$$\mathbb{E}[X^2] = \int_{\mathbb{R}} x^2 f(x)\,\mathrm{d}x\,.$$

Gilt $\mathbb{E}[X^2] < \infty$, so ist auch $\mathbb{E}[|X|] < \infty$.

Beispiel 4.37.

1. Ist $X \sim \mathcal{U}[-1,1]$, dann ist $\mathbb{E}[X] = 0$ und

$$\operatorname{Var}(X) = \mathbb{E}[X^2] = 2 \cdot \frac{1}{2}\int_0^1 x^2\,\mathrm{d}x = \Big[\frac{x^3}{3}\Big]_0^1 = \frac{1}{3}.$$

   Damit ist für $X \sim \mathcal{U}[a,b]$ die Varianz

$$\operatorname{Var}(X) = \Big(\frac{b-a}{2}\Big)^2 \cdot \frac{1}{3} = \frac{(b-a)^2}{12}.$$

2. Ist $X \sim \operatorname{Exp}(1)$, dann ist $\mathbb{E}[X] = 1$,

$$\mathbb{E}[X^2] = \int_0^\infty x^2\mathrm{e}^{-x}\,\mathrm{d}x = [x^2 \cdot (-\mathrm{e}^{-x})]_0^\infty + 2 \cdot \int_0^\infty x\mathrm{e}^{-x}\,\mathrm{d}x = 2$$

   und $\operatorname{Var}(X) = 2 - 1^2 = 1$. Damit ist für $X \sim \operatorname{Exp}(\lambda)$ die Varianz $\dfrac{1}{\lambda^2}$.

3. Ist $X \sim \mathcal{N}(0,1)$, dann ist $\mathbb{E}[X] = 0$ und
$$\begin{aligned}\operatorname{Var}(X) = \mathbb{E}[X^2] &= \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{x^2}{2}}\, \mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x \cdot \left(x\mathrm{e}^{-\frac{x^2}{2}}\right) \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \left[x \cdot \left(-\mathrm{e}^{-\frac{x^2}{2}}\right)\right]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \mathrm{e}^{-\frac{x^2}{2}}\, \mathrm{d}x = 0 + 1 = 1.\end{aligned}$$
Dann ist für $X \sim \mathcal{N}(\mu, \sigma^2)$ die Varianz σ^2. Daher ist es möglich (und üblich), von der Normalverteilung mit Erwartungswert μ und Varianz σ^2 zu sprechen.

4.6 Fast sichere Konvergenz und majorisierte Konvergenz

Definition 4.38. *Eine Folge von Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ konvergiert $\mathbb{P}$-**fast sicher**, kurz $\mathbb{P}$-f.s., gegen eine Zufallsvariable X, falls eine $\mathbb{P}$-Nullmenge $A \in \mathcal{A}$ existiert, so dass*
$$X_n(\omega) \to X(\omega) \quad \text{für alle } \omega \in A^{\complement}.$$

Wir haben also punktweise Konvergenz auf dem Komplement einer $\mathbb{P}$-Nullmenge. Für allgemeine Maße μ, und messbaren numerischen Funktionen f_n, wird dies analog als *Konvergenz μ-fast überall* bezeichnet.

Satz 4.39 (Lemma von Fatou). *Ist $(f_n) \subseteq \mathcal{M}^+$ eine Folge nicht-negativer, messbarer numerischer Funktionen, so gilt*
$$\int \liminf_{n\to\infty} f_n\, d\mu \le \liminf_{n\to\infty} \int f_n\, d\mu.$$

Beweis. Sei $f = \liminf_{n\to\infty} f_n$, dann ist $f \in \mathcal{M}^+$ nach Satz 3.25. Für $n \geq 1$ sei
$$g_n = \inf_{k\geq n} f_k \in \mathcal{M}^+.$$
Dann gilt $g_n \nearrow f$ und somit nach dem Satz 4.14 von der monotonen Konvergenz
$$\lim_{n\to\infty} \int g_n\, d\mu = \int f\, d\mu.$$
Weiter ist $g_n \le f_k$ für alle $k \geq n$, und daher auch $\int g_n\, d\mu \le \inf_{k\geq n} \int f_k\, d\mu$. Somit folgt
$$\int f\, d\mu = \lim_{n\to\infty} \int g_n\, d\mu \le \lim_{n\to\infty} \inf_{k\geq n} \int f_k\, d\mu = \liminf_{n\to\infty} \int f_k\, d\mu.$$
∎

Korollar 4.40. Ist $(f_n) \subseteq \mathcal{M}^+$ eine konvergente Folge nicht-negativer, messbarer numerischer Funktionen mit $\lim_{n\to\infty} f_n = f$, so gilt
$$\int f\, d\mu \le \liminf_{n\to\infty} \int f_n\, d\mu. \tag{4.16}$$

Es folgt nach Satz 4.14 das zweite Hauptresultat über Vertauschung von Limes-Bildung und Integration.

Satz 4.41 (Majorisierte Konvergenz, Satz von Lebesgue). *Seien $f_n, f \in \mathcal{M}$ mit $f_n \to f$ μ-f.ü.*

Existiert eine integrierbare Funktion $g \in \mathcal{M}^+$ mit

$$|f_n| \leq g,$$

dann sind f_n, f integrierbar und

$$\int f_n \, d\mu \to \int f \, d\mu, \qquad \int |f - f_n| \, d\mu \to 0.$$

Beweis. Die f_n sind nach Satz 4.16 integrierbar, und da $|f| \leq g$ μ-f.ü., auch f nach Lemma 4.29. Sei

$$M = \{f_n \not\to f\} \cup \{f_n = \infty \text{ für mind. ein } n\} \cup \{f = \infty\},$$

also $\mu(M) = 0$ nach Satz 4.18. Für $\tilde{f}_n = f_n 1_{M^\complement}$, $\tilde{f} = f 1_{M^\complement}$ gelten $\tilde{f}_n \to \tilde{f}$ punktweise überall, $\int \tilde{f}_n = \int f_n$, $\int \tilde{f} = \int f$ und $\int |\tilde{f} - \tilde{f}_n| = \int |f - f_n|$, sowie $|\tilde{f}_n| \leq g$.

Es kann daher angenommen werden, dass f_n, f reellwertig sind und $f_n \to f$ punktweise überall gilt. Sei $g_n = 2g - |f_n - f| \in \mathcal{M}^+$, dann ist $g_n \to 2g$, und mit (4.16) aus dem Korollar 4.40 zum Lemma von Fatou folgt

$$\begin{aligned} 2 \int g \, d\mu &\leq \liminf_{n\to\infty} \int g_n \, d\mu \\ &= 2 \int g - \limsup_{n\to\infty} \int |f_n - f| \, d\mu \,. \end{aligned}$$

Also folgt $\limsup_{n\to\infty} \int |f_n - f| \, d\mu = 0$ und somit Konvergenz. Da

$$\left| \int f \, d\mu - \int f_n \, d\mu \right| \leq \int |f - f_n| \, d\mu,$$

folgt der zweite Teil der Behauptung. ∎

Korollar 4.42 (beschränkte Konvergenz). Seien $f_n \in \mathcal{M}$ mit $f_n \to f$ μ-f.ü. Ist μ endlich und für ein $C > 0$ gelte $|f_n| \leq C$ für alle $n \geq 1$, dann sind $f_n, f \in \mathcal{L}_1$ und

$$\int f_n \, d\mu \to \int f \, d\mu, \qquad \int |f - f_n| \, d\mu \to 0.$$

5 Produktraum und unabhängige Zufallsvariablen

5.1 Kovarianz und Korrelation

Sind $X,Y:\Omega\to\mathbb{R}$ Zufallsvariablen auf demselben Wahrscheinlichkeitsraum mit $\mathbb{E}[X^2]<\infty$ und $\mathbb{E}[Y^2]<\infty$, dann gilt für die Zufallsvariable $Z=X+Y$ auch $\mathbb{E}[Z^2]<\infty$, da $Z^2\leq 2(X^2+Y^2)$. Also existiert nach Lemma 4.24 $\operatorname{Var}(X+Y)$ und

$$\begin{aligned}\operatorname{Var}(X+Y)&=\mathbb{E}[(X+Y-\mathbb{E}[X+Y])^2]=\mathbb{E}[((X-\mathbb{E}[X])+(Y-\mathbb{E}[Y]))^2]\\&=\mathbb{E}[(X-\mathbb{E}[X])^2]+\mathbb{E}[(Y-\mathbb{E}[Y])^2]+2\cdot\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])].\end{aligned}$$

Definition 5.1. *Sind $X,Y:\Omega\to\mathbb{R}$ Zufallsvariablen mit $\mathbb{E}[X^2]<\infty$ und $\mathbb{E}[Y^2]<\infty$, dann heißt*

$$\operatorname{Cov}(X,Y):=\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]$$

*die **Kovarianz** von X und Y.*

Satz 5.2. *Sind $X,Y,X_1,\dots,X_m,Y_1,\dots,Y_n:\Omega\to\mathbb{R}$ Zufallsvariablen auf demselben Wahrscheinlichkeitsraum mit endlichen zweiten Momenten und $a,b,a_1,\dots,a_m,b_1,\dots,b_n\in\mathbb{R}$, so gilt:*

1. $\operatorname{Cov}(X,Y)=\mathbb{E}[X\cdot Y]-\mathbb{E}[X]\cdot\mathbb{E}[Y]$,
2. $\operatorname{Cov}(X,Y)=\operatorname{Cov}(Y,X)$ *und* $\operatorname{Cov}(X,X)=\operatorname{Var}(X)$,
3. $\operatorname{Cov}(aX+b,Y)=a\operatorname{Cov}(X,Y)$,
4. $\operatorname{Cov}\left(\sum_{i=1}^m a_iX_i,\sum_{j=1}^n b_jY_j\right)=\sum_{i=1}^m\sum_{j=1}^n a_ib_j\operatorname{Cov}(X_i,Y_j)$,
5. $\operatorname{Var}(X_1+\dots+X_m)=\sum_{i=1}^m\operatorname{Var}(X_i)+2\cdot\sum_{1\leq i<j\leq m}\operatorname{Cov}(X_i,X_j)$.

Beweis. 1. Lässt sich zeigen mit

$$\begin{aligned}\operatorname{Cov}(X,Y)&=\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]=\mathbb{E}[XY-X\mathbb{E}[Y]-Y\mathbb{E}[X]+\mathbb{E}[X]\mathbb{E}[Y]]\\&=\mathbb{E}[X\cdot Y]-2\mathbb{E}[X]\cdot\mathbb{E}[Y]+\mathbb{E}[X]\cdot\mathbb{E}[Y]=\mathbb{E}[X\cdot Y]-\mathbb{E}[X]\cdot\mathbb{E}[Y]\,.\end{aligned}$$

2. Lässt sich zeigen mit

$$\begin{aligned}\operatorname{Cov}(X,Y)&=\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]=\mathbb{E}[(Y-\mathbb{E}[Y])(X-\mathbb{E}[X])]=\operatorname{Cov}(Y,X),\\\operatorname{Cov}(X,X)&=\mathbb{E}[(X-\mathbb{E}[X])(X-\mathbb{E}[X])]=\mathbb{E}[(X-\mathbb{E}[X])^2]=\operatorname{Var}(X)\,.\end{aligned}$$

3. Lässt sich zeigen mit

$$\operatorname{Cov}(aX+b,Y)=\mathbb{E}[a(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]=a\mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])]=a\operatorname{Cov}(X,Y).$$

4. Lässt sich zeigen mit

$$\begin{aligned}\operatorname{Cov}\Big(\sum_{i=1}^m a_iX_i, \sum_{j=1}^n b_jY_j\Big) &= \mathbb{E}\Big[\sum_{i=1}^m a_i(X_i-\mathbb{E}[X_i])\cdot\sum_{j=1}^n b_j(Y_j-\mathbb{E}[Y_j])\Big]\\ &= \sum_{i=1}^m\sum_{j=1}^n a_ib_j\mathbb{E}\big[(X_i-\mathbb{E}[X_i])(Y_j-\mathbb{E}[Y_j])\big] = \sum_{i=1}^m\sum_{j=1}^n a_ib_j\operatorname{Cov}(X_i,Y_j).\end{aligned}$$

5. Lässt sich zeigen mit

$$\begin{aligned}\operatorname{Var}(X_1+\cdots+X_m) &= \operatorname{Cov}(X_1+\cdots+X_m, X_1+\cdots+X_m)\\ &= \sum_{i=1}^m\sum_{j=1}^m \operatorname{Cov}(X_i,X_j) = \sum_{i=1}^m \operatorname{Cov}(X_i,X_i) + \sum_{i=1}^m\sum_{\substack{j=1\\ i\neq j}}^m \operatorname{Cov}(X_i,X_j)\\ &= \sum_{i=1}^m \operatorname{Var}(X_i) + \sum_{1\le i<j\le m}(\operatorname{Cov}(X_i,X_j)+\operatorname{Cov}(X_j,X_i))\\ &= \sum_{i=1}^m \operatorname{Var}(X_i) + 2\cdot\sum_{1\le i<j\le m}\operatorname{Cov}(X_i,X_j).\end{aligned}$$

■

Definition 5.3. *Sind* $\mathbb{E}[X^2]<\infty$ *und* $\mathbb{E}[Y^2]<\infty$, *sowie* $\operatorname{Var}(X)>0$ *und* $\operatorname{Var}(Y)>0$, *so heißt*

$$\rho(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}(X)\cdot\operatorname{Var}(Y)}} = \frac{\operatorname{Cov}(X,Y)}{\sigma(X)\cdot\sigma(Y)}$$

der ***Korrelationskoeffizient****, kurz* ***Korrelation****, zwischen* X *und* Y. *Gilt* $\rho(X,Y)=\operatorname{Cov}(X,Y)=0$, *so nennen wir* X *und* Y ***unkorreliert****.*

Wir fragen uns, was ist die *optimale lineare Prädiktion* einer Zufallsvariable Y, wenn uns die Information über eine andere Zufallsvariable X vorliege. Hier kommt $\rho(X,Y)$ eine entscheidende Bedeutung für die Vorhersage zu. Eine lineare Vorhersage $\hat{Y}$ von Y aus X ist eine Transformation $\hat{Y}=aX+b$, mit $a,b\in\mathbb{R}$. Die Skalare $a,b\in\mathbb{R}$ sollen dabei so gewählt werden, dass der quadratische Vorhersagefehler, $\mathbb{E}\big[(Y-(aX+b))^2\big]$, minimal wird.

Satz 5.4. *Für* $\operatorname{Var}(X)>0$, $\operatorname{Var}(Y)>0$ *hat das Optimierungsproblem*

$$\textit{minimiere } \mathbb{E}\big[(Y-(aX+b))^2\big] \ \textit{ über } a,b\in\mathbb{R} \tag{5.1}$$

die Lösung

$$a^* = \frac{\sigma(Y)}{\sigma(X)}\cdot\rho(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(X)}, \quad b^* = \mathbb{E}[Y]-a^*\mathbb{E}[X]$$

und der minimale Wert von (5.1) *ist*

$$M^* = \operatorname{Var}(Y)\cdot(1-\rho^2(X,Y)). \tag{5.2}$$

■

Beweis. Für festes $a\in\mathbb{R}$ wird (5.1) minimal für $b=\mathbb{E}[Y]-a\mathbb{E}[X]$. Mit $\bar{X}=X-\mathbb{E}[X]$, $\bar{Y}=Y-\mathbb{E}[Y]$ bleibt

$$h(a) = \mathbb{E}[(\bar{Y}-a\bar{X})^2] = \mathbb{E}[\bar{Y}^2]-2a\mathbb{E}[\bar{X}\cdot\bar{Y}]+a^2\mathbb{E}[\bar{X}^2] = \operatorname{Var}(Y)-2a\operatorname{Cov}(X,Y)+a^2\operatorname{Var}(X)$$

zu minimieren. Das ist eine quadratische Funktion mit Minimum an der Stelle a mit

$$h'(a) = -2\,\mathrm{Cov}(X,Y) + 2a\,\mathrm{Var}(X) \stackrel{!}{=} 0 \Rightarrow a = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(X)}.$$

Einsetzen ergibt

$$M^* = \mathrm{Var}(Y) - 2\frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(X)} + \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(X)} = \mathrm{Var}(Y)\cdot(1-\rho^2(X,Y)).$$ ■

Satz 5.5. *Sind* $\mathbb{E}[X^2] < \infty$ *und* $\mathbb{E}[Y^2] < \infty$*, sowie* $\mathrm{Var}(X) > 0$ *und* $\mathrm{Var}(Y) > 0$*, so gilt:*

1. $-1 \le \rho(X,Y) \le 1$ *und*

$$\big|\rho(X,Y)\big| = 1 \Leftrightarrow \exists a,b \in \mathbb{R} \text{ mit } \mathbb{P}(Y = aX + b) = 1,$$

2. *die* CAUCHY-SCHWARZ-*Ungleichung*

$$\big(\mathrm{Cov}(X,Y)\big)^2 \le \mathrm{Var}(X)\cdot\mathrm{Var}(Y)\,. \tag{5.3}$$

Beweis. 1. Für M^* in (5.2) ist $M^* \ge 0$, also $\rho^2(X,Y) \le 1$. Weiterhin gilt

$$\begin{aligned}|\rho(X,Y)| = 1 \Leftrightarrow M^* = 0 &\Leftrightarrow \exists a,b \in \mathbb{R} : \mathbb{E}[(Y-(aX+b))^2] = 0\\ &\Leftrightarrow \exists a,b \in \mathbb{R} : \mathbb{P}(Y = aX+b) = 1.\end{aligned}$$

2. Es gilt

$$0 \le M^* = \frac{1}{\mathrm{Var}(X)}\Big(\mathrm{Var}(X)\,\mathrm{Var}(Y) - \big(\mathrm{Cov}(X,Y)\big)^2\Big),$$

also (5.3). ■

Bemerkung. Es gilt folgende Äquivalenz:

1. Für alle Zufallsvariablen X, Y mit $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2] < \infty$ gilt

$$\big(\mathrm{Cov}(X,Y)\big)^2 \le \mathrm{Var}(X)\cdot\mathrm{Var}(Y).$$

2. Für alle Zufallsvariablen X, Y mit $\mathbb{E}[X^2]$, $\mathbb{E}[Y^2] < \infty$ gilt

$$(\mathbb{E}[XY])^2 \le \mathbb{E}\left[X^2\right]\cdot\mathbb{E}\left[Y^2\right].$$

Diese Äquivalenz ist nicht offensichtlich, aber elementar beweisbar. Eigentlich stellt vor allem letztere Ungleichung einen Spezialfall der allgemeinen CAUCHY-SCHWARZ-Ungleichung für Erwartungswerte von quadratintegrierbaren, reellwertigen Zufallsvariablen dar. Die Ungleichung (5.3) ist auch als *Varianz-Kovarianz-Ungleichung* bekannt.

5.2 Unabhängigkeit von Mengensystemen und Zufallsvariablen

Endlich viele Ereignisse $A_1,\dots,A_n$ heißen unabhängig, falls für alle $I \subseteq \{1,\dots,n\}$ gilt

$$\mathbb{P}\Big(\bigcap_{i\in I} A_i\Big) = \prod_{i\in I}\mathbb{P}(A_i).$$

Eine beliebe Familie von Ereignissen $(A_i)_{i\in I} \subset \mathcal{A}$ heißt unabhängig, falls jede endliche Teilfamilie unabhängig ist.

Mengensysteme $\mathcal{E}_1,\dots,\mathcal{E}_n \subseteq \mathcal{A}$ heißen unabhängig, falls für alle $A_1 \in \mathcal{E}_1,\dots,A_n \in \mathcal{E}_n$ die Ereignisse $A_1,\dots,A_n$ unabhängig sind. Schließlich heißt eine beliebige Familie von Mengensystemen $(\mathcal{E}_i)_{i\in I}$, $\mathcal{E}_i \subseteq \mathcal{A}$ unabhängig, falls wiederum jede endliche Teilfamilie unabhängig ist. Unabhängigkeit für beliebige Familien ist also die Anforderung an jede endliche Teilfamilie.

Bemerkung 5.6. a. Sind Mengensysteme $\mathcal{E}_1,\dots,\mathcal{E}_n \subseteq \mathcal{A}$ gegeben und gilt $\Omega \in \mathcal{E}_i$ für jedes i, so sind die Mengensysteme bereits unabhängig, falls für alle $A_1 \in \mathcal{E}_1,\dots,A_n \in \mathcal{E}_n$ gilt

$$\mathbb{P}\Big(\bigcap_{i=1}^n A_i\Big) = \prod_{i=1}^n\mathbb{P}(A_i).$$

b. Sind umgekehrt unabhängige Mengensysteme $\mathcal{E}_1,\dots,\mathcal{E}_n \subseteq \mathcal{A}$ gegeben, so sind auch die Mengensysteme $\tilde{\mathcal{E}}_i = \mathcal{E}_i \cup \{\Omega\}$, $i = 1,\dots,n$, unabhängig.

Satz 5.7. *Sind $\mathcal{E}_1,\dots,\mathcal{E}_n \subseteq \mathcal{A}$ unabhängige Mengensysteme und ist jedes $\mathcal{E}_i$ $\cap$-stabil, so sind auch $\sigma(\mathcal{E}_1),\dots,\sigma(\mathcal{E}_n)$ unabhängig.* ▪

Beweis. Der Beweis verwendet ein Dynkin-System-Argument. Wir zeigen, dass

$$\sigma(\mathcal{E}_1),\mathcal{E}_2,\dots,\mathcal{E}_n \tag{5.4}$$

unabhängig sind. Durch Iteration folgt dann die Behauptung. Zu (5.4):
O.E.d.A. sei $\Omega \in \mathcal{E}_i$ (die erzeugte σ-Algebra bleibt gleich). Seien $A_2,\dots,A_n$ mit $A_i \in \mathcal{E}_i$ und sei

$$F = A_2 \cap \ldots \cap A_n, \qquad \text{also} \qquad \mathbb{P}(F) = \mathbb{P}(A_2)\cdot\ldots\cdot\mathbb{P}(A_n).$$

Setze

$$\mathcal{L} = \{A \in \mathcal{A} : \mathbb{P}(A\cap F) = \mathbb{P}(A)\mathbb{P}(F)\}.$$

Offenbar ist $\mathcal{E}_1 \subseteq \mathcal{L}$. Wir zeigen, dass $\mathcal{L}$ ein Dynkin-System ist. Dann folgt mit Satz 3.30 $\sigma(\mathcal{E}_1) = \mathcal{D}(\mathcal{E}_1) \subseteq \mathcal{L}$, da $\mathcal{E}_1$ $\cap$-stabil, und somit die Behauptung.
$\mathcal{L}$ ist ein Dynkin-System:
1. Offenbar ist $\Omega \in \mathcal{L}$.
2. Ist $A \in \mathcal{L}$, so ist

$$\mathbb{P}(A^\complement \cap F) = \mathbb{P}(F) - \mathbb{P}(A\cap F) = \mathbb{P}(F) - \mathbb{P}(A)\mathbb{P}(F) = \mathbb{P}(F)\big(1-\mathbb{P}(A)\big) = \mathbb{P}(A^\complement)\mathbb{P}(F),$$

also $A^\complement \in \mathcal{L}$.
3. Sind schließlich $B_1,B_2,\dots,\in \mathcal{L}$ paarweise disjunkt, so auch $B_1\cap F, B_2\cap F,\dots$, daher

$$\begin{aligned}\mathbb{P}\Big(\bigcup_n B_n \cap F\Big) &= \mathbb{P}\Big(\bigcup_n (B_n\cap F)\Big)\\ &= \sum_n \mathbb{P}(B_n\cap F) = \sum_n \mathbb{P}(B_n)\mathbb{P}(F) = \mathbb{P}(F)\mathbb{P}\Big(\bigcup_n B_n\Big),\end{aligned}$$

also $\bigcup_n B_n \in \mathcal{L}$. ■

Wir werden nun die *Unabhängigkeit von Zufallsvariablen* einführen.

Definition 5.8. *Seien $(\Omega_i, \mathcal{A}_i)$ Messräume und $X_i : \Omega \to \Omega_i$ Zufallsvariablen, $i = 1, \dots, n$. Dann heißen $X_1, \dots, X_n$* ***unabhängig****, falls für alle $A_i \in \mathcal{A}_i$ die Ereignisse $\{X_i \in A_i\}$, $i = 1, \dots, n$, unabhängig sind. Eine beliebige Familie von Zufallsvariablen heißt unabhängig, falls jede endliche Teilfamilie unabhängig ist.*

Proposition 5.9. *Für Zufallsvariablen $X_i : \Omega \to \Omega_i$, $i = 1, \dots, n$, sind äquivalent:*
1. $X_1, \dots, X_n$ sind unabhängig.
2. Die σ-Algebren $\sigma(X_i) := X_i^{-1}(\mathcal{A}_i) = \{X_i^{-1}(A_i),\ A_i \in \mathcal{A}_i\}$ sind unabhängig.
3. Für alle $A_i \in \mathcal{A}_i$ gilt

$$\mathbb{P}\big(X_1 \in A_1, \dots, X_n \in A_n\big) = \mathbb{P}\big(X_1 \in A_1\big) \cdot \ldots \cdot \mathbb{P}(X_n \in A_n\big).$$

Dies folgt direkt aus den Definitionen und Bemerkung 5.6 a. (für 3.). Im nächsten Resultat charakterisieren wir Unabhängigkeit reellwertiger Zufallsvariablen durch die zugehörigen Verteilungsfunktionen.

Korollar 5.10. Sind $X_1, \dots, X_n : \Omega \to \mathbb{R}$ Zufallsvariablen, so sind $X_1, \dots, X_n$ genau dann unabhängig, falls gilt

$$\mathbb{P}\big(X_1 \le x_1, \dots, X_n \le x_n\big) = \mathbb{P}\big(X_1 \le x_1\big) \cdot \ldots \cdot \mathbb{P}(X_n \le x_n\big) \quad \forall\, x_1, \dots, x_n \in \mathbb{R},$$

also falls

$$F_{(X_1,\dots,X_n)}(x_1, \dots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n) \quad \forall\, x_1, \dots, x_n \in \mathbb{R}.$$

Beweis. Das Mengensystem $\mathcal{E} = \{(-\infty, x],\ x \in \mathbb{R}\}$ ist $\cap$-stabiler Erzeuger der Borel–σ-Algebra $\mathcal{B}$ über $\mathbb{R}$, und die Bedingung besagt, dass für $x_1, \dots, x_n \in \mathbb{R}$

$$\mathbb{P}\big(X_1 \in (-\infty, x_1], \dots, X_n \in (-\infty, x_n]\big) = \mathbb{P}\big(X_1 \in (-\infty, x_1]\big) \dots \mathbb{P}\big(X_n \in (-\infty, x_n]\big).$$

Lässt man nun $x_i \to \infty$ gehen, so erhält man die Produktformel auch für Teilfamilien $\mathcal{J} \subset \{1, \dots, n\}$, und somit die Unabhängigkeit von $X_1^{-1}(\mathcal{E}), \dots, X_n^{-1}(\mathcal{E})$. ∎

Bemerkung. Sind $X_i : \Omega \to S_i$ diskret, mit S_i abzählbar, so sind $X_1, \dots, X_n$ genau dann unabhängig, falls

$$\mathbb{P}\big(X_1 = x_1, \dots, X_n = x_n\big) = \mathbb{P}\big(X_1 = x_1\big) \cdot \ldots \cdot \mathbb{P}(X_n = x_n\big) \quad \forall\, x_i \in S_i.$$

Ist etwa $S_i = \{y_{i,1}, y_{i,2}, \dots,\}$, so ist $\mathcal{E}_i = \{\emptyset, \{y_{i,1}\}, \{y_{i,2}\}, \dots\}$ ein $\cap$-stabiler Erzeuger von $\mathcal{P}(S_i)$.

5.3 Produktmaße und der Satz von Fubini

5.3.1 Die Produkt-σ-Algebra

Definition 5.11. *Seien $(\Omega_i, \mathcal{A}_i)$, $i = 1, \dots, n$, Messräume und*

$$\Omega = \Omega_1 \times \ldots \times \Omega_n.$$

Für $i = 1, \dots, n$ sei $\pi_i : \Omega \to \Omega_i$, $x = (x_1, \dots, x_n) \mapsto x_i$, die Projektion auf die i-te Koordinate. Wir definieren

$$\begin{aligned}\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n &:= \sigma\big\{\pi_i^{-1}(\mathcal{A}_i),\ i = 1, \dots, n\big\} \\ &= \sigma\big\{A_1 \times \Omega_2 \times \ldots \times \Omega_n, \dots, \Omega_1 \times \ldots \times \Omega_{n-1} \times A_n,\ A_i \in \mathcal{A}_i, i = 1, \dots, n\big\},\end{aligned}$$

die ***Produkt-σ-Algebra*** *über Ω.*

Satz 5.12. *$\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$ ist die kleinste σ-Algebra $\mathcal{A}$ über Ω, bezüglich derer alle Projektionen $\pi_i, i = 1, \ldots, n$, $(\mathcal{A} - \mathcal{A}_i)$-messbar sind.* ▪

Beweis. Die Projektionen sind genau dann messbar, wenn $\{\pi_i^{-1}(\mathcal{A}_i),\ i = 1, \ldots, n\} \subseteq \mathcal{A}$, also $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n \subseteq \mathcal{A}$, da $\mathcal{A}$ eine σ-Algebra. ■

Satz 5.13. *Sind $\mathcal{E}_i$ Erzeuger von $\mathcal{A}_i$, $i = 1, \ldots, n$, so ist $\{\pi_i^{-1}(\mathcal{E}_i),\ i = 1, \ldots, n\}$ ein Erzeuger von $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$.* ▪

Beweis. Da $\{\pi_i^{-1}(\mathcal{E}_i),\ i = 1, \ldots, n\} \subseteq \{\pi_i^{-1}(\mathcal{A}_i),\ i = 1, \ldots, n\}$, ist

$$\sigma\{\pi_i^{-1}(\mathcal{E}_i),\ i = 1, \ldots, n\} \subseteq \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$$

klar. Umgekehrt ist für jedes i nach Satz 3.3

$$\pi_i^{-1}(\mathcal{A}_i) = \sigma\{\pi_i^{-1}(\mathcal{E}_i)\} \subseteq \sigma\{\pi_j^{-1}(\mathcal{E}_i),\ j = 1, \ldots, n\}.$$

■

Satz 5.14. *Sei $(Y, \mathcal{C})$ ein weiterer Messraum, und $f = (f_1, \ldots, f_n) : Y \to \Omega$ eine Abbildung. Dann ist f genau dann $(\mathcal{C} - \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n)$-messbar, wenn $f_i, i = 1, \ldots, n$, $(\mathcal{C} - \mathcal{A}_i)$-messbar ist.* ▪

Beweis. Ist f messbar, so auch $f_i = \pi_i \circ f$ nach Satz 5.12.

Umgekehrt ist $f^{-1}(\pi_i^{-1}(\mathcal{A}_i)) = f_i^{-1}(\mathcal{A}_i) \subseteq \mathcal{C}$, falls f_i messbar, und daher f messbar, da $\{\pi_i^{-1}(\mathcal{A}_i),\ i = 1, \ldots, n\}$ Erzeuger von $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$, und es nach Korollar 3.4 genügt, Messbarkeit auf einem Erzeuger zu testen. ■

Sind (Ω_i, d_i), $i = 1, \ldots, n$, metrische Räume mit Borel-σ-Algebren $\mathcal{B}(\Omega_i)$, und $\Omega = \Omega_1 \times \ldots \times \Omega_n$. Wir metrisieren Ω durch

$$d(x, y) = \max_{i=1,\ldots,n} d_i(x_i, y_i).$$

Die Projektionen π_i sind dann stetig. Ω besitzt einerseits eine Borel-σ-Algebra $\mathcal{B}(\Omega)$ und andererseits die Produkt-σ-Algebra der einzelnen Borel-σ-Algebren $\mathcal{B}(\Omega_1) \otimes \ldots \otimes \mathcal{B}(\Omega_n)$. Es gilt praktischerweise Übereinstimmung in separablen metrischen Räumen.

Satz 5.15. *Stets ist $\mathcal{B}(\Omega_1) \otimes \ldots \otimes \mathcal{B}(\Omega_n) \subseteq \mathcal{B}(\Omega)$. Sind die (Ω_i, d_i) separabel, so auch (Ω, d), und es gilt $\mathcal{B}(\Omega_1) \otimes \ldots \otimes \mathcal{B}(\Omega_n) = \mathcal{B}(\Omega)$.* ▪

Beweis. Seien $\mathcal{O}_i$ die offenen Mengen in Ω_i und $\mathcal{O}$ die offenen Mengen in Ω. Da π_i stetig ist, gilt $\pi_i^{-1}(\mathcal{O}_i) \subseteq \mathcal{O}$, und daher nach Satz 5.13, dass

$$\mathcal{B}(\Omega_1) \otimes \ldots \otimes \mathcal{B}(\Omega_n) = \sigma\{\pi_i^{-1}(\mathcal{O}_i),\ i = 1, \ldots, n\} \subseteq \sigma(\mathcal{O}) = \mathcal{B}(\Omega).$$

Seien nun die Ω_i separabel, etwa $M_i \subseteq \Omega_i$ dicht und abzählbar in Ω_i. Dann ist

$$M = \{(x_1, \ldots, x_n)^T, \qquad x_i \in M_i\}$$

dicht in Ω, insbesondere ist auch Ω separabel. Um dies nachzuweisen, genügt es zu zeigen, dass jede offene Kugel $B_\varepsilon(x)$ in Ω einen Punkt in M enthält. Nun ist für $x = (x_1, \ldots, x_n)^T$

$$B_\varepsilon(x) = B_\varepsilon(x_1) \times \ldots \times B_\varepsilon(x_n),$$

und wir können $y_i \in B_\varepsilon(x_i)$ in M_i auswählen und haben dann $(y_1, \ldots, y_n)^T \in B_\varepsilon(x)$. Dann ist

$$B_\varepsilon(x) = \left(\pi_1^{-1}(B_\varepsilon(x_1)) \cap \ldots \cap \pi_n^{-1}(B_\varepsilon(x_n))\right) \in \left(\mathcal{B}(\Omega_1) \otimes \ldots \otimes \mathcal{B}(\Omega_n)\right),$$

und da wegen der Separabilität von Ω die offenen Kugeln $\mathcal{B}(\Omega)$ erzeugen, folgt die Gleichheit. ■

Insbesondere ist also $\mathcal{B}^d = \mathcal{B} \otimes \ldots \otimes \mathcal{B}$.

5.3.2 Produkte von Maßräumen

Seien $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ zwei Maßräume. Das Ziel ist nachfolgend, einen Produktraum

$$(\Omega, \mathcal{A}, \mu) = (\Omega_1, \mathcal{A}_1, \mu_1) \otimes (\Omega_2, \mathcal{A}_2, \mu_2)$$

zu konstruieren mit
1. $\Omega = \Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \ \omega_1 \in \Omega_1, \ \omega_2 \in \Omega_2\}$.
2. $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ die Produkt-σ-Algebra, wobei wir beachten, dass

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{H}), \quad \mathcal{H} := \left\{A_1 \times A_2 : \ A_1 \in \mathcal{A}_1, \ A_2 \in \mathcal{A}_2\right\},$$

und $\mathcal{H}$ ein $\cap$-stabiler Erzeuger von $\mathcal{A}_1 \otimes \mathcal{A}_2$,
3. $\mu = \mu_1 \otimes \mu_2$ ein möglichst eindeutig bestimmtes Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit

$$\mu(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2), \qquad A_1 \times A_2 \in \mathcal{H}.$$

Wir betrachten nachfolgend Schnitte von Mengen und Funktionen: Sei $A \subseteq \Omega = \Omega_1 \times \Omega_2$.
Schnitte von A:
Für $x_1 \in \Omega_1$ setze $A_{x_1} := \{x_2 \in \Omega_2 : \ (x_1, x_2) \in A\}$.
Für $x_2 \in \Omega_2$ setze $A^{x_2} := \{x_1 \in \Omega_1 : \ (x_1, x_2) \in A\}$.
Schnitte einer Funktion $f : \Omega \to \overline{\mathbb{R}}$:
Für $x_1 \in \Omega_1$ setze $f_{x_1}(x_2) = f(x_1, x_2)$, also $f_{x_1} : \Omega_2 \to \overline{\mathbb{R}}$.
Für $x_2 \in \Omega_2$ setze $f^{x_2}(x_1) = f(x_1, x_2)$, also $f^{x_2} : \Omega_1 \to \overline{\mathbb{R}}$.
Für $x_1 \in \Omega_1$ betrachte die Einbettung

$$\tau_{x_1} : \Omega_2 \to \Omega, \qquad \tau_{x_1}(x_2) = (x_1, x_2).$$

Für $A \subseteq \Omega$, $f : \Omega \to \overline{\mathbb{R}}$ ist dann

$$A_{x_1} = \tau_{x_1}^{-1}(A), \qquad f_{x_1} = f \circ \tau_{x_1}, \tag{5.5}$$

insbesondere vertauschen x_1-Schnittbildung und Mengenoperationen.

Satz 5.16. *1. Ist $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$, so sind $A_{x_1} \in \mathcal{A}_2$ für alle $x_1 \in \Omega_1$ und $A^{x_2} \in \mathcal{A}_1$ für alle $x_2 \in \Omega_2$.*
2. Ist $f : \Omega_1 \times \Omega_2 \to \overline{\mathbb{R}}$ $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \overline{\mathcal{B}})$-messbar, so sind f_{x_1} $(\mathcal{A}_2 - \overline{\mathcal{B}})$-messbar für alle $x_1 \in \Omega_1$ und f^{x_2} $(\mathcal{A}_1 - \overline{\mathcal{B}})$-messbar für alle $x_2 \in \Omega_2$ ▪

Beweis. Zu 1.: Wegen (5.5) ist zu zeigen: Für alle $x_1 \in \Omega_1$ ist τ_{x_1} $(\mathcal{A}_2 - \mathcal{A}_1 \otimes \mathcal{A}_2)$-messbar. Es genügt, die Messbarkeit auf einem Erzeuger zu testen. Wir zeigen daher $\tau_{x_1}^{-1}(\mathcal{H}) \subseteq \mathcal{A}_2$:
Ist $A = A_1 \times A_2 \in \mathcal{H}$, so ist $\tau_{x_1}^{-1}(A) = A_{x_1} = A_2$ falls $x_1 \in A_1$, und $\tau_{x_1}^{-1}(A) = \emptyset$ sonst. In jedem Fall ist

$\tau_{x_1}^{-1}(A) \in \mathcal{A}_2$.
Zu 2.: Nach 1. ist τ_{x_1} $(\mathcal{A}_2 - \mathcal{A}_1 \otimes \mathcal{A}_2)$-messbar, und f $(\mathcal{A}_1 \otimes \mathcal{A}_2 - \bar{\mathcal{B}})$- messbar nach Voraussetzung, also $f_{x_1} = f \circ \tau_{x_1}$ $(\mathcal{A}_2 - \bar{\mathcal{B}})$-messbar. ∎

Satz 5.17. *Ist $(\Omega_2, \mathcal{A}_2, \mu_2)$ σ-endlich, so ist für jedes $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ die Abbildung*

$$x_1 \mapsto \mu_2(A_{x_1}), \quad (\Omega_1 \to [0,\infty]),$$

eine $\mathcal{A}_1$-messbare Funktion und $\rho : \mathcal{A}_1 \otimes \mathcal{A}_2 \to [0,\infty]$,

$$\rho(A) = \int_{\Omega_1} \mu_2(A_{x_1})\, d\mu_1(x_1), \quad A \in \mathcal{A}_1 \otimes \mathcal{A}_2, \tag{5.6}$$

ist ein Maß auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit

$$\rho(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2). \tag{5.7}$$

Beweis. 1. Messbarkeit: Für $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$ sei $f_A(x_1) = \mu_2(A_{x_1})$, und

$$\mathcal{M} := \left\{ A \in \mathcal{A}_1 \otimes \mathcal{A}_2 : \ f_A \text{ ist } \mathcal{A}_1 - \text{messbar} \right\}.$$

Wir wollen zeigen, dass $\mathcal{M} = \mathcal{A}_1 \otimes \mathcal{A}_2$. Da

$$f_{A_1 \times A_2}(x_1) = \mu_2(A_2) \cdot 1_{A_1}(x_1), \tag{5.8}$$

ist $\mathcal{H} \subseteq \mathcal{M}$.
Fall A. Ist $\mu_2(\Omega_2) < \infty$, so ist $\mathcal{M}$ ein Dynkin-System:

- es ist $\Omega \in \mathcal{H} \subseteq \mathcal{M}$,

- sind $(A^{(n)})_{n\geq 1} \subseteq \mathcal{M}$ paarweise disjunkt und $A = \bigcup_{n=1}^{\infty} A^{(n)}$, so ist für $x_1 \in \Omega_1$

$$\begin{aligned} f_A(x_1) &= \mu_2\left(\left(\bigcup_{n=1}^{\infty} A^{(n)}\right)_{x_1}\right) = \mu_2\left(\bigcup_{n=1}^{\infty} A_{x_1}^{(n)}\right) \\ &= \sum_{n=1}^{\infty} \mu_2(A_{x_1}^{(n)}) = \sum_{n=1}^{\infty} f_{A^{(n)}}(x_1), \end{aligned} \tag{5.9}$$

und die Messbarkeit überträgt sich auf den punktweisen Limes.

- ist $A \in \mathcal{M}$, so auch $A^{\complement}$, denn

$$f_{A^{\complement}}(x_1) = \mu_2(A_{x_1}^{\complement}) = \mu_2(\Omega_2) - \mu_2(A_{x_1}) = \mu_2(\Omega_2) - f_A(x_1).$$

Da $\mathcal{M}$ den $\cap$-stabilen Erzeuger $\mathcal{H}$ von $\mathcal{A}_1 \otimes \mathcal{A}_2$ enthält, folgt

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\mathcal{H}) = \mathcal{D}(\mathcal{H}) \subseteq \mathcal{M},$$

also für diesen Fall $\mathcal{M} = \mathcal{A}_1 \otimes \mathcal{A}_2$.
Fall B. Ist μ_2 σ-endlich, $B_n \subseteq B_{n+1}$ mit $\bigcup_n B_n = \Omega_2$, $\mu_2(B_n) < \infty$, so sind $\mu_{2,n}(B) := \mu_2(B \cap B_n)$, $B \in \mathcal{A}_2$, endliche Maße auf Ω_2, und $f_{A,n}(x_1) = \mu_{2,n}(A_{x_1})$ messbar, und es gilt

$$f_{A,n}(x_1) \to f_A(x_1) \qquad \text{punktweise in } x_1\,.$$

2. Gleichung (5.7): Klar mit (5.8) und (5.6):

$$\rho(A_1 \times A_2) = \int_{\Omega_1} \mu_2(A_2) \cdot 1_{A_1}(x_1)\, d\mu_1(x_1) = \mu_1(A_1) \cdot \mu_2(A_2).$$

3. ρ ist σ-additiv: Sind $(A^{(n)})_{n\geq 1} \subseteq \mathcal{M}$ paarweise disjunkt und $A = \bigcup_{n=1}^{\infty} A^{(n)}$, so ist mit (5.9)

$$\begin{aligned}
\rho(A) &= \int_{\Omega_1} f_A(x_1)\, d\mu_1(x_1) = \int_{\Omega_1} \sum_{n=1}^{\infty} f_{A^{(n)}}(x_1)\, d\mu_1(x_1) \\
&= \sum_{n=1}^{\infty} \int_{\Omega_1} f_{A^{(n)}}(x_1)\, d\mu_1(x_1) \qquad \text{(monotone Konvergenz)} \\
&= \sum_{n=1}^{\infty} \rho\big(A^{(n)}\big).
\end{aligned}$$

■

Ist $(\Omega_1, \mathcal{A}_1, \mu_1)$ nicht σ-endlich, so muss ρ in (5.7) nicht eindeutig bestimmt sein, vgl. Elstrodt, Kapitel V §1.

Satz 5.18. *Sind $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ σ-endlich, so existiert genau ein Maß $\mu_1 \otimes \mu_2$ auf $\mathcal{A}_1 \otimes \mathcal{A}_2$ mit*

$$\mu_1 \otimes \mu_2(A_1 \times A_2) = \mu_1(A_1) \cdot \mu_2(A_2), \quad A_1 \times A_2 \in \mathcal{H}, \tag{5.10}$$

und es gilt für $A \in \mathcal{A}_1 \otimes \mathcal{A}_2$

$$\mu_1 \otimes \mu_2(A) = \int_{\Omega_1} \mu_2(A_{x_1})\, d\mu_1(x_1) = \int_{\Omega_2} \mu_1(A^{x_2})\, d\mu_2(x_2). \tag{5.11}$$

*$\mu_1 \otimes \mu_2$ heißt das **Produktmaß von** μ_1 **und** μ_2 und ist auch σ-endlich.* ▪

Beweis. Ist ρ wie im ersten Teil von (5.11) definiert, so ist ρ nach Satz 5.17 ein Prämaß auf $\mathcal{H}$, d.h. es erfüllt dort die Eigenschaften eines Maßes. Weiter ist ρ auf $\mathcal{H}$ σ-endlich, denn sind $A_n \subseteq A_{n+1} \in \mathcal{A}_1$, $\mu_1(A_n) < \infty$, $\bigcup_{n=1}^{\infty} A_n = \Omega_1$ sowie $B_n \subseteq B_{n+1} \in \mathcal{A}_2$, $\mu_1(B_n) < \infty$, $\bigcup_n B_n = \Omega_2$, so sind

$$A_n \times B_n \subseteq A_{n+1} \times B_{n+1} \in \mathcal{H}, \qquad \bigcup_n A_n \times B_n = \Omega, \qquad \text{und} \qquad \rho\big(A_n \times B_n\big) < \infty.$$

Daher hat ρ eine eindeutige Fortsetzung (Maßeindeutigkeit und Maßfortsetzung), die nach Satz 5.17 durch (5.11) (Symmetriegründe) gegeben ist. ■

5.3.3 Der Satz von Fubini

Seien im folgenden $(\Omega_1, \mathcal{A}_1, \mu_1)$ und $(\Omega_2, \mathcal{A}_2, \mu_2)$ zwei σ-endliche Maßräume und $(\Omega, \mathcal{A}, \mu) = (\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2, \mu_1 \otimes \mu_2)$ der Produktraum. Wir zeigen zunächst den Satz für nicht-negative messbare numerische Funktionen.

Satz 5.19. *Ist $f : \Omega \to [0, \infty]$ eine $\mathcal{A}$-messbare Funktion, so sind*

$$\begin{aligned}
x_1 &\mapsto \int_{\Omega_2} f(x_1, x_2)\, d\mu_2 \qquad \textit{eine } \mathcal{A}_1 - \textit{messbare Funktion}, \\
x_2 &\mapsto \int_{\Omega_1} f(x_1, x_2)\, d\mu_1 \qquad \textit{eine } \mathcal{A}_2 - \textit{messbare Funktion},
\end{aligned}$$

und es gilt

$$\int_\Omega f\,d\mu = \int_{\Omega_1}\Big(\int_{\Omega_2} f(x_1,x_2)\,d\mu_2(x_2)\Big)d\mu_1(x_1)$$
$$= \int_{\Omega_2}\Big(\int_{\Omega_1} f(x_1,x_2)\,d\mu_1(x_1)\Big)d\mu_2(x_2).$$

Beweis. Für Indikatorfunktionen $f = \mathbf{1}_A, A \in \mathcal{A}$, ist dies Formel (5.11). Mit Linearität folgt die Aussage dann bereits für alle $f \in \mathcal{T}^+(\Omega,\mathcal{A})$. Für $f \in \mathcal{M}^+(\Omega,\mathcal{A})$ existiert eine wachsende Folge von Treppenfunktionen $(f^{(n)}) \subseteq \mathcal{T}^+(\Omega,\mathcal{A})$ mit $f^{(n)} \uparrow f$. Nach Satz 5.16 ist $f_{x_1} \in \mathcal{M}^+(\Omega_2,\mathcal{A}_2)$ und außerdem sind $f_{x_1}^{(n)} \in \mathcal{T}^+(\Omega_2,\mathcal{A}_2)$ mit $f_{x_1}^{(n)} \uparrow f_{x_1}$. Nach Definition des Integrals gilt

$$\int_{\Omega_2} f_{x_1}^{(n)}\,d\mu_2 \uparrow \int_{\Omega_2} f_{x_1}\,d\mu_2\,,$$

und es folgt, dass die rechte Seite als Grenzwert wiederum $(\mathcal{A}_1 - \overline{\mathcal{B}})$-messbar ist. Mit der Definition des Integrals und der Aussage des Satzes für Treppenfunktionen folgt

$$\begin{aligned}\int_\Omega f\,d\mu &= \lim_{n\to\infty}\int_\Omega f^{(n)}\,d\mu\\ &= \lim_{n\to\infty}\int_{\Omega_1}\Big(\int_{\Omega_2} f_{x_1}^{(n)}\,d\mu_2\Big)d\mu_1\\ &= \int_{\Omega_1}\Big(\lim_{n\to\infty}\int_{\Omega_2} f_{x_1}^{(n)}\,d\mu_2\Big)d\mu_1 && \text{(monotone Konvergenz)}\\ &= \int_{\Omega_1}\Big(\int_{\Omega_2} f_{x_1}\,d\mu_2\Big)d\mu_1\,.\end{aligned}$$

Entsprechend schließt man unter vertauschten Rollen von μ_1 und μ_2. ■

Satz 5.20 (Satz von Fubini). *Ist $f:\Omega\to\overline{\mathbb{R}}$ integrierbar, so sind $f_{x_1}:\Omega_2\to\overline{\mathbb{R}}$ μ_2-integrierbar μ_1-f.ü., $f^{x_2}:\Omega_1\to\overline{\mathbb{R}}$ μ_1-integrierbar μ_2-f.ü.*

$$A = \Big\{x_1\in\Omega_1:\ f_{x_1} \text{ ist nicht } \mu_2-\text{integrierbar}\Big\}\in\mathcal{A}_1,$$
$$B = \Big\{x_2\in\Omega_2:\ f^{x_2} \text{ ist nicht } \mu_1-\text{integrierbar}\Big\}\in\mathcal{A}_2,$$

$$x_1 \mapsto \int_{\Omega_2} f(x_1,x_2)\,d\mu_2(x_2) \quad \text{ist} \quad \mu_1-\text{integrierbar über } A^{\complement},$$
$$x_2 \mapsto \int_{\Omega_1} f(x_1,x_2)\,d\mu_1(x_1) \quad \text{ist} \quad \mu_2-\text{integrierbar über } B^{\complement},$$

und es gilt

$$\int_\Omega f\,d\mu = \int_{A^{\complement}}\Big(\int_{\Omega_2} f(x_1,x_2)\,d\mu_2(x_2)\Big)d\mu_1(x_1)$$
$$= \int_{B^{\complement}}\Big(\int_{\Omega_1} f(x_1,x_2)\,d\mu_1(x_1)\Big)d\mu_2(x_2).$$

Beweis. Mit f ist auch $|f|$ integrierbar und mit Satz 5.19:

$$\int_\Omega |f|\,d\mu = \int_{\Omega_1}\Big(\int_{\Omega_2}|f|\,d\mu_2\Big)\,d\mu_1 < \infty\,.$$

$\Big(\int_{\Omega_2}|f|\,d\mu_2\Big) \in \mathcal{M}(\Omega_1,\mathcal{A}_1)$ nach Satz 5.19 und aus der Endlichkeit des Integrals folgt $\int_{\Omega_2}|f|\,d\mu_2 < \infty$ μ_1-fast überall. Die Ausnahmemenge sei

$$A = \Big\{x_1 \in \Omega_1 : \int_{\Omega_2}|f_{x_1}(x_2)|d\mu_2(x_2) = \infty\Big\}\,,$$

eine μ_1-Nullmenge. Nun können wir f in Positiv- und Negativteil zerlegen; und die Aussage auf Satz 5.19 zurückführen. Für $x_1 \in A^\complement$ zerlegen wir

$$\int_{\Omega_2} f(x_1,x_2)\,d\mu_2(x_2) = \int_{\Omega_2} f^+(x_1,x_2)\,d\mu_2(x_2) - \int_{\Omega_2} f^-(x_1,x_2)\,d\mu_2(x_2)\,,$$

wobei die Integrale rechts Funktionen (in x_1) aus $\mathcal{M}^+(\Omega_1,\mathcal{A}_1)$ darstellen. Diese sind μ_1-integrierbar und wir schließen mit $\mu_1(A) = 0$:

$$\begin{aligned}
&\int_{A^\complement}\Big(\int_{\Omega_2} f(x_1,x_2)\,d\mu_2(x_2)\Big)\,d\mu_1(x_1)\\
&= \int_{\Omega_1}\Big(\int_{\Omega_2} f^+(x_1,x_2)\,d\mu_2(x_2)\Big)\,d\mu_1(x_1) - \int_{\Omega_1}\Big(\int_{\Omega_2} f^-(x_1,x_2)\,d\mu_2(x_2)\Big)\,d\mu_1(x_1)\\
&= \int_\Omega f^+\,d\mu - \int_\Omega f^-\,d\mu = \int_\Omega f\,d\mu\,.
\end{aligned}$$

Entsprechend schließt man unter vertauschten Rollen von μ_1 und μ_2. ■

Korollar 5.21 (Tonelli)**.** Ist $f : \Omega \to \overline{\mathbb{R}}$ eine $\mathcal{A}$-messbare Funktion, und ist eines der Integrale

$$\int_\Omega |f|\,d\mu, \qquad \int_{\Omega_1}\Big(\int_{\Omega_2}|f(x_1,x_2)|\,d\mu_2(x_2)\Big)d\mu_1(x_1),$$
$$\int_{\Omega_2}\Big(\int_{\Omega_1}|f(x_1,x_2)|\,d\mu_1(x_1)\Big)d\mu_2(x_2)$$

endlich, so ist f integrierbar, und es gilt Satz 5.20.

Ein für die Stochastik sehr nützliches Beispiel zur Anwendung des Satzes von Fubini ist die Berechnung von Erwartungswerten und Momenten über Tail-Wahrscheinlichkeiten.

Satz 5.22. *Sei* $(\Omega,\mathcal{A},\mathbb{P})$ *ein Wahrscheinlichkeitsraum,* $X : \Omega \to [0,\infty)$ *eine nicht-negative Zufallsvariable,* $\phi : [0,\infty) \to [0,\infty)$ *monoton wachsend, stetig, auf* $(0,\infty)$ *stetig differenzierbar und* $\phi(0) = 0$. *Dann gilt für alle* $A \in \mathcal{A}$

$$\int_A \phi(X)\,d\mathbb{P} = \int_0^\infty \phi'(t)\,\mathbb{P}\big(\{X \geq t\} \cap A\big)\,\mathrm{d}t. \tag{5.12}$$

Beweis. Ohne Einschränkung der Allgemeinheit sei $A = \Omega$. Sonst könnten wir für $\mathbb{P}(A) > 0$ zu $\mathbb{P}_A(B) = \mathbb{P}(B\cap A)/\mathbb{P}(A)$ übergehen, $B \in \mathcal{A}$. Falls $\mathbb{P}(A) = 0$, ist die Formel klar. Nach dem Satz von der monotonen Konvergenz gilt

$$\int_{(0,a]}\phi'(t)\,\mathrm{d}t = \lim_{n\to\infty}\int_{[1/n,a]}\phi'(t)\,\mathrm{d}t = \phi(a) - \lim_{n\to\infty}\phi(1/n) = \phi(a)\,.$$

Somit ist

$$\int_{(0,X(\omega)]} \phi'(t)\,\mathrm{d}t = \phi\big(X(\omega)\big), \qquad \omega \in \Omega$$

und daher

$$\begin{aligned}\int_\Omega \phi \circ X \, d\mathbb{P} &= \int_\Omega \Big(\int_{(0,X(\omega)]} \phi'(t)\,\mathrm{d}t \Big)\, d\mathbb{P} \\ &= \int_\Omega \Big(\int_{(0,\infty)} \mathbf{1}_{(0,X(\omega)]}(t)\phi'(t)\,\mathrm{d}t \Big)\, d\mathbb{P}.\end{aligned}$$

Da die Menge $\{(\omega,t) \in \Omega \times (0,\infty) : t \leq X(\omega)\}$ $\mathcal{A} \otimes \mathcal{B}$-messbar ist (als Urbild von $\{(x,y) \in (0,\infty)^2 : x \geq y\}$ unter der messbaren Abbildung $(\omega,t) \mapsto (X(\omega),t)$), ist der Integrand in obiger Formel $(\mathcal{A} \otimes \mathcal{B} - \mathcal{B})$ messbar und nicht-negativ. Somit folgt mit dem ersten Satz von Fubini, Satz 5.19, dass

$$\begin{aligned}\int_\Omega \phi \circ X \, d\mathbb{P} &= \int_{(0,\infty)} \int_\Omega \mathbf{1}_{\{(\omega,t)\in\Omega\times(0,\infty):t\leq X(\omega)\}} \phi'(t)\, d\mathbb{P}\,\mathrm{d}t \\ &= \int_{(0,\infty)} \phi'(t)\,\mathbb{P}\big(t \leq X\big)\,\mathrm{d}t.\end{aligned}$$

■

Bemerkung. Die Funktion $t \mapsto \mathbb{P}(X \geq t)$ ist monoton fallend und linksstetig, und hat daher höchstens abzählbar viele Sprungstellen, nämlich gerade bei $\mathbb{P}(X \geq t) > \mathbb{P}(X > t)$. Daher ist $\mathbb{P}(X \geq t) = \mathbb{P}(X > t)$ Lebesgue-fast überall, und somit stimmen auch die Integrale überein:

$$\int_0^\infty \phi'(t)\,\mathbb{P}\big(\{X \geq t\} \cap A\big)\,\mathrm{d}t = \int_0^\infty \phi'(t)\,\mathbb{P}\big(\{X > t\} \cap A\big)\,\mathrm{d}t.$$

Beispiel 5.23. Sei $\phi(t) = t^p$, $p > 0$, und X eine reellwertige Zufallsvariable. Dann ist nach (5.12)

$$\mathbb{E}\big[|X|^p\big] = \int_0^\infty p t^{p-1}\,\mathbb{P}\big(\{|X| \geq t\}\big)\,\mathrm{d}t.$$

Insbesondere folgt

$$\mathbb{E}[|X|] = \int_0^\infty \mathbb{P}\big(\{|X| \geq t\}\big)\,\mathrm{d}t.$$

Für $X \sim \mathrm{Exp}(\lambda)$ folgt etwa $\mathbb{E}[X] = \int_0^\infty e^{-\lambda x}\,\mathrm{d}x = -\lambda^{-1}(e^{-\infty} - e^0) = \lambda^{-1}$.

Es kann hieraus auch auf die Existenz von Erwartungswerten geschlossen werden: Gilt zum Beispiel für eine reellwertige Zufallsvariable $X : \Omega \to \mathbb{R}$

$$\mathbb{P}(|X| \geq t) \leq C t^{-\alpha}, \quad t \geq 1,$$

für ein $\alpha > 0$, so gilt für $0 < \beta < \alpha$

$$\begin{aligned}\mathbb{E}\big[|X|^\beta\big] &= \int_0^\infty \beta t^{\beta-1}\,\mathbb{P}(|X| \geq t)\,\mathrm{d}t \\ &\leq \beta\Big(\int_0^1 t^{\beta-1}\,\mathrm{d}t + C\int_1^\infty t^{-\alpha-1+\beta}\,\mathrm{d}t\Big) \leq 1 + C\beta/(\alpha-\beta).\end{aligned}$$

Ferner ergibt ein Vergleich von Integral und Summe, dass

$$\mathbb{E}[|X|] < \infty \Leftrightarrow \sum_{n=1}^\infty \mathbb{P}(|X| \geq n) < \infty. \tag{5.13}$$

Falls X Werte in $\mathbb{N}_0$ hat, so gilt dass $\mathbb{E}[X] = \sum_{n=1}^\infty \mathbb{P}(X \geq n)$.

5.4 Unabhängige Zufallsvariablen und Produkträume

Sind $(\Omega_i, \mathcal{A}_i)$ Messräume, so ist nach dem letzten Kapitel der *Produktmessraum* gegeben durch
1. die Grundmenge $\Omega_1 \times \ldots \times \Omega_n$,
2. die σ-Algebra $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n = \sigma(\mathcal{H})$, $\mathcal{H} = \{A_1 \times \ldots \times A_n,\ A_i \in \mathcal{A}_i\}$.
Das Mengensystem $\mathcal{H}$ ist dabei ein $\cap$-stabiler Erzeuger von $\mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n$.

Ist $(\Omega, \mathcal{A})$ ein weiterer Messraum und $X_i : \Omega \to \Omega_i$, $i = 1, \ldots, n$ Abbildungen, so sind äquivalent:
1. X_i ist $(\mathcal{A} - \mathcal{A}_i)$-messbar, $i = 1, \ldots, n$.
2. $X = (X_1, \ldots, X_n)^T : \Omega \to \Omega_1 \times \ldots \times \Omega_n$ ist $(\mathcal{A} - \mathcal{A}_1 \otimes \ldots \otimes \mathcal{A}_n)$-messbar.

Ist also insbesondere $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, und sind $X_i : \Omega \to \Omega_i$ Zufallsvariablen, so ist der Vektor $X = (X_1, \ldots, X_n)^T$ eine Zufallsvariable mit Werten in dem Produktraum. Seine Verteilung P_X heißt *gemeinsame Verteilung* der Zufallsvariablen $X_1, \ldots, X_n$.

Sind $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ Wahrscheinlichkeitsräume, $i = 1, \ldots, n$, so ist der *Produktwahrscheinlichkeitsraum*

$$(\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \otimes \ldots \otimes (\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$$

der Produktmessraum mit dem *Produktmaß* $\bigotimes_{i=1}^n \mathbb{P}_i = \mathbb{P}_1 \otimes \ldots \otimes \mathbb{P}_n$ als Wahrscheinlichkeitsmaß:
3. $\bigotimes_{i=1}^n \mathbb{P}_i\big(A_1 \times \ldots \times A_n\big) = \mathbb{P}_1(A_1) \cdot \ldots \cdot \mathbb{P}_n(A_n), \qquad A_i \in \mathcal{A}_i$.

Es existiert und ist durch die Eigenschaft 3. eindeutig bestimmt. Aus den Definitionen und Satz 5.9 folgt dann unmittelbar

Satz 5.24. *Für Zufallsvariablen $X_i : \Omega \to \Omega_i$, $i = 1, \ldots, n$, sind äquivalent:*
1. $X_1, \ldots, X_n$ sind unabhängig.
2. Die gemeinsame Verteilung von $(X_1, \ldots, X_n)$ auf $\Omega_1 \times \ldots \times \Omega_n$ ist das Produktmaß:

$$\mathbb{P}_{(X_1, \ldots, X_n)} = \mathbb{P}_{X_1} \otimes \ldots \otimes \mathbb{P}_{X_n}.$$

Bemerkung.
- Unabhängigkeit hängt nur von der gemeinsamen Verteilung ab.
- Unabhängigkeit legt zusammen mit den univariaten Verteilungen der X_i die gemeinsame Verteilung von $(X_1, \ldots, X_n)^\mathrm{T}$ fest.

Für absolutstetige, reellwertige Zufallsvariablen ergibt sich folgender wichtiger Zusammenhang.

Satz 5.25 (Produktdichte)**.** *Sei $\mathbf{X} = (X_1, \ldots, X_n)^\mathrm{T}$, mit reellwertigen Zufallsvariablen X_i, mit der Dichte $f : \mathbb{R}^n \to \mathbb{R}$. Sei weiterhin f_i die Dichte von X_i für alle $i = 1, \ldots, n$. Dann sind äquivalent:*

1. *$X_1, \ldots, X_n$ sind unabhängig.*
2. *Es gilt für $\mathbf{x} \in \mathbb{R}^n$ Lebesgue-fast überall*

$$f(\mathbf{x}) = f_1(x_1) \cdot \ldots \cdot f_n(x_n).$$

Beweis. 1. Gilt für $\mathbf{x} \in \mathbb{R}^n$ fast überall $f(\mathbf{x}) = f_1(x_1) \cdot \ldots \cdot f_n(x_n)$, so folgt für $I_1, \ldots, I_n \in \mathcal{B}$

$$\begin{aligned}
\mathbb{P}(X_1 \in I_1, \ldots, X_n \in I_n) = \mathbb{P}(\mathbf{X} \in I_1 \times \cdots \times I_n) &= \int_{I_1 \times \cdots \times I_n} f(\mathbf{x})\,\mathrm{d}\mathbf{x} \\
&= \int_{I_1} \cdots \int_{I_n} f((x_1, \ldots, x_n)^\mathrm{T})\,\mathrm{d}x_1 \cdots \mathrm{d}x_n \\
&= \int_{I_1} \cdots \int_{I_n} f_1(x_1) \cdot \ldots \cdot f_n(x_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_n
\end{aligned}$$

$$
\begin{aligned}
&= \int_{I_1} f_1(x_1)\,\mathrm{d}x_1 \cdot \dots \cdot \int_{I_n} f_n(x_n)\,\mathrm{d}x_n \\
&= \mathbb{P}(X_1 \in I_1) \cdot \dots \cdot \mathbb{P}(X_n \in I_n),
\end{aligned}
$$

also sind $X_1, \dots, X_n$ unabhängig.

2. Sind $X_1, \dots, X_n$ unabhängig, so gilt für $I_1, \dots, I_n \in \mathcal{B}$

$$
\begin{aligned}
\mathbb{P}(\mathbf{X} \in I_1 \times \dots \times I_n) &= \mathbb{P}(X_1 \in I_1, \dots, X_n \in I_n) \\
&= \mathbb{P}(X_1 \in I_1) \cdot \dots \cdot \mathbb{P}(X_n \in I_n) = \int_{I_1} f_1(x_1)\,\mathrm{d}x_1 \cdot \dots \cdot \int_{I_n} f_n(x_n)\,\mathrm{d}x_n \\
&= \int_{I_1 \times \dots \times I_n} f_1(x_1) \cdot \dots \cdot f_n(x_n)\,\mathrm{d}\mathbf{x}.
\end{aligned}
$$

Damit ist $f(\mathbf{x}) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$ als Dichte von $\mathbf{X}$ fast überall eindeutig bestimmt. ■

Beispiel 5.26. 1. Ist $\mathbf{X} = (X_1, \dots, X_n)^{\mathrm{T}} \sim U([a,b])$ uniform verteilt auf dem Quader

$$
[\mathbf{a}, \mathbf{b}] = \{\mathbf{x} \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i \text{ für } i = 1, \dots, n\}, \quad \mathbf{a}, \mathbf{b} \in \mathbb{R}^n, \quad a_i < b_i \text{ für } i = 1, \dots, n,
$$

so ist die Dichte f von $\mathbf{X}$

$$
f(\mathbf{x}) = \frac{1}{\prod_{i=1}^n (b_i - a_i)} \prod_{i=1}^n 1_{[a_i,b_i]}(x_i) = \prod_{i=1}^n \frac{1_{[a_i,b_i]}(x_i)}{b_i - a_i} = \prod_{i=1}^n f_i(x_i),
$$

wobei

$$
f_i(x_i) = \frac{1_{[a_i,b_i]}(x_i)}{b_i - a_i}, \quad i = 1, \dots, n
$$

je die Dichte von X_i ist. Also sind $X_1, \dots, X_n$ unabhängig.

2. Ist $\mathbf{Z} = (X,Y)^\top \sim U(K)$ uniform verteilt auf dem Kreis $K = \{(x,y)^\top \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$, so ist mit den Dichten f, f_X, f_Y von $\mathbf{Z}, X, Y$ für $(x,y)^\top \in (-1,1) \setminus K$ stets

$$
0 = \frac{1}{\pi} 1_{\{x^2+y^2 \leq 1\}} = f(x,y) \neq f_X(x) f_Y(y) = \frac{4}{\pi^2} \underbrace{\sqrt{1-x^2}}_{>0} \underbrace{\sqrt{1-y^2}}_{>0} > 0,
$$

also nicht fast überall $f(x,y) = f_X(x) f_Y(y)$. Damit sind X, Y nicht unabhängig.

Satz 5.27. *Sind $X_1, \dots, X_n : \Omega \to \Omega'$ unabhängige Zufallsvariablen und $h_1, \dots, h_n : \Omega' \to \mathbb{R}$ messbar, so sind auch $h_1(X_1), \dots, h_n(X_n)$ unabhängig.* ▪

Beweis. Dies folgt aus

$$
\sigma(h_j \circ X_j) = (h_j \circ X_j)^{-1}(\mathcal{B}) = X_j^{-1}(h_j^{-1}(\mathcal{B})) \subseteq \sigma(X_j),
$$

da wegen Proposition 5.9 Teilmengen(systeme) von unabhängigen Mengensystemen stets wieder unabhängig sind. ■

Wir sehen also, dass Funktionen unabhängiger Zufallsvariablen wieder unabhängig sind.

Satz 5.28 (Blockungslemma)**.** *Sind $X_1, \dots, X_n : \Omega \to \mathbb{R}$ unabhängige Zufallsvariablen, $r \in \{1, \dots, n-1\}$, $g : \mathbb{R}^r \to \mathbb{R}$ und $h : \mathbb{R}^{n-r} \to \mathbb{R}$, so sind $Y = g(X_1, \dots, X_r)$ und $Z = h(X_{r+1}, \dots, X_n)$ unabhängig.* ▪

Beweis. Es gilt, dass

$$\sigma(X_1,\dots,X_r) = \sigma\Big(\bigcup_{j=1}^{r}\sigma(X_j)\Big),$$

sowie

$$\sigma(X_{r+1},\dots,X_n) = \sigma\Big(\bigcup_{j=r+1}^{n}\sigma(X_j)\Big).$$

Mit Satz 5.27 führen wir den Satz damit auf folgendes Lemma zurück, welches besagt, dass disjunkte Blöcke unabhängiger, $\cap$-stabiler Mengensysteme unabhängig sind. ∎

Lemma 5.29. *Ist $(\mathcal{E}_i)_{i\in I}$, $\mathcal{E}_i \subseteq \mathcal{A}$ eine unabhängige Familie von $\cap$-stabilen Mengensystemen, $(I_j)_{j\in J}$ eine disjunkte Zerlegung von I, und $\mathcal{A}_j = \sigma\big(\bigcup_{i\in I_j}\mathcal{E}_i\big)$, so sind auch $(\mathcal{A}_j)_{j\in J}$ unabhängig.*

Beweis. Für $j\in J$ sei

$$\tilde{\mathcal{E}}_j = \Big\{\bigcap_{k\in K} E_k,\ K\subseteq I_j \text{ endlich}, E_k\in\mathcal{E}_k, k\in K\Big\}.$$

Wir zeigen, dass
1. $\sigma(\tilde{\mathcal{E}}_j) = \mathcal{A}_j$,
2. $\tilde{\mathcal{E}}_j$ ist $\cap$-stabil,
3. Die $\tilde{\mathcal{E}}_j$, $j\in J$, sind unabhängig.

Zu 1.: $\tilde{\mathcal{E}}_j \subseteq \sigma(\mathcal{E}_i, i\in I_j)$ klar. Umgekehrt: Da $\bigcup_{i\in I_j}\mathcal{E}_i \subseteq \tilde{\mathcal{E}}_j$.

Zu 2.: Seien $K_1, K_2 \in I_j$ endlich,

$$\tilde{E} = \bigcap_{k\in K_1} E_k, \qquad \tilde{F} = \bigcap_{l\in K_2} F_l, \qquad E_k\in\mathcal{E}_k, F_l\in\mathcal{E}_l.$$

Dann folgt

$$\tilde{E}\cap\tilde{F} = \bigcap_{k\in K_1\setminus K_2} E_k \cap \bigcap_{l\in K_2\setminus K_1} F_l \cap \bigcap_{k\in K_2\cap K_1}(E_k\cap F_k) \in \tilde{\mathcal{E}}_j,$$

da $E_k\cap F_k\in\mathcal{E}_k$, da $\mathcal{E}_k$ $\cap$-stabil.

Zu 3.: Wir können $\Omega\in\mathcal{E}_i$, $i\in I$ annehmen. Sind dann $j_1,\dots,j_n\in J$,

$$\tilde{E}_{j_l} = \bigcap_{k\in K_l} E_k, \qquad E_k\in\mathcal{E}_k, k\in K_l, \quad K_l\subseteq I_{j_l}, \quad l=1,\dots,n,$$

so sind insbesondere $K_1,\dots,K_n$ paarweise disjunkt, und wegen der Unabhängigkeit der $\mathcal{E}_i$ folgt

$$\begin{aligned}\mathbb{P}\Big(\bigcap_{l=1}^{n}\tilde{E}_{j_l}\Big) &= \mathbb{P}\Big(\bigcap_{l=1}^{n}\bigcap_{k\in K_l}E_k\Big) = \prod_{l=1}^{n}\prod_{k\in K_l}\mathbb{P}(E_k)\\ &= \prod_{l=1}^{n}\mathbb{P}\Big(\bigcap_{k\in K_l}E_k\Big) = \prod_{l=1}^{n}\mathbb{P}(\tilde{E}_{j_l})\end{aligned}$$

Dann folgt die Unabhängigkeit für jede beliebige endliche Teilfamilie mit Satz 5.7. Also folgt die Unabhängigkeit der $\tilde{\mathcal{E}}_j$, $j\in J$, und aus 1.-3. der Satz. ∎

Aus dem Beweis ergibt sich, dass eine analoge Aussage wie im vorangehenden Satz auch für Unterteilungen in mehr als zwei Blöcke gilt.

Satz 5.30 (Faltung)**.** *Seien $X,Y:\Omega\to\mathbb{R}^d$ unabhängige Zufallsvariablen. Dann gilt für $A,B\in\mathcal{B}^d$*

$$\mathbb{P}(X+Y\in B, Y\in A)=\int_A P_X(B-y)\,dP_Y(y),$$

insbesondere

$$\mathbb{P}(X+Y\in B)=\int_{\mathbb{R}^d} P_X(B-y)\,dP_Y(y)=\int_{\mathbb{R}^d} P_Y(B-x)\,dP_X(x)\,. \tag{5.14}$$

Beweis. Nach der Transformationsformel (4.13) folgt

$$\begin{aligned}\mathbb{P}(X+Y\in B, Y\in A)&=\int_{\mathbb{R}^{2d}}\mathbf{1}_B(x+y)\,\mathbf{1}_A(y)\,d(P_X\otimes P_Y)(x,y)\\&=\int_A\Big(\int_{\mathbb{R}^d}\mathbf{1}_{B-y}(x)\,dP_X(x)\Big)dP_Y(y)\\&=\int_A P_X(B-y)\,dP_Y(y)\,.\end{aligned}$$
∎

Die Verteilung P_{X+Y} von $X+Y$ in (5.14) nennt man die *Faltung $P_X * P_Y$* der Verteilungen P_X und P_Y.

Korollar 5.31. Sind die Zufallsvariablen $X,Y:\Omega\to\mathbb{Z}$ unabhängig, so gilt

$$\mathbb{P}(X+Y=k)=\sum_{j\in\mathbb{Z}}\mathbb{P}(X=j)\,\mathbb{P}(Y=k-j)\ ,\ k\in\mathbb{Z}.$$

Sind $X,Y:\Omega\to\mathbb{N}_0$ unabhängig, so ist

$$\mathbb{P}(X+Y=k)=\sum_{j=0}^{k}\mathbb{P}(X=j)\,\mathbb{P}(Y=k-j)\ ,\ k\in\mathbb{N}_0.$$

Beispiel 5.32. Sind $X\sim\mathrm{Poi}(\lambda)$ und $Y\sim\mathrm{Poi}(\mu)$ unabhängig, so gilt $X+Y\sim\mathrm{Poi}(\lambda+\mu)$.

Für Dichten erhält man die folgende *Faltungsformel*.

Korollar 5.33. Sind $X,Y:\Omega\to\mathbb{R}$ unabhängige, reellwertige und absolutstetige Zufallsvariablen mit den Dichten f und g. Dann ist auch $X+Y$ absolutstetig mit der Dichte

$$(f*g)(z)=\int_{\mathbb{R}} f(z-t)g(t)\,\mathrm{d}t\,.$$

Bemerkung. Es gilt $(f*g)(z)=(g*f)(z)$.

Beispiel 5.34.

1. Sind $X\sim U([0,1])$ und $Y\sim U([0,1])$ unabhängig mit Dichte f, so ist die Dichte von $X+Y$ die in Abbildung 5.1 gezeigte Funktion

$$g(x)=(f*f)(x)=(1-|x-1|)\,1_{[0,2]}(x).$$

2. Sind $X\sim\mathcal{N}(\mu_1,\sigma_1^2)$ und $Y\sim\mathcal{N}(\mu_2,\sigma_2^2)$ unabhängig, so gilt $X+Y\sim\mathcal{N}(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$. Das kann mit der Faltungsformel bewiesen werden. In der Stochastik II bekommen wir das einfacher, indem wir die Addition als lineare Transformation von $(X,Y)^{\mathrm{T}}$ auffassen mit einem Satz über lineare Transformation der mutlivariaten Normalverteilung.

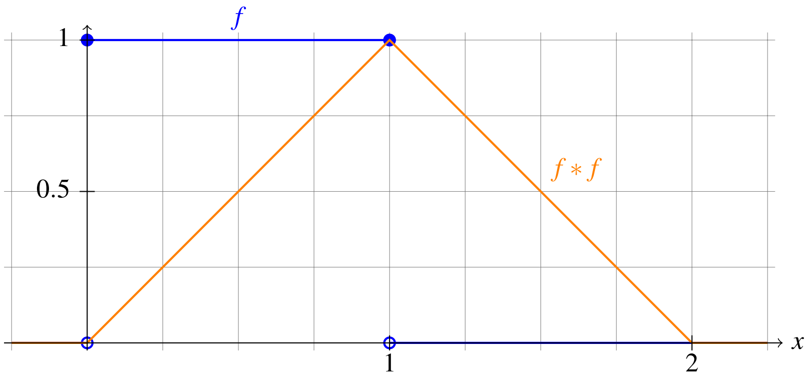


Abbildung 5.1: Verteilung der Summe zweier unabhängiger $U([0,1])$-Zufallsvariablen

Zufallsvariablen $X_1,\dots,X_n : \Omega \to \mathbb{R}$ heißen *unabhängig identisch verteilt*, abgekürzt *u.i.v.*, falls sie unabhängig sind und ihre Verteilungen $\mathbb{P}_{X_1} = \ldots = \mathbb{P}_{X_n}$ übereinstimmen.
Es sind $X_1,\dots,X_n : \Omega \to \mathbb{R}$ u.i.v. genau dann, wenn ihre gemeinsame Verteilung das n-fache Produkt $\mathbb{P}_{X_1}^{\otimes n}$ auf $\big(X_1(\Omega)\big)^n$ ist, vgl. Satz 5.24.

Nachfolgend betrachten wir Unabhängigkeit und Erwartungswerte.

Satz 5.35. *Seien $X : \Omega \to \Omega_1$, $Y : \Omega \to \Omega_2$ unabhängige Zufallsvariablen. Sind $f : \Omega_1 \to \mathbb{R}$ und $g : \Omega_2 \to \mathbb{R}$ messbar, und gilt $f(X), g(Y) \geq 0$ oder $f(X)$ und $g(Y)$ integrierbar, so ist*

$$\mathbb{E}\big[f(X)g(Y)\big] = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)]\,. \tag{5.15}$$

Beweis. 1. $f(X), g(Y) \geq 0$: Es ist

$$\begin{aligned}\mathbb{E}\big[f(X)\,g(Y)\big] &= \int\int f(x)g(y)\,dP_X(x)\,dP_Y(y)\\ &= \int g(y)\int f(x)\,dP_X(x)\,dP_Y(y)\\ &= \int g(y)\mathbb{E}[f(X)]\,dP_Y(y) = \mathbb{E}[f(X)]\,\mathbb{E}[g(Y)]\,.\end{aligned}$$

2. Die obige Rechnung zeigt für $|f|$ und $|g|$, dass $\mathbb{E}\big[|f(X)g(Y)|\big] = \mathbb{E}\big[|f(X)|\big]\,\mathbb{E}\big[|g(Y)|\big]$. Dies zeigt, dass $f(x)g(y)$ bzgl. des Produktmaßes integrierbar ist, und das obige Argument kann mit Fubini wiederholt werden. ∎

Bemerkung. Durch Betrachten von Indikatorfunktionen $f = \mathbf{1}_A$, $g = \mathbf{1}_B$ folgt, dass X und Y genau dann unabhängig sind, falls für alle messbaren, beschränkten f, g (5.15) gilt.

Korollar 5.36. Sind $X_1,\dots,X_n$ unabhängig und gilt $X_i \geq 0$, $i = 1,\dots,n$, oder X_i integrierbar, $i = 1,\dots,n$, so gilt

$$\mathbb{E}\Big[\prod_{i=1}^n X_i\Big] = \prod_{i=1}^n \mathbb{E}[X_i]\,.$$

Satz 5.37. 1. *Sind $X,Y:\Omega\to\mathbb{R}$ unabhängig mit $\mathbb{E}[X^2]<\infty$ und $\mathbb{E}[Y^2]<\infty$, so gilt*

$$\operatorname{Cov}(X,Y)=0\,.$$

2. *Für unabhängige $X_1,\dots,X_n:\Omega\to\mathbb{R}$ mit $\mathbb{E}[X_i^2]<\infty$ für $i\in\{1,\dots,n\}$, gilt*

$$\operatorname{Var}(X_1+\ldots+X_n)=\sum_{i=1}^n\operatorname{Var}(X_i)\,.$$

Beweis. 1. Nach dem Korollar gilt, dass

$$\operatorname{Cov}(X,Y)=\mathbb{E}[X\cdot Y]-\mathbb{E}[X]\cdot\mathbb{E}[Y]=\mathbb{E}[X]\cdot\mathbb{E}[Y]-\mathbb{E}[X]\cdot\mathbb{E}[Y]=0\,.$$ ■

2. Mit 1. folgt unmittelbar aus Satz 5.2 (5.) die Aussage.

Nach dem vorangehenden Satz sind unabhängige Zufallsvariablen mit existierenden zweiten Momenten unkorreliert. Die Umkehrung gilt im Allgemeinen nicht, also unkorrelierte Zufallsvariablen müssen nicht unabhängig sein. Betrachte als Gegenbeispiel für zwei unabhängige Zufallsvariablen X und Y mit $\operatorname{Var}(X)=\operatorname{Var}(Y)$, etwa die Augenzahlen beim zweifachen Wurf eines fairen Würfels, deren Summe $(X+Y)$ und Differenz $(X-Y)$. Dann gilt $\operatorname{Cov}(X+Y,X-Y)=\operatorname{Var}(X)-\operatorname{Var}(Y)=0$. Unabhängigkeit gilt dagegen nicht.

6 Asymptotische Gesetze der Stochastik

6.1 TSCHEBYSCHEV-Ungleichung und schwaches Gesetz der großen Zahlen

Satz 6.1 (MARKOV-Ungleichung)**.** *Ist* $X : \Omega \to \mathbb{R}$ *eine Zufallsvariable und* $p, c > 0$*, dann gilt*

$$\mathbb{P}(|X| \geq c) \leq \frac{1}{c^p} \cdot \mathbb{E}\big[|X|^p \cdot 1_{\{|X| \geq c\}}\big] \leq \frac{1}{c^p} \cdot \mathbb{E}[|X|^p] .$$

Beweis. Setzen wir

$$Y = c^p \cdot 1_{\{|X| \geq c\}}, \quad Z = |X|^p \cdot 1_{\{|X| \geq c\}},$$

dann gilt für $\omega \in \Omega$

$$\begin{aligned} |X(\omega)| < c &\Rightarrow Y(\omega) = 0 = Z(\omega), \\ |X(\omega)| \geq c &\Rightarrow Y(\omega) = c^p \leq |X(\omega)|^p = Z(\omega), \end{aligned}$$

und somit $Y \leq Z$. Es folgt

$$c^p \cdot \mathbb{P}(|X| \geq c) = \mathbb{E}[Y] \leq \mathbb{E}[Z] = \mathbb{E}[|X|^p \cdot 1_{\{|X| \geq c\}}] ,$$

also die erste Ungleichung. Die zweite ergibt sich nun direkt aus $Z \leq |X|^p$. ∎

Satz 6.2 (TSCHEBYSCHEV-Ungleichung)**.** *Ist* $X : \Omega \to \mathbb{R}$ *eine Zufallsvariable mit* $\mathbb{E}[X^2] < \infty$ *und* $\varepsilon > 0$*, so ist*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \operatorname{Var}(X) .$$

Beweis. Setzen wir in die MARKOV-Ungleichung für die Zufallsvariable $Z = |X - \mathbb{E}[X]|$ die Werte $c = \varepsilon$ und $p = 2$ ein, so erhalten wir

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) = \mathbb{P}(|Z| \geq c) \leq \frac{1}{c^2} \cdot \mathbb{E}[Z^2] = \frac{1}{\varepsilon^2} \operatorname{Var}(X) .$$ ∎

Bemerkung. Setzen wir in der TSCHEBYSCHEV-Ungleichung $\varepsilon = \sigma(X) \cdot t$, so erhalten wir

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \sigma(X) \cdot t) \leq \frac{1}{t^2} .$$

Definition 6.3. *Eine Folge* $(X_n)_{n \in \mathbb{N}}$ *reellwertiger Zufallsvariablen konvergiert* $\mathbb{P}$***-stochastisch*** *gegen eine Zufallsvariable* X*, falls für alle* $\varepsilon > 0$ *gilt:*

$$\lim_{n \to \infty} \mathbb{P}\big(|X_n - X| \geq \varepsilon\big) = 0.$$

Die Definition lässt sich analog auch auf Zufallsvariablen mit Werten in metrischen Räumen und Abstandsfunktion $|-|$ erweitern.

Satz 6.4 (**Schwaches Gesetz der großen Zahlen (SGGZ)**). *Sind $X_1, \dots, X_n$ unabhängig, identisch verteilte, reellwertige Zufallsvariablen mit $\mathbb{E}[X_1^2] < \infty$, so gilt:*

$$\forall \varepsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}[X_1] \right| \geq \varepsilon \right) = 0.$$

Mit Definition 6.3 ausgedrückt: Der Mittelwert $n^{-1}\sum_{i=1}^{n} X_i$ von u.i.v. reellwertigen Zufallsvariablen mit endlicher Varianz konvergiert stochastisch gegen den Erwartungswert.

Beweis. Wir setzen $S_n = \sum_{i=1}^{n} X_i$. Da die X_i identisch verteilt sind, ist

$$\mathbb{E}[S_n] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n\mathbb{E}[X_1]\,.$$

und daher

$$\mathbb{E}\left[\frac{S_n}{n}\right] = \mathbb{E}[X_1], \quad \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{\operatorname{Var}(X_1)}{n}\,.$$

Es folgt mit der TSCHEBYSCHEV-Ungleichung:

$$\mathbb{P}\left(\left| \frac{S_n}{n} - \mathbb{E}[X_1] \right| \geq \varepsilon \right) \leq \frac{1}{\varepsilon^2} \operatorname{Var}\left(\frac{S_n}{n}\right) = \frac{1}{\varepsilon^2} \cdot \frac{\operatorname{Var}(X_1)}{n}\,,$$

also

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \frac{S_n}{n} - \mathbb{E}[X_1] \right| \geq \varepsilon \right) \leq \frac{\operatorname{Var}(X_1)}{\varepsilon^2} \lim_{n\to\infty} \frac{1}{n} = 0\,.$$

■

Tatsächlich benötigen wir für das SGGZ mit der TSCHEBYSCHEV-Ungleichung nur, dass die quadratintegrierbaren Zufallsvariablen mit identischem Erwartungswert unkorreliert sind.

6.2 Der Satz von DE MOIVRE-LAPLACE

Für $X_1, X_2, \dots$ unabhängig identisch verteilt und $S_n = X_1 + \cdots + X_n$ lässt sich unter den Voraussetzungen von Satz 6.4 ein stärkeres Resultat herleiten, nämlich eine asymptotische Verteilung der geeignet reskalierten Summenfolge. Zunächst betrachten wir BERNOULLI-verteilte X_i.

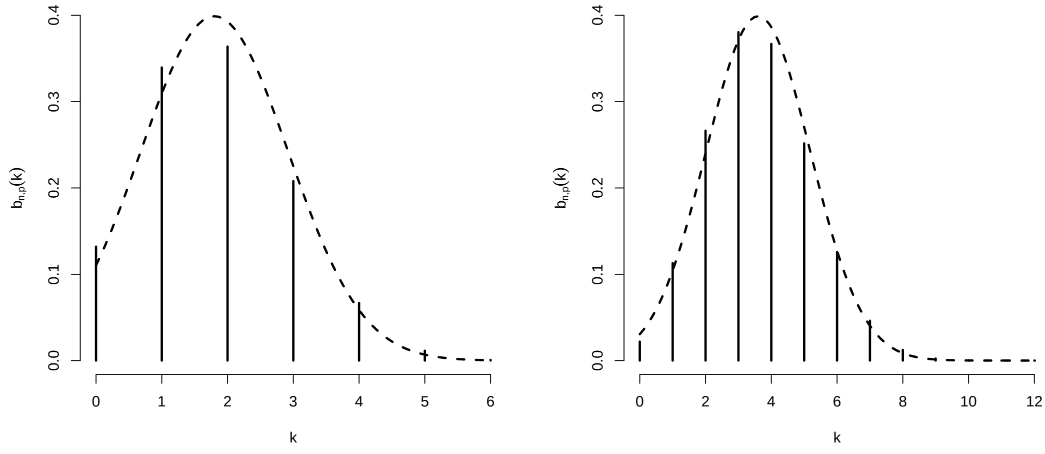
Satz 6.5 (Lokaler Grenzwertsatz für die Binomialverteilung). *Seien $X_1, X_2, \dots$ unabhängig identisch BERNOULLI-verteilt, so gilt*

$$\lim_{n\to\infty} \sup_{k\in\mathbb{Z}} \left| \sqrt{np(1-p)}\,\mathbb{P}(S_n = k) - \varphi\left(\frac{k-np}{\sqrt{np(1-p)}}\right) \right| = 0, \tag{6.1}$$

wobei $\varphi(t) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}}$. Hier ist $S_n \sim \operatorname{Bin}(n,p)$ mit $p \in (0,1)$, also

$$\mathbb{P}(S_n = k) = b_{n,p}(k) = \begin{cases} \binom{n}{k} p^k (1-p)^{n-k}, & \textit{falls } k \in \{0, \dots, n\}, \\ 0, & \textit{sonst.} \end{cases}$$

Wir werden den lokalen Grenzwertsatz im folgenden Abschnitt nur im Falle der BERNOULLI-Verteilung beweisen. Es gilt ein analoger Satz auch allgemeiner für Zufallsvariablen mit Werten in $\mathbb{Z}$ unter sehr



(a) $n = 6$ (b) $n = 12$

Abbildung 6.1: Skalierte Wahrscheinlichkeiten $\sqrt{n}\sigma b_{n,p}(k)$ der Binomialverteilung für $p = 0.3$ und unterschiedliche n, jeweils zusammen mit der Approximation $x \mapsto \varphi((x-np)/\sqrt{np(1-p)})$ (gestrichelt).

allgemeinen Voraussetzungen, siehe Davis and McDonald (1995). Wir schreiben für die Standardabweichung nachfolgend $\sigma = \sqrt{p(1-p)}$. φ ist die Dichte der Standard-Normalverteilung aus Beispiel 3.49.

Bemerkung. Die $\sqrt{n}$-Skalierung in (6.1) kann dadurch motiviert werden, dass sie eine Geschwindigkeit für das Gesetz der großen Zahlen angibt, siehe Satz 6.4. Setzen wir nämlich $\varepsilon = t/\sqrt{n}$ für $t \in (0,\infty)$ in die TSCHEBYSCHEV-Ungleichung ein, so erhalten wir Schranken welche nicht von n abhängen. Für $\varepsilon \ll \sqrt{n}$ erhalten wir dagegen stochastische Konvergenz von $\mathbb{P}(|S_n/n - \mathbb{E}[X_1]| \geq \varepsilon)$ gegen null. Dies liegt daran, dass $\mathbb{V}\mathrm{ar}(S_n/\sqrt{n})$ unabhängig von n wird, während $\mathbb{V}\mathrm{ar}(n^\beta S_n)$ gegen unendlich geht für $\beta > -1/2$ und gegen null für $\beta < -1/2$.

Lemma 6.6. *Sind $l_n, m_n \in \mathbb{R}$ mit $l_n \leq m_n$ mit*

$$\lim_{n\to\infty} \frac{\left(\frac{l_n - np}{\sqrt{n}\sigma}\right)^3}{\sqrt{n}} = \lim_{n\to\infty} \frac{\left(\frac{m_n - np}{\sqrt{n}\sigma}\right)^3}{\sqrt{n}} = 0,$$

so folgt mit $\sigma = \sqrt{p(1-p)}$

$$\lim_{n\to\infty} \sup_{l_n \leq k \leq m_n} \left| \frac{\sqrt{n}\sigma b_{n,p}(k)}{\varphi\left(\frac{k-np}{\sqrt{n}\sigma}\right)} - 1 \right| = 0.$$

Notation. Wir definieren

$$a_n \sim b_n :\Leftrightarrow \lim_{n\to\infty} \frac{a_n}{b_n} = 1.$$

Im Beweis verwenden wir die STIRLINGsche Formel (vgl. z.Bsp. §20, Satz 6 in O. Forster *Analysis 1*)

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{\mathrm{e}}\right)^n$$

Beweis. Ist $l_n \le k_n \le m_n$, so gilt

$$\lim_{n\to\infty} \frac{\left(\frac{k_n-np}{\sqrt{n}\sigma}\right)^3}{\sqrt{n}} = \lim_{n\to\infty} \frac{1}{\sqrt{n}}\left(\sqrt{n}\frac{\frac{k_n}{n}-p}{\sigma}\right)^3 = \lim_{n\to\infty} \frac{n}{\sigma^3}\left(\frac{k_n}{n}-p\right)^3 = 0,$$

also insbesondere $\lim_{n\to\infty} \frac{k_n}{n} = p$, $\lim_{n\to\infty} k_n = \infty$ und $\lim_{n\to\infty}(n-k_n) = \infty$ (wegen $0 < p < 1$). Mit der STIRLING-Formel folgt

$$\begin{aligned} b_{n,p}(k_n) &= \frac{n!}{k_n!(n-k_n)!} p^{k_n}(1-p)^{n-k_n} \\ &\sim \frac{\sqrt{2\pi n}(\frac{n}{\mathrm{e}})^n}{\sqrt{2\pi k_n}(\frac{k_n}{\mathrm{e}})^{k_n}\sqrt{2\pi(n-k_n)}(\frac{n-k_n}{\mathrm{e}})^{n-k_n}} p^{k_n}(1-p)^{n-k_n} \\ &= \frac{1}{\sqrt{2\pi}}\sqrt{\frac{n}{k_n(n-k_n)}}\left(\frac{np}{k_n}\right)^{k_n}\left(\frac{(1-p)n}{n-k_n}\right)^{n-k_n}. \end{aligned}$$

Wegen $k_n \sim np$ und $(n-k_n) \sim n(1-p)$ ist

$$\sqrt{\frac{n}{k_n(n-k_n)}} = \frac{1}{\sqrt{n}} \frac{1}{\sqrt{\frac{k_n}{n}}\sqrt{\frac{n-k_n}{n}}} \sim \frac{1}{\sqrt{n}\sigma}.$$

Setzen wir nun

$$\chi(n,k_n) = \left(\frac{np}{k_n}\right)^{k_n}\left(\frac{nq}{n-k_n}\right)^{n-k_n} = \left(\frac{p}{t}\right)^{nt}\left(\frac{q}{1-t}\right)^{n(1-t)}, \quad q = 1-p, \quad t = t_n = \frac{k_n}{n},$$

so erhalten wir

$$\log\chi(n,k_n) = n\Big(t\log\frac{p}{t} + (1-t)\log\frac{q}{1-t}\Big) = -n\Big(t\log\frac{t}{p} + (1-t)\log\frac{1-t}{q}\Big) =: -ng(t).$$

Die Ableitungen und spezielle Werte sind

$$\begin{aligned} g'(t) &= \log\frac{t}{p} - \log\frac{1-t}{q}, \quad g''(t) = \frac{1}{t} + \frac{1}{1-t}, \\ g(p) &= g'(p) = 0, \quad g''(p) = \frac{1}{p} + \frac{1}{q} = \frac{1}{pq} = \frac{1}{\sigma^2}. \end{aligned}$$

Somit ist

$$g(t) = \frac{(t-p)^2}{2pq} + \mathcal{O}\big((t-p)^3\big) \text{ für } t \to p,$$

also per Definition

$$\exists c, \varepsilon > 0\ \forall p-\varepsilon < x < p+\varepsilon : \left|g(t) - \frac{(t-p)^2}{2pq}\right| \le c|t-p|^3.$$

Es ist

$$\lim_{n\to\infty} n(t-p)^3 = \lim_{n\to\infty} n\left(\frac{k_n}{n}-p\right)^3 = \sigma^3 \lim_{n\to\infty} \frac{\left(\frac{k_n-np}{\sqrt{n}\sigma}\right)^3}{\sqrt{n}} = 0,$$

also

$$\lim_{n\to\infty}\Big(-\log\chi(n,k_n) - n\frac{(t-p)^2}{2pq}\Big) = 0.$$

Es folgt nun

$$1 = \lim_{n\to\infty} \exp\Big(-\log \chi(n,k_n) - n\frac{(t-p)^2}{2pq}\Big) = \lim_{n\to\infty} \frac{\exp\Big(-n\frac{(t-p)^2}{2pq}\Big)}{\chi(n,k_n)},$$

also

$$\chi(n,k_n) \sim \exp\Big(-n\frac{(t-p)^2}{2pq}\Big) = \exp\Big(-\frac{\Big(\frac{k_n-np}{\sqrt{n}\sigma}\Big)^2}{2}\Big)$$

und zusammengefasst

$$b_{n,p}(k_n) \sim \frac{1}{\sqrt{2\pi n}\sigma}\exp\Big(-\frac{1}{2}\Big(\frac{k_n-np}{\sqrt{n}\sigma}\Big)^2\Big) = \frac{1}{\sqrt{n}\sigma}\varphi\Big(\frac{k_n-np}{\sqrt{n}\sigma}\Big) \Leftrightarrow \lim_{n\to\infty}\frac{\sqrt{n}\sigma b_{n,p}(k_n)}{\varphi\Big(\frac{k_n-np}{\sqrt{n}\sigma}\Big)} = 1.$$

Da das gleichmäßig für $l_n \le k_n \le m_n$ gilt, folgt die Behauptung. ■

Mit dem Lemma folgern wir nachfolgend den lokalen Grenzwertsatz für die Binomialverteilung:

$$\lim_{n\to\infty}\sup_{k\in\mathbb{Z}}\Big|\sqrt{np(1-p)}b_{n,p}(k) - \varphi\Big(\frac{k-np}{\sqrt{np(1-p)}}\Big)\Big| = 0.$$

Beweis. Seien $l_n, m_n \in \mathbb{Z}$, sodass

$$\lim_{n\to\infty}\frac{\Big(\frac{l_n-np}{\sqrt{n}\sigma}\Big)^3}{\sqrt{n}} = \lim_{n\to\infty}\frac{\Big(\frac{m_n-np}{\sqrt{n}\sigma}\Big)^3}{\sqrt{n}} = 0, \quad \lim_{n\to\infty}\Big(\frac{l_n-np}{\sqrt{n}\sigma}\Big) = -\infty, \quad \lim_{n\to\infty}\Big(\frac{m_n-np}{\sqrt{n}\sigma}\Big)^3 = \infty$$

und der Modus (Maximalstelle) von $b_{n,p}(\cdot)$ zwischen l_n und m_n liegt (siehe dazu die folgende Bemerkung). Dann gilt nach dem Lemma, da φ beschränkt ist, dass

$$\lim_{n\to\infty}\sup_{l_n\le k\le m_n}\Big|\sqrt{np(1-p)}b_{n,p}(k) - \varphi\Big(\frac{k-np}{\sqrt{n}\sigma}\Big)\Big| = 0, \quad \lim_{n\to\infty}\frac{\sqrt{np(1-p)}b_{n,p}(l_n)}{\varphi\Big(\frac{l_n-np}{\sqrt{n}\sigma}\Big)} = 1.$$

Da aus $\lim_{n\to\infty}\Big(\frac{l_n-np}{\sqrt{n}\sigma}\Big) = -\infty$ schon $\lim_{n\to\infty}\varphi\Big(\frac{l_n-np}{\sqrt{n}\sigma}\Big) = 0$ folgt, ist nun

$$\lim_{n\to\infty}\sqrt{np(1-p)}b_{n,p}(l_n) = 0.$$

Da für alle $k \le l_n$ auch $b_{n,p}(k) \le b_{n,p}(l_n)$ gilt, ist dann auch

$$\lim_{n\to\infty}\sup_{k\le l_n}(\sqrt{np(1-p)}b_{n,p}(x)) = \lim_{n\to\infty}\sup_{x\le\frac{l_n-np}{\sqrt{n}\sigma}}\varphi(x) = 0.$$

Mit analoger Argumentation gilt das auch für $k \ge m_n$ und $x \ge \frac{m_n-np}{\sqrt{n}\sigma}$. Es folgt die Behauptung. ■

Bemerkung. 1. Ist

$$\forall n \in \mathbb{N}: \Big|\frac{k_n-np}{\sqrt{n}\sigma}\Big| \le c \in \mathbb{R},$$

so folgt

$$np - c\sqrt{n}\sigma \le k_n \le np + c\sqrt{n}\sigma.$$

2. Ist dagegen

$$\lim_{n\to\infty}\frac{\Big(\frac{k_n-np}{\sqrt{n}\sigma}\Big)^3}{\sqrt{n}} = 0,$$

also zum Beispiel

$$\frac{\left|\frac{k_n-np}{\sqrt{n}\sigma}\right|^3}{\sqrt{n}} = \frac{1}{\log n} \Rightarrow |k_n-np|^3 \leq \frac{n^2\sigma^3}{\log n}$$

bzw.

$$np - \frac{n^{\frac{2}{3}}\sigma}{(\log n)^{\frac{1}{3}}} \leq k_n \leq np + \frac{n^{\frac{2}{3}}\sigma}{(\log n)^{\frac{1}{3}}}.$$

Wir setzen also zum Beispiel

$$m_n = \lfloor np + n^{\frac{7}{12}} \rfloor, \quad l_n = \lfloor np - n^{\frac{7}{12}} \rfloor.$$

3. Der (von n abhängende) Modus von $b_{n,p}(\cdot)$ lässt sich bestimmen als maximale Lösung der Ungleichung
$$\frac{b_{n,p}(k)}{b_{n,p}(k+1)} = \frac{\binom{n}{k}p^k(1-p)^{n-k}}{\binom{n}{k+1}p^{k+1}(1-p)^{n-k-1}} = \frac{k+1}{n-k}\frac{q}{p} \leq 1.$$
Diese ist erfüllt genau dann, wenn mit $c = qp^{-1}$ gilt
$$k+1 \leq \frac{n-k}{c} \Leftrightarrow k \leq \frac{\frac{n}{c}-1}{1+\frac{1}{c}} = np-q,$$
also solange $k \leq \lfloor np-q \rfloor$. Dieser Wert ist also der Modus.

Satz 6.7 (Satz von de Moivre-Laplace)**.** *Es seien* $X_1, X_2, \ldots$ *unabhängig* Bernoulli-p-*verteilt mit* $p \in (0,1)$ *und somit* $S_n = X_1 + \cdots + X_n \sim \mathrm{Bin}(n,p)$. *Bezeichne* $S_n^\star = \frac{S_n-np}{\sqrt{np(1-p)}}$. *Dann gilt für alle* $a<b$: $\lim_{n\to\infty} \mathbb{P}(a \leq S_n^\star \leq b) = \Phi(b) - \Phi(a) = \int_a^b \varphi(x)\,\mathrm{d}x$. ▪

Beweis. Zunächst ist

$$a \leq \frac{S_n-np}{\sqrt{np(1-p)}} \leq b \;\Leftrightarrow\; a\sqrt{np(1-p)}+np \leq S_n \leq b\sqrt{np(1-p)}+np.$$

Mit $\alpha_n = \lceil a\sqrt{np(1-p)}+np \rceil$, sowie $\beta_n = \lfloor b\sqrt{np(1-p)}+np \rfloor$, gilt

$$\left|\frac{\alpha_n-np}{\sqrt{np(1-p)}} - a\right| \leq \frac{1}{\sqrt{np(1-p)}},$$
$$\left|\frac{\beta_n-np}{\sqrt{np(1-p)}} - b\right| \leq \frac{1}{\sqrt{np(1-p)}},$$

und Lemma 6.6 ist anwendbar. Es folgt, dass für eine Folge (ε_n) mit $\varepsilon_n \to 0$, für $\alpha_n \leq k_n \leq \beta_n$ gilt:

$$1-\varepsilon_n < \frac{\binom{n}{k_n}p^{k_n}(1-p)^{n-k_n}\sqrt{np(1-p)}}{\varphi\left(\frac{k_n-np}{\sqrt{np(1-p)}}\right)} < 1+\varepsilon_n.$$

Damit folgt für $R_n = \sum_{k=\alpha_n}^{\beta_n} \varphi\left(\frac{k_n-np}{\sqrt{np(1-p)}}\right)\frac{1}{\sqrt{np(1-p)}}$:

$$(1-\varepsilon_n)R_n \leq \mathbb{P}\big(a \leq S_n^\star \leq b\big) \leq (1+\varepsilon_n)R_n.$$

R_n approximiert als Riemann-Summe das Integral $\int_{u_n}^{o_n} \varphi(x)\,\mathrm{d}x$ mit den Grenzen $u_n = \frac{\alpha_n-1/2-np}{\sqrt{np(1-p)}}$ und

$o_n = \frac{\beta_n + 1/2 - np}{\sqrt{np(1-p)}}$. Mit

$$\int_{u_n}^{o_n} \varphi(x)\,\mathrm{d}x = \Phi(o_n) - \Phi(u_n)$$

und da $o_n \to b$ und $u_n \to a$, folgt das Resultat. ∎

Bemerkung. S_n nimmt Werte aus $\{0,\dots,n\}$ an, für konstantes l ist

$$\lim_{n\to\infty} \mathbb{P}(S_n \le l) = \lim_{n\to\infty} \mathbb{P}\Big(S_n^\star \le \frac{l-np}{\sqrt{np(1-p)}}\Big) = 0.$$

Ersetzt man hier die Konstante durch ein von n abhängiges l_n mit $\lim_{n\to\infty} l_n = \infty$, so dass $x_n = \frac{l_n - np}{\sqrt{np(1-p)}} \to x \in [-\infty,\infty]$ konvergiert, erhält man

$$\lim_{n\to\infty} \mathbb{P}(S_n \le l_n) = \lim_{n\to\infty} \mathbb{P}\Big(S_n^\star \le \frac{l_n-np}{\sqrt{np(1-p)}}\Big) = \Phi(x).$$

Bemerkung (Stetigkeitskorrektur). Sei $\sigma = \sqrt{np(1-p)}$. Nach dem Satz können wir approximieren

$$\mathbb{P}(S_n \le l) = \mathbb{P}(S_n^\star \le (l-np)\cdot\sigma^{-1}) \approx \Phi((l-np)\cdot\sigma^{-1}) = \Phi\Big(\frac{l-np}{\sqrt{np(1-p)}}\Big).$$

Für endliche n ist die Approximation aus dem Beweis

$$\mathbb{P}(S_n \le l) = \mathbb{P}(S_n^\star \le (l-np)\cdot\sigma^{-1}) \approx \Phi((l-np+\tfrac{1}{2})\cdot\sigma^{-1}) = \Phi\Big(\frac{l-np+\frac{1}{2}}{\sqrt{np(1-p)}}\Big)$$

aber besser. Insbesondere ist für $k,l \in \{0,\dots,n\}$ und $k \le l$, für ausreichend große n, eine gute Approximation gegeben durch

$$\mathbb{P}(k \le S_n \le l) = \mathbb{P}(S_n \le l) - \mathbb{P}(S_n \le k-1) \approx \Phi\Big(\frac{l-np+\frac{1}{2}}{\sqrt{np(1-p)}}\Big) - \Phi\Big(\frac{k-np-\frac{1}{2}}{\sqrt{np(1-p)}}\Big).$$

Abbildungen 6.2 und 6.3 visualisieren die Approximation von DE MOIVRE-LAPLACE.

Beispiel 6.8.

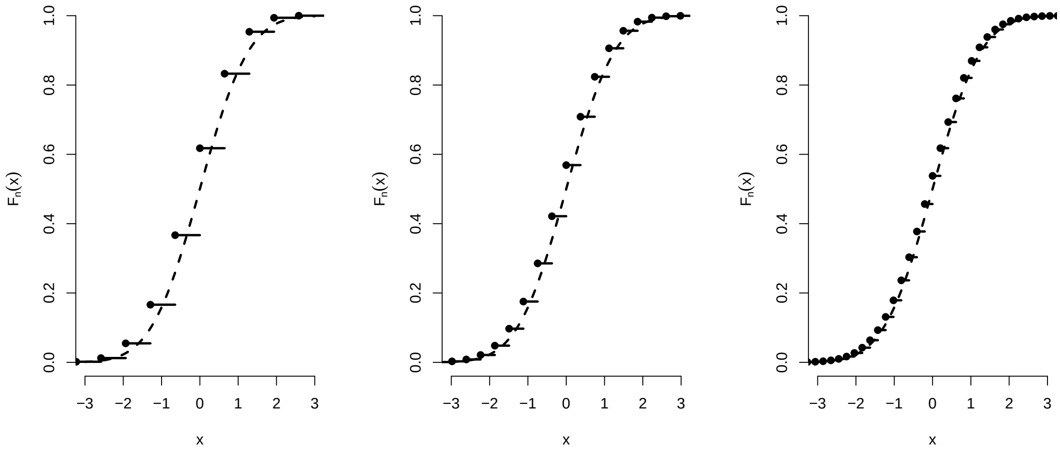
1. Mit $n = 20$, $p = 0{,}4$ und $k = 6$, $l = 10$, berechnen wir

$$\begin{aligned}\mathbb{P}(6 \le S_{20} \le 10) &\approx 2\Phi\Big(\frac{10-8}{\sqrt{20\cdot 0{,}4\cdot 0{,}6}}\Big) - 1 \approx 0{,}6387,\\ &\approx 2\Phi\Big(\frac{10-8+\frac{1}{2}}{\sqrt{20\cdot 0{,}4\cdot 0{,}6}}\Big) - 1 \approx 0{,}7461,\\ \text{exakt } &\approx 0{,}7469.\end{aligned}$$

2. Mit $n = 100$, $p = 0{,}4$ und $k = 35$, $l = 45$ berechnen wir

$$\begin{aligned}\mathbb{P}(35 \le S_{100} \le 45) &\approx 2\Phi\Big(\frac{45-40}{\sqrt{100\cdot 0{,}4\cdot 0{,}6}}\Big) - 1 \approx 0{,}6926,\\ &\approx 2\Phi\Big(\frac{45-40+\frac{1}{2}}{\sqrt{100\cdot 0{,}4\cdot 0{,}6}}\Big) - 1 \approx 0{,}7384,\\ \text{exakt } &\approx 0{,}7486.\end{aligned}$$

Betrachten wir eine Folge $X_1, X_2, \dots$ von unabhängigen, identisch verteilten, reellwertigen Zufallsvariablen mit $\mathbb{E}[X_1^2] < \infty$, $\mu = \mathbb{E}[X_1]$ und $\sigma^2 = \mathrm{Var}(X_1) > 0$. Sei $S_n = X_1 + \ldots + X_n$, dann ist $\mathbb{E}[S_n] = n\cdot\mu$



(a) $n = 10$ (b) $n = 30$ (c) $n = 100$

Abbildung 6.2: Verteilungsfunktion F_n^* von $S_n^\star$ für $p = 0.6$ und unterschiedliche n, jeweils zusammen mit der Verteilungsfunktion Φ der Standard-Normalverteilung (gestrichelt).

und $\operatorname{Var}(S_n) = n \cdot \sigma^2$. Wir betrachten das asymptotische Verhalten für $n \to \infty$. Dazu erweitern wir das schwache Gesetz der großen Zahlen

$$\forall \varepsilon > 0 : \lim_{n\to\infty} \mathbb{P}\left(\left| \frac{1}{n} S_n - \mu \right| \geq \varepsilon \right) = 0$$

zum *Zentralen Grenzwertsatz*, der die Verteilung von S_n/n um μ herum beschreibt. Wir standardisieren S_n linear zu $S_n^\star$ mit $\mathbb{E}[S_n^\star] = 0$ und $\operatorname{Var}(S_n^\star) = 1$, so dass

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

Die Funktion $\Phi(x)$ bezeichnet die Verteilungsfunktion der Standard-Normalverteilung.

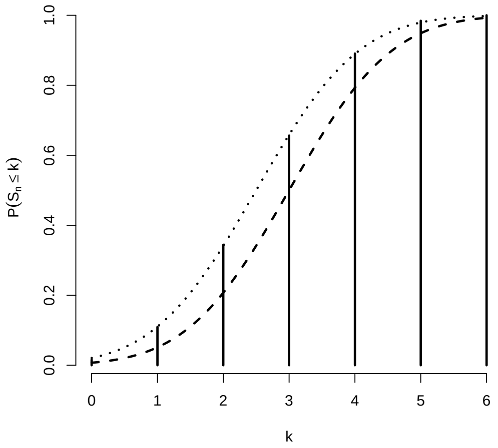
Definition 6.9. *Eine Folge reellwertiger Zufallsvariablen $(X_n)_{n\in\mathbb{N}}$ **konvergiert** gegen eine Zufallsvariable X, $X_n \xrightarrow{d} X$, falls die Folge der Verteilungsfunktionen F_n von X_n punktweise gegen die Verteilungsfunktion F von X konvergiert, in allen Verteilungen Stetigkeitsstellen von F, also falls für alle $x \in \mathbb{R}$ in denen F stetig ist, gilt*

$$F_n(x) \to F(x), \quad n \to \infty.$$

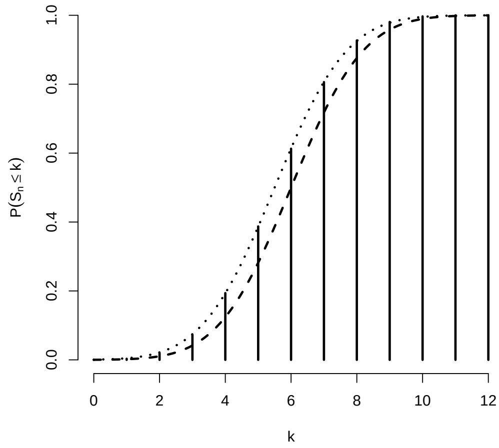
Da es bei der Konvergenz in Verteilung nur auf die Verteilungen ankommt, können wir statt einer Grenzzufallsvariable auch eine Verteilung schreiben und dabei unsere eingeführten Bezeichnungen verwenden, zum Beispiel $\mathcal{N}(0,1)$ für die Standard-Normalverteilung.

Satz 6.10 (Zentraler Grenzwertsatz). *Gelten die oben genannten Voraussetzungen, so gilt*

$$\forall\, b \in \mathbb{R} : \quad \lim_{n\to\infty} \mathbb{P}(S_n^\star \leq b) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} \mathrm{e}^{-\frac{x^2}{2}}\, \mathrm{d}x = \Phi(b)\,.$$



(a) $n = 6$



(b) $n = 12$

Abbildung 6.3: Kumulierte Wahrscheinlichkeiten $\mathbb{P}(S_n \leq k)$ der Binomialverteilung $\mathrm{Bin}(n, 0.5)$ für unterschiedliche n, jeweils zusammen mit der Approximation durch die Normalverteilung ohne (gestrichelt) und mit (gepunktet) Stetigkeitskorrektur.

Mit Definition 6.9 formuliert gilt also

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{d} \mathcal{N}(0,1)\,.$$

■

Bemerkung. Für den zentralen Grenzwertsatz gibt es verschiedene Beweise, unter anderem

1. den Standardbeweis des allgemeinen zentralen Grenzwertsatzes über sogenannte charakteristische Funktionen (wird in Stochastik II behandelt).

2. ein direkter, elementarer Beweis wird im Buch von Krengel (2002), §12.3, entwickelt.

Bemerkung. 1. Die numerische Approximation von $\mathrm{Bin}(n, p)$ durch den zentralen Grenzwertsatz von DE MOIVRE-LAPLACE ist heutzutage eher unnötig, da die Werte auch für große n mit Computern exakt bestimmt werden können. Daher ist dieses Verfahren praktisch nicht mehr besonders relevant, liefert aber theoretisch Aufschluss über die Größenordnung der Varianz $1/n$.

2. Zusätzlich zur *Konvergenzrate* $n^{-1/2}$ ist das theoretische Resultat der Normalapproximation grundlegend und sehr relevant, vor allem auch die allgemeine Form des Zentralen Grenzwertsatzes, da sie unabhängig von der speziellen Verteilung der X_i ist.

3. Da Φ stetig ist folgt über

$$\begin{aligned}\liminf_{n\to\infty} \mathbb{P}(S_n^\star < b) &\geq \limsup_{\varepsilon\searrow 0} \lim_{n\to\infty} \mathbb{P}(S_n^\star \leq b-\varepsilon) = \limsup_{\varepsilon\searrow 0} \Phi(b-\varepsilon)\\ &= \Phi(b) = \lim_{n\to\infty} \mathbb{P}(S_n^\star \leq b) \geq \limsup_{n\to\infty} \mathbb{P}(S_n^\star < b)\end{aligned}$$

weiterhin

$$\lim_{n\to\infty} \mathbb{P}(S_n^\star < b) = \Phi(b), \quad \lim_{n\to\infty} \mathbb{P}(a \leq S_n^\star \leq b) = \Phi(b) - \Phi(a).$$

4. Der allgemeine Zentrale Grenzwertsatz 6.10 enthält als Spezialfall auch wieder den Satz von DE MOIVRE-LAPLACE 6.7.

6.3 Starkes Gesetz der großen Zahlen

6.3.1 Null-Eins-Gesetze

Null-Eins-Gesetze charakterisieren Klassen von Ereignissen, die stets entweder Wahrscheinlichkeit 0 oder 1 haben. Wir diskutieren hier das 0-1-Gesetz für terminale Ereignisse von Kolmogorov, sowie die beiden Lemmata von Borel-Cantelli. Das erste Borel-Cantelli Lemma ist grundlegend, um fast sichere Konvergenz von Folgen von Zufallsvariablen zu zeigen.

Seien $X_1, X_2, \ldots : \Omega \to \mathbb{R}$ Zufallsvariablen, und

$$\mathcal{F}'_n = \sigma\big(X_k,\ k \geq n\big)$$

die σ-Algebra der Zukunft ab dem Zeitpunkt n, und

$$\mathcal{T} = \bigcap_{n=1}^{\infty} \mathcal{F}'_n$$

die σ-Algebra der *terminalen Ereignisse* der Folge (X_k). Intuitiv enthält $\mathcal{T}$ alle Ereignisse, deren Eintreten durch Ändern endlicher vieler X_i nicht beeinflusst wird.

Beispiel. Seien $B_n \in \mathcal{B}$, dann

$$\{X_n \in B_n \text{ unendlich oft (u.o.) }\} = \bigcap_{n=1}^{\infty} \bigcup_{k \geq n} \{X_k \in B_k\} \in \mathcal{T}.$$

Sind insbesondere $A_n \in \mathcal{A}$, und $X_n = \mathbf{1}_{A_n}$, dann $\{A_n \text{ u.o.}\} \in \mathcal{T}$.

Satz 6.11 (0-1-Gesetz von Kolmogorov). *Sind $X_1, X_2, \ldots$ unabhängig, so gilt für alle $A \in \mathcal{T}$:*

$$\mathbb{P}(A) = 0 \qquad \textit{oder} \qquad \mathbb{P}(A) = 1.$$

$\mathcal{T}$ heißt in dem Fall auch $\mathbb{P}$-trivial.

Beweis. Wir zeigen, dass $A \in \mathcal{T}$ von sich selbst unabhängig ist. Dann folgt

$$\mathbb{P}(A) = \mathbb{P}(A \cap A) = \big(\mathbb{P}(A)\big)^2, \qquad \text{also } \mathbb{P}(A) \in \{0,1\}.$$

Die σ-Algebren $\sigma(X_1, \ldots, X_n)$ und $\sigma(X_k,\ k > n)$ sind unabhängig. Da $\mathcal{T} \subset \sigma(X_k,\ k > n)$ für alle n, sind

$$\bigcup_{n=1}^{\infty} \sigma(X_1, \ldots, X_n) \qquad \text{und} \qquad \mathcal{T} \qquad \text{unabhängig.}$$

Da $\bigcup_{n=1}^{\infty} \sigma(X_1, \ldots, X_n)$ ein $\cap$-stabiler Erzeuger von $\sigma(X_1, X_2, \ldots)$, folgt mit Satz 5.7, dass

$$\sigma(X_1, X_2, \ldots) \qquad \text{und} \qquad \mathcal{T} \qquad \text{unabhängig.}$$

Da $\mathcal{T} \subset \sigma(X_1, X_2, \ldots)$, folgt die Unabhängigkeit von $A \in \mathcal{T}$ von sich selbst. ∎

Korollar 6.12. Sind (A_n) unabhängig, so ist

$$\mathbb{P}\big(A_n \text{ u.o.}\big) \in \{0,1\}.$$

Wir fragen uns, wann gilt welche der beiden Möglichkeiten?

Satz 6.13 (Erstes Lemma von Borel-Cantelli). *Ist $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, so ist $\mathbb{P}(A_n \text{ u.o.}) = 0$.*

Bemerke, dass hier keine Unabhängigkeit der (A_n) vorausgesetzt wird.

Beweis. Sei $B_n = \bigcup_{k \geq n} A_k$, so ist

$$\{A_n \text{ u.o.}\} = \bigcap_{n=1}^{\infty} B_n,$$

und $B_{n+1} \subseteq B_n$, daher $\mathbb{P}(B_n) \to \mathbb{P}(A_n \text{ u.o.})$. Da

$$\mathbb{P}(B_n) \leq \sum_{k \geq n} \mathbb{P}(A_k) \to 0, \qquad n \to \infty$$

folgt die Behauptung. ∎

Satz 6.14 (Zweites Lemma von Borel-Cantelli). *Sind (A_n) unabhängig und gilt $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$, so ist $\mathbb{P}(A_n \text{ u.o.}) = 1$.*

Beweis. Da $1 - x \leq e^{-x}$, folgt mit der Unabhängigkeit der A_n für $0 < M < N$:

$$\begin{aligned}
\mathbb{P}\Big(\bigcap_{k=M}^{N} A_k^{\complement}\Big) &= \prod_{k=M}^{N} \big(1 - \mathbb{P}(A_k)\big) \\
&\leq \prod_{k=M}^{N} \exp\big(-\mathbb{P}(A_k)\big) = \exp\Big(-\sum_{k=M}^{N} \mathbb{P}(A_k)\Big) \to 0, \quad N \to \infty.
\end{aligned}$$

Daher ist $\mathbb{P}\big(\bigcup_{k=M}^{\infty} A_k\big) = 1$ für alle M, und somit auch (Stetigkeit von oben)

$$\mathbb{P}(A_n \text{ u.o.}) = \mathbb{P}\Big(\bigcap_{M=1}^{\infty} \bigcup_{k=M}^{\infty} A_k\Big) = 1.$$

∎

Als Folgerung beweisen wir zunächst eine Umkehrung des starken Gesetzes der großen Zahlen.

Satz 6.15. *Sind $X_1, X_2, \ldots$ unabhängig identisch verteilt mit $\mathbb{E}\big[|X_i|\big] = \infty$, dann gilt für $S_n = X_1 + \ldots + X_n$*

$$\mathbb{P}\Big(\lim_{n\to\infty} \frac{S_n}{n} \text{ existiert in } \mathbb{R}\Big) = 0.$$

Beweis. Das beschriebene Ereignis ist messbar, denn es gilt nach dem Cauchy-Kriterium

$$\Big\{\lim_{n\to\infty} \frac{S_n}{n} \text{ existiert in } \mathbb{R}\Big\} = \bigcap_{\varepsilon>0, \varepsilon\in\mathbb{Q}} \bigcup_{n=1}^{\infty} \bigcap_{n \leq k < l}^{\infty} \big\{|S_k/k - S_l/l| < \varepsilon\big\}.$$

Wir zeigen zunächst, dass $\mathbb{P}(|X_n| \geq n \;\text{ u.o.}) = 1$. Da nach Voraussetzung $\mathbb{E}\big[|X_i|\big] = \infty$, folgt nach (5.13) und da die X_i identisch verteilt sind

$$\infty = \sum_{n=1}^{\infty} \mathbb{P}(|X_1| \geq n) = \sum_{n=1}^{\infty} \mathbb{P}(|X_n| \geq n).$$

Damit folgt $\mathbb{P}(|X_n| \geq n \;\text{ u.o.}) = 1$ mit dem zweiten Borel-Cantelli-Lemma.
Nun ist

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1}.$$

Sei $A = \{\lim_{n\to\infty} \frac{S_n}{n}$ existiert in $\mathbb{R}\}$ und $B = \{|X_n| \geq n \;\text{ u.o.}\}$. Für $\omega \in A \cap B$ ist $\frac{S_n(\omega)}{n(n+1)} \to 0$, und da auf B gilt $|X_{n+1}|/(n+1) \geq 1$ *u.o.*, folgt auf der Schnittmenge

$$\left|\frac{S_n}{n} - \frac{S_{n+1}}{n+1}\right| \geq 1/2 \qquad u.o.,$$

was $\omega \in A$ widerspricht. Daher ist $A \cap B = \emptyset$, und da $\mathbb{P}(B) = 1$, folgt $\mathbb{P}(A) = 0$, also die Aussage des Satzes. ■

6.3.2 Starkes Gesetz der großen Zahlen und Etemadis Beweis

Satz 6.16 (Starkes Gesetz der großen Zahlen). *Sind $X_1, X_2, \ldots$ unabhängig identisch verteilt mit $\mathbb{E}\big[|X_1|\big] < \infty$, so gilt für $S_n = \sum_{i=1}^n X_i$, dass*

$$\frac{S_n}{n} \to \mathbb{E}[X_1] \qquad \textit{fast sicher}.$$

Bemerkung. 1. Zusammen mit Satz 6.15 folgt, dass für X_i u.i.v. S_n/n genau dann gegen eine Konstante $a \in \mathbb{R}$ fast sicher konvergiert, wenn $\mathbb{E}\big[|X_i|\big] < \infty$, und dann ist $a = \mathbb{E}[X_1]$.

2. Im Beweis wird nur die paarweise Unabhängigkeit der X_i benutzt, der Satz gilt also unter dieser schwächeren Voraussetzung.

3. Der folgende Beweis geht zurück auf Etemadi (1981), das Resultat auf Kolmogorov. Wir beginnen mit einem Lemma.

Lemma 6.17. *Sei $Y_k = X_k \mathbf{1}_{\{|X_k| \leq k\}}$, und $T_n = Y_1 + \ldots + Y_n$, dann gilt*

$$\frac{S_n - T_n}{n} \to 0 \qquad \textit{fast sicher}.$$

Beweis. Da $\mathbb{E}\big[|X_1|\big] < \infty$ und da die X_i identisch verteilt sind, folgt mit (5.13), dass

$$\sum_{k\geq 1} \mathbb{P}(|X_k| \geq k) = \sum_{k\geq 1} \mathbb{P}(|X_1| \geq k) < \infty.$$

Mit dem ersten Lemma von Borel-Cantelli folgt

$$\mathbb{P}(|X_k| > k \quad \text{u.o.}) = 0,$$

also

$$\mathbb{P}(X_k \neq Y_k \quad \text{u.o.}) = 0,$$

und auf $\{X_k \neq Y_k \quad \text{u.o.}\}^\complement$ gilt $(S_n - T_n)/n \to 0$. ■

Beweis von Satz 6.16. Da X_n^+ und X_n^- die Voraussetzungen des Satzes erfüllen, und

$$S_n = X_1^+ + \ldots + X_n^+ - (X_1^- + \ldots + X_n^-)$$

folgt aus den Aussagen für X_n^+ und X_n^- die Behauptung für X_n. Daher kann o.E.d.A. $X_n \geq 0$ angenommen werden. Wir zeigen die Behauptung also o.E.d.A. unter der Annahme $X_n \geq 0$ für die $Y_n = X_n \mathbf{1}_{\{X_n \leq n\}}$, und T_n wie in Lemma 6.17.

Die neue Idee von Etemadi (1981) war es, die fast sichere Konvergenz von $T_{n_k}/n_k \to \mathbb{E}[X_1]$ mit Borel-Cantelli entlang exponentiell wachsender Teilfolgen $n_k = [\alpha^k]$ für $\alpha > 1$, zu zeigen und dann auf die fast sichere Konvergenz der ganzen Folge $T_n/n \to \mathbb{E}[X_1]$ zu schließen. Mit dem Lemma 6.17 überträgt sich diese auf $S_n/n \to \mathbb{E}[X_1]$ mit einem Monotonieargument, da T_n wegen der Nichtnegativität der Summanden monoton wächst. Wir zeigen in Lemma 6.18, dass

$$\sum_{k=1}^{\infty} \mathbb{P}\big(|T_{n_k} - \mathbb{E}[T_{n_k}]| \geq \varepsilon n_k\big) < \infty$$

und damit nach dem ersten Lemma von Borel-Cantelli:

$$\frac{T_{n_k} - \mathbb{E}[T_{n_k}]}{n_k} \to 0 \qquad \text{fast sicher.}$$

Da nach dem Satz von der monotonen Konvergenz $\mathbb{E}[Y_k] \to \mathbb{E}[X_1]$ gilt, folgt

$$\frac{\mathbb{E}[T_{n_k}]}{n_k} = \frac{1}{n_k} \sum_{m=1}^{n_k} \mathbb{E}[Y_m] \to \mathbb{E}[X_1]\,.$$

Somit folgt

$$\frac{T_{n_k}}{n_k} \to \mathbb{E}[X_1] \qquad \text{fast sicher.}$$

Für $m \in \mathbb{N}$ wähle $k = k(m)$ mit $n_{k(m)} \leq m < n_{k(m)+1}$. Wir nutzen nun, dass die $X_i \geq 0$, denn dann folgt $T_{n_{k(m)}} \leq T_m \leq T_{n_{k(m)+1}}$ und damit

$$\frac{T_{n_{k(m)}}}{n_{k(m)+1}} \leq \frac{T_m}{m} \leq \frac{T_{n_{k(m)+1}}}{n_{k(m)}}\,.$$

Mit $m \to \infty$ geht auch $k(m) \to \infty$. Bilde $\liminf$ in der linken Ungleichung und $\limsup$ in der rechten, und nutze $n_{k(m)}/n_{k(m)+1} \to 1/\alpha$, dann folgt

$$\mu/\alpha \leq \liminf_m T_m/m \leq \limsup_m T_m/m \leq \alpha\mu\,.$$

Da $\alpha > 1$ beliebig, folgt

$$\mu \leq \liminf_m T_m/m \leq \limsup_m T_m/m \leq \mu\,,$$

also die Konvergenz von T_m/m, und somit auch die von S_m/m. ∎

Lemma 6.18. *Mit der obigen Notation aus dem Beweis von Satz 6.16 gilt für alle* $\varepsilon > 0$

$$\sum_{k=1}^{\infty} \mathbb{P}\big(|T_{n_k} - \mathbb{E}[T_{n_k}]| \geq \varepsilon n_k\big) < \infty\,.$$

Beweis. Mit der Tschebyschev-Ungleichung folgt, dass

$$\begin{aligned}\sum_{k=1}^{\infty} \mathbb{P}\big(|T_{n_k} - \mathbb{E}[T_{n_k}]| \geq \varepsilon n_k\big) &\leq \sum_{k=1}^{\infty} \frac{\mathbb{V}\mathrm{ar}(T_{n_k})}{\varepsilon^2 n_k^2} \\ &= \frac{1}{\varepsilon^2} \sum_{k=1}^{\infty} \frac{1}{n_k^2} \sum_{m=1}^{n_k} \mathbb{V}\mathrm{ar}(Y_m)\end{aligned}$$

$$= \frac{1}{\varepsilon^2} \sum_{m=1}^{\infty} \mathbb{V}\mathrm{ar}(Y_m) \sum_{k:n_k \geq m} \frac{1}{n_k^2} \,.$$

Sei l minimal mit $n_l \geq m$, dann ist für $l \leq k$: $n_k \geq n_l \alpha^{k-l}/2 \geq m\alpha^{k-l}/2$. Es folgt

$$\sum_{k:n_k \geq m} \frac{1}{n_k^2} \leq \frac{1}{m^2} 4 \sum_{k=1}^{\infty} \alpha^{-k} = \frac{4}{m^2(1-\alpha^{-1})} \,.$$

Es genügt also zu zeigen, dass gilt

$$\sum_{k=1}^{\infty} \frac{\mathbb{V}\mathrm{ar}(Y_k)}{k^2} \leq 4\mathbb{E}\big[|X_1|\big] < \infty \,. \tag{6.2}$$

Da $Y_k \leq k$, ist mit (5.12)

$$\begin{aligned}
\mathbb{V}\mathrm{ar}(Y_k) \leq \mathbb{E}[Y_k^2] &= \int_0^\infty 2y\mathbb{P}(Y_k > y)\,\mathrm{d}y \\
&= \int_0^k 2y\mathbb{P}(Y_k > y)\,\mathrm{d}y \\
&\leq \int_0^k 2y\mathbb{P}(X_1 > y)\,\mathrm{d}y \,,
\end{aligned}$$

und

$$\begin{aligned}
\sum_{k=1}^{\infty} \frac{\mathbb{V}\mathrm{ar}(Y_k)}{k^2} &\leq \sum_{k=1}^{\infty} \frac{1}{k^2} \int_0^\infty \mathbf{1}_{y<k}\, 2y\,\mathbb{P}(|X_1| > y)\,\mathrm{d}y \\
&= \int_0^\infty \Big(2y \sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{1}_{y<k}\Big)\, \mathbb{P}(|X_1| > y)\,\mathrm{d}y \,.
\end{aligned}$$

Die Behauptung folgt dann durch die Ungleichung

$$2y \sum_{k=1}^{\infty} \frac{1}{k^2} \mathbf{1}_{y<k} \leq 4 \qquad \text{für alle } y \geq 0 \,.$$

Um diese zu zeigen, bemerken wir zunächst für $m \geq 2$, dass

$$\sum_{k \geq m} \frac{1}{k^2} \leq \int_{m-1}^{\infty} \frac{1}{x^2}\,\mathrm{d}x = \frac{1}{m-1} \,.$$

Die obige Summe beginnt mit $[y]+1$, wobei $[y]$ die größte ganze Zahl $\leq y$ ist.
Ist nun $y \geq 1$, so beginnt die Summe mit $[y]+1 \geq 2$, und somit

$$2y \sum_{k>y} \frac{1}{k^2} \leq \frac{2y}{[y]} \leq 4 \,,$$

da $y/[y] \leq 2$ für $y \geq 1$. Ist $0 \leq y < 1$, so ist

$$2y \sum_{k>y} \frac{1}{k^2} \leq 2\Big(1 + \sum_{k \geq 2} \frac{1}{k^2}\Big) \leq 4 \,.$$

■

6.3.3 Der Satz von Glivenko-Cantelli

Seien $X_1, X_2, \ldots$ u.i.v. reellwertige Zufallsvariablen mit Verteilungsfunktion F. Wir bezeichnen mit

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{(-\infty,x]}(X_k) = \frac{1}{n}\#\{k:\ X_k \le x\}$$

die empirische Verteilungsfunktion. Es ist mit das grundlegendste Problem der nichtparametrischen Statistik eine unbekannte Verteilung aus u.i.v. Beobachtungen zu schätzen. Dazu wird typischerweise die Verteilungsfunktion F durch die empirische Verteilungsfunktion F_n geschätzt. Da $Z_k = \mathbf{1}_{(-\infty,x]}(X_k)$ unabhängig und Bernoulli-verteilt sind mit Erfolgswahrscheinlichkeit $\mathbb{P}(X_k \le x) = F(x)$, gilt nach dem starken Gesetz der großen Zahlen für alle $x \in \mathbb{R}$

$$F_n(x) \to F(x) \qquad \text{fast sicher.}$$

Der folgende Satz besagt, dass die Konvergenz sogar gleichmäßig gilt. Der Satz wird manchmal auch als Hauptsatz der Statistik bezeichnet.

Satz 6.19 (Satz von Glivenko-Cantelli). *Seien $X_1, X_2, \ldots$ u.i.v. mit Verteilungsfunktion F, so gilt*

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \to 0 \qquad \textit{fast sicher.}$$

Beweis. Zunächst ist wegen der rechtsseitigen Stetigkeit von Verteilungsfunktionen $\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = \sup_{x\in\mathbb{Q}} |F_n(x) - F(x)|$, also insbesondere messbar. Wir benutzen die Quantilsfunktion

$$F^{-1}(t) = \inf\{x:\ F(x) \ge t\}$$

und auch den linksseitigen Limes

$$F(x-) := \lim_{y\to x, y<x} F(y) = \mathbb{P}(X_1 < x).$$

Wegen der Rechtsstetigkeit von F gilt dann

$$F\big(F^{-1}(t)\big) \ge t \qquad \text{und} \qquad F\big(F^{-1}(t)-\big) \le t. \tag{6.3}$$

Weiter beachten wir, dass auch gilt

$$F_n(x-) = \frac{1}{n}\sum_{k=1}^{n} \mathbf{1}_{(-\infty,x)}(X_k) \to \mathbb{P}(X_1 < x) = F(x-) \qquad \text{fast sicher}$$

für jedes $x \in \mathbb{R}$.
Sei nun $N \in \mathbb{N}$ beliebig, und sei $x_{k,N} = F^{-1}(k/N)$, $k = 1, \ldots, N-1$, und $x_{0,N} = -\infty$, $x_{N,N} = \infty$. Für $k = 1, \ldots, N$ ist dann

$$F(x_{k-1,N}) \ge \frac{k-1}{N}, \qquad F(x_{k,N}-) \le \frac{k}{N}.$$

Falls $x_{k,N} > x_{k-1,N}$, ist damit wegen der Monotonie von F,

$$0 \le F(x_{k,N}-) - F(x_{k-1,N}) \le 1/N.$$

Wegen der punktweisen Konvergenz von $F_n(x-)$ und $F_n(x)$ können wir n_0 so groß wählen, dass für $n \ge n_0$:

$$|F_n(x_{k,N}) - F(x_{k,N})| \le 1/N, \qquad |F_n(x_{k,N}-) - F(x_{k,N}-)| \le 1/N, \qquad k = 1, \ldots, N-1.$$

Die Ungleichungen gelten trivialerweise auch für $k = 0$ und $k = N$. Ist nun $x \in \mathbb{R}$ beliebig, etwa $x \in [x_{k-1,N}, x_{k,N})$, so folgt für $n \geq n_0$, dass

$$\begin{aligned} F_n(x) &\leq F_n(x_{k,N}-) \leq F(x_{k,N}-) + 1/N \leq F(x_{k-1,N}) + 2/N \leq F(x) + 2/N, \\ F_n(x) &\geq F_n(x_{k-1,N}) \geq F(x_{k-1,N}) - 1/N \geq F(x_{k,N}-) - 2/N \geq F(x) - 2/N, \end{aligned}$$

also die Behauptung. ■

Literaturverzeichnis

Bauer, H. (2002). *Wahrscheinlichkeitstheorie, 5th edition.* de Gruyter.

Bibinger, M. und Holzmann, H. (2017). *Maß- und Integrationstheorie*. Vorlesungsskript, Universität Marburg.

Billingsley, P. (1994). *Probability and Measure*. Wiley.

Breiman, L. (2007). *Probability*. Siam.

Davis, B. and McDonald, D. (1995). *An elementary proof of the local central limit theorem. J Theor Probab 8, 693–701*. https://doi.org/10.1007/BF02218051

Döring, L. (2021) *Stochastik I*. Skript

Durrett, R. (2010). *Probability. Theory and examples. 4th ed.*, Cambridge: Cambridge University Press.

Elstrodt, J. (1996). *Maß- und Integrationstheorie*, Springer.

Etemadi, N. (1981). *An elementary proof of the strong law of large numbers, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 55, 119-122. doi:10.1007/BF01013465

Georgii, H.-O. (2007). *Stochastik: Einführung in die Wahrscheinlichkeitstheorie und Statistik*. de Gruyter.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications, Volume 1*. Wiley.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications, Volume 2*. Wiley.

Jacod, J. and Protter, P. (2000). *Probability Essentials.* Springer.

Klenke, A. (2008). *Wahrscheinlichkeitstheorie.* Berlin: Springer.

Krengel, U. (2002). *Einführung in die Wahrscheinlichkeitstheorie und Statistik. 6th Edition.* Braunschweig: Vieweg.

Werner, D. (2011). *Funktionalanalysis.* Berlin: Springer.