

Vorlesung Stochastik II

Carsten Hartmann

Version vom 2.10.2014

Inhaltsverzeichnis

Vorwort	2
1 Grundlagen der Maß- und Integrationstheorie	3
1.1 Maßproblem, Mengensysteme, Maßraum	3
1.2 Konstruktion von (Wahrscheinlichkeits-)Maßen	5
1.3 Integration messbarer Funktionen	11
1.4 Konvergenzsätze der Maßtheorie (starke Konvergenz)	16
1.5 Produktmaße	21
2 Summen von unabhängigen Zufallsvariablen	24
2.1 Wahrscheinlichkeitsverteilungen und -dichten	24
2.2 Verteilungskonvergenz (schwache Konvergenz)	26
2.3 Unabhängigkeit	34
2.4 Grenzwertsätze	36
2.5 Mittlere und große Abweichungen	43
3 Abhängige Zufallsvariable: Markovketten	51
3.1 Bedingte Wahrscheinlichkeiten und Erwartungswerte	51
3.2 Markovketten auf abzählbarem Zustandsraum	53
3.3 Stationäre Verteilungen von Markovketten	58
3.4 Asymptotisches Verhalten und Kopplung von Markovketten	65
3.5 Satz von Perron-Frobenius	68
3.6 Markovketten-Monte-Carlo	72
4 Zeitdiskrete Martingale	77
4.1 Bedingte Wahrscheinlichkeiten und Erwartungswerte, Forts.	77
4.2 Spielsysteme und stochastische Integrale	82
4.3 Gleichgradige Integrierbarkeit	85
4.4 Stoppzeiten und Stoppsätze	88
4.5 Optimales Stoppen	91
Literatur	96

Vorwort

Dieses Skript basiert auf der Vorlesung “Stochastik II”, die ich an der FU Berlin zum ersten Mal im Sommersemester 2012 gehalten habe. Inhaltlich knüpft es dort an, wo in der Vorlesung “Elementare Stochastik” die meisten Fragen unbeantwortet bleiben mussten – bei der präzisen maßtheoretischen Formulierung von Wahrscheinlichkeiten. Auch wenn dabei eine gewisse Vertrautheit im Umgang mit Wahrscheinlichkeiten vorausgesetzt wird, so versucht dieses Skript dennoch, ein einigermaßen vollständiges Bild der “echten” Wahrscheinlichkeitstheorie zu vermitteln, indem alle wesentlichen Definitionen und Sätze noch einmal vorgestellt werden.

1 Grundlagen der Maß- und Integrationstheorie

Zunächst stellen wir einige Vorüberlegungen an. Wir betrachten eine beliebige nichtleere Menge Ω , die Menge aller *Elementarereignisse* $\omega \in \Omega$, und ein System \mathcal{A} von Teilmengen $A \subset \Omega$, *Ereignissystem* genannt.

Unter einem Wahrscheinlichkeitsmaß, das jedem Ereignis, sprich: jedem Element von \mathcal{A} , eine Zahl zwischen 0 und 1 zuordnet, verstehen wir eine Mengenfunktion $P: \mathcal{A} \rightarrow \bar{\mathbf{R}} := \mathbf{R} \cup \{-\infty, +\infty\}$ mit den folgenden Eigenschaften:

1. *Normiertheit*: $P(\Omega) = 1$.
2. *Positivität*: $0 \leq P(A) \leq 1$ für alle $A \in \mathcal{A}$.
3. *σ -Additivität*: Für eine Folge $A_1, A_2, A_3, \dots \in \mathcal{A}$ von paarweise disjunkten Elementen gilt

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Zwei zentrale Fragen, die in der Vorlesung ‘Elementare Stochastik’ unbeantwortet geblieben sind, waren zum einen die Frage nach der geeigneten Wahl von \mathcal{A} – bzw. die Frage, warum die Potenzmenge 2^Ω sich im Allgemeinen nicht als Ereignissystem eignet – sowie die Frage, ob die etwas unnatürliche Unterscheidung zwischen abzählbaren und überabzählbaren Mengen von Elementarereignissen wirklich notwendig ist.

1.1 Maßproblem, Mengensysteme, Maßraum

Angenommen wir wollten den Flächeninhalt Großbritanniens auf einer Landkarte bestimmen. Dazu könnten wir beispielsweise die Fläche mit hinreichend kleinen, möglicherweise unendlich vielen Quadraten überdecken und die (leicht zu berechnenden) Flächeninhalte der Quadrate aufsummieren (siehe Abb. 1). Etwas allgemeiner gesagt suchen wir eine σ -additive Mengenfunktion $\mu: \mathcal{B} \rightarrow [0, \infty)$ auf einem Mengensystem $\mathcal{B} \subset 2^\Omega$, die invariant unter Kongruenztransformationen¹ ist; da die Nullfunktion $\mu = 0$ diese Bedingungen erfüllt, fordern wir zusätzlich $\mu([0, 1]^n) = 1$ für das n -dimensionale Einheitsintervall $[0, 1]^n = [0, 1] \times \dots \times [0, 1]$.

Das *Maßproblem* besteht gerade darin, eine solche Mengenfunktion μ auf der Potenzmenge von $\Omega \subset \mathbf{R}^n$ zu finden. Dass das Maßproblem keine Lösung hat, besagt der folgende Satz von Vitali.²

Satz 1.1 (Vitali, 1905) *Das Maßproblem ist für kein $n \in \mathbf{N}$ lösbar.*

Beweis. Wir beschränken uns auf den Fall $n = 1$ und verweisen ansonsten auf die einschlägige Literatur, z.B. [7, Satz III.3.3]. O.B.d.A. betrachten wir das Einheitsintervall $I = [0, 1] \subset \mathbf{R}$, wobei wir die beiden Endpunkte miteinander identifizieren (mit anderen Worten: wir betrachten einen eindimensionalen Torus der Länge 1). Um einen Widerspruchsbeweis zu führen, nehmen wir an, dass ein translationsinvariantes, σ -additives und normiertes Maß auf 2^I existiert. Nach Voraussetzung gilt $\mu([0, 1]) = 1$, und Translationsinvarianz bedeutet

$$\mu(B + x \bmod 1) = \mu(B) \quad \forall B \subset I,$$

wobei wir die Notation $B + x = \{x + y : y \in B\}$ verwenden. Die Idee des Beweises ist es, eine geeignete Partition von I zu finden, aus der sich eine Menge konstruieren lässt, der kein eindeutiges Maß zugewiesen werden kann. Dazu bilden wir eine Äquivalenzklasse von Punkten, die alle rationalen Abstand zueinander haben; sei also A_x die Äquivalenzklasse aller Punkte, die durch die Äquivalenzrelation

$$x \sim y \quad \Leftrightarrow \quad x - y \in [0, 1] \cap \mathbf{Q}$$

bestimmt ist. Für zwei beliebige Punkte $x, y \in I$ gilt dann also entweder $A_x = A_y$ oder $A_x \cap A_y = \emptyset$. Aus der Partition $P := \{A_x : x \in I\}$ der Menge I wählen wir mit Hilfe des Auswahlaxioms je einen Repräsentanten pro Äquivalenzklasse aus, also $x \in A_x, y \in A_y, \dots$. Es sei nun V die Menge all dieser Repräsentanten und

$$V_s = V + x \bmod 1, \quad s \in [0, 1] \cap \mathbf{Q}.$$

¹ Zwei Mengen $A, B \in \mathcal{C} \mathbf{R}^n$ werden *kongruent* genannt, wenn ein Vektor $q \in \mathbf{R}^n$ und eine Drehmatrix $Q \in O(n)$ existieren, so dass $B = Q(A) + q$, wobei $Q(A)$ eine kompakte Schreibweise für $\{Qx : x \in A\}$ ist.

² Giuseppe Vitali (1875–1932), ital. Mathematiker

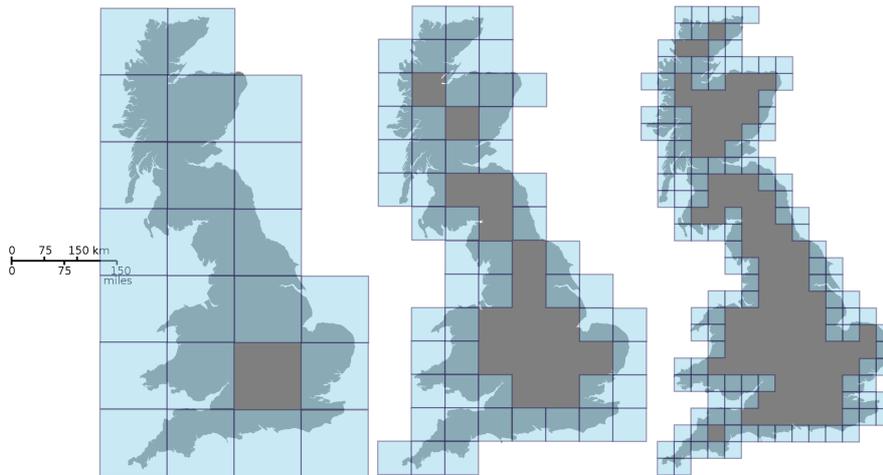


Abb. 1 Disjunkte Überdeckung einer Menge durch Quadrate

Offenbar sind die V_s alle disjunkt, denn es gilt $V_s \cap V_{s'} = \emptyset$ für alle $s \neq s' \pmod 1$. Ferner ist

$$\bigcup_{s \in [0,1] \cap \mathbf{Q}} V_s = I,$$

so dass wir aus den Eigenschaften des Maßes folgern können, dass

$$1 = \mu(I) = \sum_{s \in [0,1] \cap \mathbf{Q}} \mu(V_s) = \sum_{s \in [0,1] \cap \mathbf{Q}} \mu(V).$$

Dabei haben wir im ersten Schritt die Normiertheit, im zweiten die σ -Additivität und im dritten die Translationsinvarianz des Maßes ausgenutzt. Da im letzten Ausdruck aber eine unendliche Summe über alle rationalen Punkte $s \in [0,1] \cap \mathbf{Q}$ steht, kann auf der rechten Seite der Gleichung nur 0 oder unendlich stehen, wodurch die Aussage des Satzes für $n = 1$ bewiesen wäre. \square

Da die Forderungen, die wir an unser Maß gestellt haben, allesamt sinnvoll erscheinen, bleibt uns an dieser Stelle nichts anderes übrig, als den Definitionsbereich von μ einzuschränken, sprich: die Wahl des Systems von Teilmengen $\mathcal{B} \subset 2^\Omega$, denen ein Maß oder ein Volumen zugeordnet werden kann, zu überdenken. Tatsächlich stellt sich heraus, dass die Potenzmenge zu groß ist, und, wie wir sehen werden, ist die σ -Algebra der Borelmengen ein natürlicher Kandidat bei der Lösung des Maßproblems.

Definition 1.2 (Algebra, σ -Algebra) Es sei Ω eine nichtleere Menge und \mathcal{A} ein System von Teilmengen auf Ω . \mathcal{A} heißt *Algebra*, wenn

1. $\emptyset \in \mathcal{A}$,
2. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$,
3. $A \in \mathcal{A} \Rightarrow A^c = \Omega \setminus A \in \mathcal{A}$.

\mathcal{A} heißt *σ -Algebra*, wenn zusätzlich gilt, dass

4. $A_1, A_2, A_3, \dots \in \mathcal{A} \Rightarrow \bigcup_n A_n \in \mathcal{A}$.

Anmerkung 1.3 Wie man leicht mit Hilfe der de Morganschen Regeln beweist, enthalten σ -Algebren neben abzählbaren auch endliche Vereinigungen, Relativkomplemente sowie beliebige Durchschnitte.

Anmerkung 1.4 Eine weitere hilfreiche, leicht zu beweisende Eigenschaft von σ -Algebren ist ihre *Durchschnittstabilität*: Der Schnitt zweier σ -Algebren \mathcal{A} und \mathcal{B} ist wieder eine σ -Algebra (siehe [10, Satz 1.11]).

Ein wichtiger Begriff ist der der *erzeugten σ -Algebra*. Ist beispielsweise $A \in \Omega$ irgendeine Teilmenge, so ist $\{\emptyset, \Omega, A, A^c\}$ die kleinste σ -Algebra, die A enthält; sie wird *die von A erzeugte σ -Algebra* genannt und A ihr *Erzeuger*. Bezogen auf Mengensysteme ergeben sich die nachstehenden Definitionen.

Definition 1.5 (erzeugte σ -Algebra, Erzeuger) Es sei $\mathcal{E} \subset 2^\Omega$ ein Mengensystem und Σ die Menge aller σ -Algebren, die \mathcal{E} enthalten. Die σ -Algebra

$$\sigma(\mathcal{E}) = \bigcap_{\mathcal{F} \in \Sigma} \mathcal{F}$$

ist die kleinste σ -Algebra, die \mathcal{E} enthält und heißt *die von \mathcal{E} erzeugte σ -Algebra*. \mathcal{E} wird *Erzeuger von $\sigma(\mathcal{E})$* genannt.

Definition 1.6 (σ -Algebra der Borelmengen) Sei Ω ein topologischer Raum und \mathcal{O} das System der offenen Teilmengen von Ω . Dann heißt

$$\mathcal{B}(\Omega) = \sigma(\mathcal{O})$$

Borelsche σ -Algebra über Ω ; die Elemente von $\mathcal{B}(\Omega)$ werden *Borelmengen* genannt.

Borelmengen von \mathbf{R} sind praktisch alle Mengen, die sich mit Hilfe elementarer Mengensymbole wie “ \cup ”, “ \cap ”, “ \setminus ”, “ $[\cdot, \cdot)$ ” usw. hinschreiben lassen.

Satz 1.7 (Erzeuger der Borel- σ -Algebra) *Jedes der folgenden Mengensysteme ist Erzeuger von $\mathcal{B}(\mathbf{R}^n)$:*

$$\mathcal{O}^n = \{U \subset \mathbf{R}^n : U \text{ offen}\}$$

$$\mathcal{C}^n = \{A \subset \mathbf{R}^n : A \text{ abgeschlossen}\}$$

$$\mathcal{I}^n = \{(a, b] = (a_1, b_1] \times \dots \times (a_n, b_n] : a_i, b_i \in \mathbf{R}\}$$

$$\mathcal{I}_\infty^n = \{(-\infty, c] : c \in \mathbf{R}^n\}$$

Beweis. ÜA

□

1.2 Konstruktion von (Wahrscheinlichkeits-)Maßen

Das Konzept der erzeugten σ -Algebra erlaubt es uns, bestimmte Eigenschaften von Maßen nicht auf der gesamten σ -Algebra, sondern nur auf dem Erzeuger (also beispielsweise auf den halboffenen Intervallen $(a, b] \subset \mathbf{R}$, $a, b \in \mathbf{R}$) nachzuweisen und diese dann auf das Mengensystem der σ -Algebra hochzuziehen. Was damit gemeint ist, wollen wir nun genauer untersuchen. Zunächst einige Definitionen.

Definition 1.8 (Messraum, messbare Menge) Es sei Ω nichtleer und \mathcal{B} eine σ -Algebra über Ω . Das Paar (Ω, \mathcal{B}) heißt *Messraum*; die Elemente von \mathcal{B} heißen *messbare Mengen*.

Definition 1.9 (Maß, σ -endliches Maß, Maßraum) Es sei (Ω, \mathcal{E}) ein Messraum. Eine Mengenfunktion $\mu : \mathcal{E} \rightarrow \bar{\mathbf{R}}$ heißt *Maß auf \mathcal{E}* , wenn sie folgende Eigenschaften erfüllt:

1. $\mu(\emptyset) = 0$ (*Nulltreue*),
2. $\mu(E) \geq 0$ für alle $E \in \mathcal{E}$,
3. $\mu(\cup_n E_n) = \sum_{n=1}^\infty \mu(E_n)$ für jede Folge $E_1, E_2, E_3, \dots \in \mathcal{E}$ von paarweise disjunkten Teilmengen.

Das Maß μ wird *σ -endlich* genannt, wenn es eine Folge E_1, E_2, E_3, \dots in \mathcal{E} gibt, für die $E_n \uparrow \Omega$ gilt und $\mu(E_n) < \infty$ für alle $n \in \mathbf{N}$. Das Tripel $(\Omega, \mathcal{B}, \mu)$ wird *Maßraum* genannt.

Bekannte Beispiele sind das Diracmaß (auch *Punktmaß* oder *Punktmasse* genannt)

$$\delta_\omega : \mathcal{E} \rightarrow \{0, 1\}, \quad \delta_\omega(E) = \begin{cases} 0, & \omega \notin E \\ 1, & \omega \in E \end{cases}$$

oder das Zählmaß

$$\mu : \mathcal{A} \rightarrow \mathbf{N} \cup \{\infty\}, \quad \mu(A) = \begin{cases} |A|, & A \text{ endlich} \\ \infty, & \text{sonst.} \end{cases}$$

Wie man leicht einsieht, erfüllt auch das Nullmaß $\mu = 0$ die obige Definition.

Die Definition des Maßraums ähnelt bereits stark der Definition des Wahrscheinlichkeitsraums (Ω, \mathcal{B}, P) für ein Wahrscheinlichkeitsmaß P . Ein Wahrscheinlichkeitsraum ist nun schlicht ein Maßraum (Ω, \mathcal{B}, P) mit $P(\Omega) = 1$; die Normiertheit impliziert zusammen mit der Additivität des Maßes P die Nulltreue $P(\emptyset) = 0$.

Elementare Eigenschaften von Maßen

Einige wesentliche Eigenschaften, die sich aus der σ -Additivität ergeben, fasst der folgende Satz zusammen.

Lemma 1.10 (Folgerungen aus der σ -Additivität) *Sei $(\Omega, \mathcal{E}, \mu)$ ein Maßraum und $E, F, E_n \in \mathcal{E}$, $n \in \mathbf{N}$.*

1. Endliche Additivität: *Sind E, F disjunkt, so gilt*

$$\mu(E \cup F) = \mu(E) + \mu(F).$$

2. Subtraktivität: *Ist $E \subset F$ und $\mu(E) < \infty$, so gilt*

$$\mu(F \setminus E) = \mu(F) - \mu(E).$$

3. Monotonie: *Im Falle $E \subset F$ ist $\mu(E) \leq \mu(F)$.*

4. Sub- σ -Additivität: *Es gilt*

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu(E_n)$$

Beweis. Mit der Folge $E, F, \emptyset, \emptyset, \dots$ folgt aus der σ -Additivität zusammen mit der Nulltreue sofort die endliche Additivität. Damit ergibt sich sofort die Subtraktivität, denn für $E \subset F$ ist $F = F \setminus E \cup E$ eine disjunkte Zerlegung von F und aus

$$\mu(F) = \mu(F \setminus E) + \mu(E) \geq 0$$

ergeben sich in einem Zug die Subtraktivität und die Monotonie. Um die Sub- σ -Additivität zu zeigen, schreiben wir die Vereinigung der E_n als disjunkte Zerlegung

$$\bigcup_{n=1}^{\infty} E_n = \bigcup_{n=1}^{\infty} \left(E_n \setminus \bigcup_{k=1}^{n-1} E_k \right)$$

und erhalten somit wegen der Subtraktivität

$$\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \mu\left(\bigcup_{n=1}^{\infty} \left(E_n \setminus \bigcup_{k=1}^{n-1} E_k \right)\right) = \sum_{n=1}^{\infty} \mu\left(E_n \setminus \bigcup_{k=1}^{n-1} E_k \right) \leq \sum_{n=1}^{\infty} \mu(E_n)$$

□

Eine weitere wichtige Konsequenz der σ -Additivität von Maßen ist ihre Stetigkeit.

Satz 1.11 *Sei $(\Omega, \mathcal{E}, \mu)$ ein Maßraum und $E, E_n \in \mathcal{E}$, $n \in \mathbf{N}$. Dann gilt:*

1. Stetigkeit von unten: *Aus $E_n \uparrow E$ folgt $\mu(E_n) \uparrow \mu(E)$.*

2. Stetigkeit von oben: *Aus $E_n \downarrow E$ und $\mu(E_1) < \infty$ folgt $\mu(E_n) \downarrow \mu(E)$.*

Beweis. ÜA

□

Fortsetzbarkeit von Maßen auf eine σ -Algebra

Wir kommen zu einem wichtigen Baustein bei der Konstruktion eines eindeutigen Maßes auf den Borelmengen, dem Fortsetzungs- und Eindeutigkeitssatz von Carathéodory.³ Als Mengensystem \mathcal{F} wählen wir einen Erzeuger der Borelschen σ -Algebra, den wir mit der algebraischen Struktur eines Rings versehen.

Definition 1.12 (Ring) Ein System \mathcal{F} von Teilmengen über einer Menge Ω heißt *Ring*, wenn

1. $\emptyset \in \mathcal{F}$,
2. $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$,
3. $A, B \in \mathcal{F} \Rightarrow B \setminus A \in \mathcal{F}$.

³ Konstantinos Karatheodori (1873–1950), griech. Mathematiker

Aus der Definition 1.2 folgt zusammen mit den de Morganschen Regeln, dass jede Algebra auch ein Ring ist.

Satz 1.13 (Carathéodory, 1914) *Jedes σ -endliche Maß μ auf einem Ring \mathcal{E} (d.h., es existieren $E_n \in \mathcal{E}$ mit $E_n \uparrow \Omega$ und $\mu(E_n) < \infty$ für alle $n \in \mathbf{N}$) lässt sich eindeutig auf die von \mathcal{E} erzeugte σ -Algebra fortsetzen, d.h., es existiert ein eindeutiges Maß μ^* auf $\sigma(\mathcal{E})$, so dass $\mu^*(E) = \mu(E)$ für alle $E \in \mathcal{E}$.*

Beweis. Wir werden den Beweis nur skizzieren und wie zuvor auf die einschlägige Literatur verweisen.

1. Zunächst definieren wir ein sogenanntes äußeres Maß durch

$$\mu^*: 2^\Omega \rightarrow \bar{R}, \quad \mu^*(E) = \inf_{\cup_n E_n \supset E} \sum_{n=1}^{\infty} \mu(E_n), \quad (1.1)$$

wobei $E_n \in \mathcal{E}$ und wir ohne Einschränkung der Allgemeinheit annehmen können, dass die Überdeckung von $E \subset \cup_n E_n$ durch die E_n disjunkt ist (ansonsten wählen wir statt der $E_n \in \mathcal{E}$ die Mengen $E_n \setminus (E_1 \cup \dots \cup E_{n-1}) \in \mathcal{E}$).

2. Von μ^* lässt sich zeigen, dass $\mu^*(E) = \mu(E)$ für alle $E \in \mathcal{E}$ und dass es sub- σ -additiv ist, d.h., dass

$$\mu^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mu^*(E_n). \quad (1.2)$$

Die erste Eigenschaft ergibt sich im wesentlichen direkt aus (1.1), indem man sich zunächst klarmacht, dass $\mu^*(E) \leq \mu(E)$ und dass wegen der σ -Additivität von μ umgekehrt $\mu^*(E) \geq \mu(E)$ gelten muss; folglich ist $\mu^*(E) = \mu(E)$ für alle $E \in \mathcal{E}$. Um die Sub- σ -Additivität zu beweisen, nehmen wir an, dass $\sum_n \mu^*(E_n) < \infty$ ist – andernfalls gilt die Ungleichung (1.2) trivialerweise – und betrachten eine Folge $\{F_k^n\}$, $k \in \mathbf{N}$ von Überdeckungen von E_n für jeweils festes $n \in \mathbf{N}$. Die Überdeckungen seien so gewählt, dass

$$\sum_{k=1}^{\infty} \mu^*(F_k^n) \leq \mu^*(E_n) + \frac{\epsilon}{2^n}, \quad \epsilon > 0.$$

Summieren über n liefert wegen der Eigenschaften der geometrischen Reihe sofort

$$\sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mu^*(F_k^n) \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \epsilon.$$

Mit der Überdeckungseigenschaft der $\{F_k^n\}$ folgt nun

$$\mu^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mu^*(F_k^n) \leq \sum_{n=1}^{\infty} \mu^*(E_n) + \epsilon,$$

und der Grenzwert $\epsilon \rightarrow 0$ liefert die Sub- σ -Additivität des äußeren Maßes.

3. Wir definieren eine Menge $B \subset \Omega$ als μ^* -messbar, wenn

$$\mu^*(A) \geq \mu^*(A \cap B) + \mu^*(A \cap B^c) \quad \forall A \in 2^\Omega \quad (1.3)$$

und zeigen dass die μ^* -messbaren Mengen eine σ -Algebra \mathcal{B}^* bilden und dass μ^* auf \mathcal{B}^* σ -additiv ist.

Wir werden als erstes zeigen, dass die μ^* -messbaren Mengen eine Algebra bilden. Dazu machen wir uns zunächst klar, dass (1.3) wegen der Subadditivität von μ^* gleichbedeutend ist mit der Aussage

$$\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c) \quad \forall A \in 2^\Omega \quad (1.4)$$

d.h., B ist μ^* -messbar, wenn Sie jede Teilmenge $A \subset \Omega$ in disjunkte Teilmengen zerlegt, auf denen μ^* additiv ist. Damit sehen wir sofort, dass $\emptyset \in \mathcal{B}^*$; aufgrund der Symmetrie von (1.4) liegt mit jeder Menge $B \in \mathcal{B}^*$ auch ihr Komplement in \mathcal{B}^* . Für zwei Mengen $B_1, B_2 \in \mathcal{B}^*$ gilt nun

$$\begin{aligned} \mu^*(A) &\geq \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c) \\ &\geq \mu^*(A \cap B_1) + \mu^*(A \cap B_1^c \cap B_2) + \mu^*(A \cap B_1^c \cap B_2^c) \\ &\geq \mu^*((A \cap B_1) \cup (A \cap B_1^c \cap B_2)) + \mu^*(A \cap B_1^c \cap B_2^c) \\ &= \mu^*((A \cap (B_1 \cup B_2)) \cup (A \cap (B_1 \cup B_2)^c)), \end{aligned}$$

wobei wir in den ersten beiden Ungleichungen die Definition (1.3) und in der dritten Ungleichung die Subadditivität von μ^* ausgenutzt haben – die letzte Zeile folgt aus der Anwendung der de Morganschen Regeln und der Distributivität der Mengenoperationen “ \cap ” und “ \cup ”. Damit ist $B_1 \cup B_2 \in \mathcal{B}^*$.

Abgeschlossenheit von \mathcal{B}^* bezüglich abzählbaren Vereinigungen und σ -Additivität von μ^* zeigt man in einem Zuge – durch Induktion: Sei E_1, E_2, \dots eine Folge disjunkter Mengen in \mathcal{E} mit $E = \cup_n E_n$. Wenn wir in (1.4) für A die Menge $A \cap (E_1 \cup E_2)$ und für B die Menge E_1 einsetzen, so folgt aus

$$\mu^*(A \cap (E_1 \cup E_2)) = \mu^*(A \cap E_1) + \mu^*(A \cap E_2)$$

per Induktion, dass

$$\mu^* \left(A \cap \bigcup_{n=1}^m E_n \right) = \sum_{n=1}^m \mu^*(A \cap E_n) \quad \forall m \in \mathbf{N}.$$

Da wir soeben bewiesen haben, dass $\cup_{n=1}^m E_n \in \mathcal{B}^*$, ergibt sich zusammen mit (1.3), dass

$$\begin{aligned} \mu^*(A) &\geq \mu^* \left(A \cap \bigcup_{n=1}^m E_n \right) + \mu^* \left(A \cap \left(\bigcup_{n=1}^m E_n \right)^c \right) \\ &\geq \sum_{n=1}^m \mu^*(A \cap E_n) + \mu^*(A \cap E^c), \end{aligned}$$

insbesondere also

$$\mu^*(A) \geq \sum_{n=1}^{\infty} \mu^*(A \cap E_n) + \mu^*(A \cap E^c).$$

Andererseits folgt aus der Sub- σ -Additivität des äußeren Maßes, dass

$$\begin{aligned} \sum_{n=1}^{\infty} \mu^*(A \cap E_n) + \mu^*(A \cap E^c) &\geq \mu^*(A \cap E) + \mu^*(A \cap E^c) \\ &\geq \mu^*(A). \end{aligned}$$

Aus der Kombination der letzten beiden Ungleichungen können wir also schließen, dass

$$\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A \cap E_n) + \mu^*(A \cap E^c) = \mu^*(A \cap E) + \mu^*(A \cap E^c).$$

Die zweite Gleichheitszeichen drückt dabei aus, dass $E = \cup_n E_n$ in \mathcal{B}^* liegt, die erste Gleichung besagt, dass das äußere Maß μ^* auf \mathcal{B}^* σ -additiv ist – voilà!

- Es bleibt zu zeigen, dass $\mathcal{B}^* \supset \sigma(\mathcal{E})$. Das ist aber klar, denn wir haben bewiesen, dass \mathcal{B}^* eine σ -Algebra ist, die \mathcal{E} enthält. Also ist mit $\mathcal{B}^* \supset \mathcal{E}$ auch $\mathcal{B}^* \supset \sigma(\mathcal{E})$ und μ^* ist die gesuchte Fortsetzung von μ auf $\sigma(\mathcal{E})$.
- Die Eindeutigkeit des fortgesetzten Maßes ergibt sich aus einer Eigenschaft monotoner Klassen. Dazu nehmen wir an, es würden zwei unterschiedliche Fortsetzungen μ_1 und μ_2 auf $\sigma(\mathcal{E})$ existieren und definieren

$$\mathcal{M} = \{M \in \sigma(\mathcal{E}) : \mu_1(M) = \mu_2(M)\}.$$

Aus der Stetigkeit von Maßen folgt nun, dass wenn M_n eine aufsteigende Kette von Teilmengen ist (d.h. $M_1 \subset M_2 \subset \dots$), auch ihre Vereinigung $\cup_n M_n$ in \mathcal{M} liegt bzw. für eine absteigenden Kette $M_1 \supset M_2 \supset \dots$ ihr Schnitt $\cap_n M_n$. Eine solche Familie $\mathcal{M} \subset 2^\Omega$ heißt *monotone Klasse*. Nach dem Satz über monotone Klassen [7, Satz I.6.2] gilt aber, dass die kleinste monotone Klasse, die \mathcal{E} enthält, gerade $\sigma(\mathcal{E})$ ist und somit

$$\mu_1(E) = \mu_2(E) = \mu^*(E) \quad \forall E \in \sigma(\mathcal{E}).$$

□

Anmerkung 1.14 In einer etwas allgemeineren Fassung des Satzes von Carathéodory ist das Maß μ nur auf einem Halbring statt auf einem Ring definiert. Man spricht in diesem Fall von einem *Prämaß*. (Jeder Ring ist auch ein Halbring.) Der allgemeine Fall wird in [7, §II.4] diskutiert.

Was haben wir nun gewonnen? Offenbar löst der Satz von Carathéodory nicht unser Maßproblem, sondern führt es auf seine Lösung auf einem Mengenring zurück. Die weitere Strategie zur Definition eines eindeutigen Wahrscheinlichkeitsbegriffs für Borelmengen wird sein, den Halbring der halboffenen Teilmengen $(a, b] \subset \mathbf{R}^n$ etwas aufzupeppen, so dass er zu einem Ring wird und darauf dann unser Maß zu definieren.

Lebesgue-Maß

Die Tatsache, dass wir den Intervallen problemlos einen Inhalt bzw. ein Volumen zuordnen können und dass die halboffenen Intervalle die σ -Algebra der Borelmengen erzeugen, erklärt die Bedeutung der Borel- σ -Algebra für die Maßtheorie. Wir beschränken uns zunächst auf den Fall $n = 1$. Sei also $\mathcal{I} = \mathcal{I}^1$ das System der halboffenen Intervalle $I_{a,b} = (a, b]$ auf \mathbf{R} ; einen Ring \mathcal{F} erhalten wir, wenn wir zu \mathcal{I} die disjunkten endlichen Vereinigungen der Intervalle $I_{a,b}$ und die leere Menge hinzunehmen. Da die von \mathcal{F} erzeugte σ -Algebra $\sigma(\mathcal{F})$ gerade die Borel- σ -Algebra $\mathcal{B}(\mathbf{R})$ über \mathbf{R} ist, besagt Satz 1.13, dass ein eindeutiges Maß λ auf den Borelmengen existiert, das jedem Intervall $I_{a,b}$ seinen elementargeometrischen Inhalt $\lambda(I_{a,b}) = b - a$, sprich: seine Länge zuordnet. Die letzte Aussage können wir als Definition des n -dimensionalen Lebesgue-Maßes heranziehen:

Definition 1.15 (Lebesgue-Maß, Lebesgue-Borel-Maß) Das Maß $\lambda^n: \mathcal{B}(\mathbf{R}^n) \rightarrow [0, \infty]$, das jedem n -dimensionalen Intervall $I_{a,b}^n = (a_1, b_1] \times \dots \times (a_n, b_n] \subset \mathbf{R}^n$ sein Volumen

$$\lambda^n(I_{a,b}^n) = \prod_{i=1}^n (b_i - a_i),$$

zuordnet heißt *Lebesgue-Maß* bzw. *Lebesgue-Borel-Maß*. Im Fall $n = 1$ schreiben wir einfach $\lambda^1 = \lambda$.

Beispiel 1.16 Dass abzählbare Mengen Lebesgue-Nullmengen sind, folgt sofort aus der Stetigkeit des Maßes von oben und der σ -Additivität. Wegen

$$\lambda(\{x\}) = \lim_{n \rightarrow \infty} \lambda((x - 1/n, x]) = \lim_{n \rightarrow \infty} 1/n = 0$$

gilt für jede abzählbare Vereinigung V von Punktengen, dass $\lambda(V) = 0$. Insbesondere gilt

$$\lambda(\mathbf{Q}) = \sum_{x \in \mathbf{Q}} \lambda(\{x\}) = 0.$$

Da das Lebesgue-Maß all die geforderten Eigenschaften eines Volumenmaßes hat, ist es die Lösung unseres Maßproblems, sofern wir die Einschränkung auf Borelmengen akzeptieren. Doch zurück zum Fall $n = 1$ und zu den Wahrscheinlichkeiten: Wie wir soeben gesehen haben, hat das Lebesgue-Maß die Eigenschaft, dass neben der leeren Menge auch jede höchstens abzählbare Menge das Maß 0 hat; dies wird bei der Betrachtung von *fast sicheren* Ereignissen eine wichtige Rolle spielen wird.

Anmerkung 1.17 Das Lebesgue-Maß ist σ -endlich, aber nicht endlich, denn aus der Monotonie folgt, dass

$$\lambda(\mathbf{R}) \geq \lambda((0, n]) = n, \quad n \in \mathbf{N},$$

und damit $\lambda(\mathbf{R}) = \infty$.

Wahrscheinlichkeitsmaß, maßerzeugende Funktion

Wir werden nun zeigen, dass eine Eins-zu-eins-Beziehung zwischen Wahrscheinlichkeitsmaßen P – im Falle der Gleichverteilung sind das gerade die normierten Lebesgue-Maße – und einer bestimmten Klasse rechtsstetiger Funktionen, den *Verteilungsfunktionen*, existiert. Dazu sei $F: \mathbf{R} \rightarrow \mathbf{R}$ eine Funktion die monoton wachsend ist (im Sinne von *nicht fallend*); für F gelte

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1.$$

und wir definieren ein endlich additives Wahrscheinlichkeitsmaß durch die Abbildung

$$P(I_{a,b}) = F(b) - F(a)$$

zunächst für beliebige Intervalle $I_{a,b} \in \mathcal{I}$ bzw. als Summe

$$P\left(\bigcup_{n=1}^m I_{a_n, b_n}\right) = \sum_{n=1}^m (F(b_n) - F(a_n))$$

für die endlichen, disjunkten Vereinigungen, die in \mathcal{F} enthalten sind. Das folgende Resultat beleuchtet den Zusammenhang zwischen Verteilungsfunktion und Wahrscheinlichkeiten etwas genauer.

Satz 1.18 (frei nach Lebesgue, 1902) *P ist σ -additiv auf \mathcal{B} genau dann, wenn F rechtsseitig stetig ist. Zu jedem σ -additiven Wahrscheinlichkeitsmaß existiert also genau eine rechtsseitig stetige Funktion F mit*

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow \infty} F(x) = 1,$$

so dass

$$F(x) = P(I_{-\infty, x}) \quad \forall x \in \mathbf{R}$$

und umgekehrt.

Beweis. Wir stellen zunächst $I_{a,b} = (a, b]$ als abzählbare, disjunkte Vereinigung

$$I_{a,b} = \bigcup_{n=1}^{\infty} I_n, \quad I_n = (a_n, b_n]$$

dar. Wegen der endlichen Additivität von P gilt dann

$$P\left(\bigcup_{n=1}^{\infty} I_n\right) = \sum_{n=1}^m P(I_n) + P\left(\bigcup_{n=m}^{\infty} I_n\right) \geq \sum_{n=1}^m P(I_n). \quad (1.5)$$

“ \Rightarrow ” Ist also P σ -additiv, so muss gelten, dass $P(B_n) \rightarrow 0$ für jede absteigende Kette $B_1 \supset B_2 \supset \dots \in \mathcal{B}$ mit $B_n \downarrow \emptyset$, woraus sofort die Stetigkeit von F von rechts folgt.

“ \Leftarrow ” Um umgekehrt die σ -Additivität von P aus der rechtsseitigen Stetigkeit zu folgern, verfahren wir wie folgt:

1. Da die Ungleichung (1.5) für alle $m \in \mathbf{N}$ gilt, ist

$$P(I_{a,b}) \geq \sum_{n=1}^{\infty} P(I_n). \quad (1.6)$$

2. Sei nun $\epsilon > 0$ fest. Wegen der rechtsseitigen Stetigkeit von F gibt es ein $\alpha \in (a, b]$, so dass $F(\alpha) - F(a) \leq \epsilon/2$. Mit dem gleichen Argument können wir zu jedem Teilintervall $I_n = (a_n, b_n]$ ein $\beta_n > b_n$ finden mit $F(\beta_n) - F(b_n) \leq \epsilon 2^{-n-1}$. Offenbar ist $[\alpha, b] \subset \cup_n (a_n, \beta_n)$, und der Satz von Heine-Borel sagt uns, dass es für $[\alpha, b]$ eine endliche Überdeckung durch offene Mengen gibt, d.h. es gibt ein $m \in \mathbf{N}$, so dass

$$[\alpha, b] \subset \bigcup_{n=1}^m (a_n, \beta_n) \subset \bigcup_{n=1}^m (a_n, b_n].$$

Wegen $(\alpha, b] \subset [\alpha, b]$ gilt das gleichermaßen, wenn wir den linken Randpunkt auslassen, so dass wir aus der (endlichen) Subadditivität von P folgern können, dass

$$P(I_{\alpha, b}) \leq \sum_{n=1}^m P(I_{a_n, \beta_n}). \quad (1.7)$$

Aus der rechtsseitigen Stetigkeit von F bei a und b_n erhalten wir ferner die Abschätzungen

$$P(I_{a_n, \beta_n}) \leq P(I_{a_n, b_n}) + \frac{\epsilon}{2^{n+1}} \quad (1.8)$$

sowie

$$P(I_{\alpha, b}) \leq P(I_{\alpha, b}) + \frac{\epsilon}{2} \quad (1.9)$$

3. Aus (1.7)–(1.9) folgt sodann

$$\begin{aligned} P(I_{a,b}) &\leq \sum_{n=1}^m P(I_{a_n, \beta_n}) + \frac{\epsilon}{2} \\ &\leq \sum_{n=1}^m P(I_{a_n, b_n}) + \frac{\epsilon}{2^{n+1}} + \frac{\epsilon}{2} \\ &\leq \sum_{n=1}^{\infty} P(I_{a_n, b_n}) + \sum_{n=1}^{\infty} \frac{\epsilon}{2^{n+1}} + \frac{\epsilon}{2} \\ &= \sum_{n=1}^{\infty} P(I_{a_n, b_n}) + \epsilon. \end{aligned}$$

und somit

$$P(I_{a,b}) \geq \sum_{n=1}^{\infty} P(I_n) \tag{1.10}$$

Die Ungleichungen (1.6) und (1.10) liefern zusammen die σ -Additivität von P .

□

Anmerkung 1.19 Die maßerzeugende Funktion $F: \mathbf{R} \rightarrow \mathbf{R}$ aus Satz 1.18 heißt *Verteilungsfunktion*. Ihre Eindeutigkeit folgt aus der Tatsache, dass zwei Kandidaten F_1, F_2 für F sich höchstens um eine Konstante c unterscheiden können, d.h., $F_1 = F_2 + c$; nun gilt aber $F_1(x), F_2(x) \rightarrow 1$ für $x \rightarrow \infty$, d.h., $c = 0$.

Anmerkung 1.20 Die Aussage des Satzes 1.18 gilt analog für das Lebesgue-Maß λ . Wird λ durch eine rechtsseitig stetige, monotone Funktion F dargestellt, spricht man auch vom *Lebesgue-Stieltjes-Maß*.

Beispiel 1.21 Das Maß der Gleichverteilung auf einer Grundmenge $\Omega \subset \mathbf{R}$ mit $\lambda(\Omega) < \infty$ ist das normierte Lebesgue-Maß

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

Ist also $\Omega = [c, d]$, so gilt insbesondere

$$P(I_{a,b}) = F(b) - F(a), \quad I_{a,b} \subset [c, d]$$

mit

$$F(x) = \begin{cases} 0, & x \leq c \\ \frac{x-c}{d-c}, & c < x < d \\ 1, & x \geq d \end{cases}$$

1.3 Integration messbarer Funktionen

Im folgenden sei $(\Omega, \mathcal{E}, \mu)$ ein Maßraum. Wir wollen nun den Maßbegriff, mit dessen Hilfe sich Volumina von beliebigen Borelmengen im Prinzip berechnen lassen, auf Maße erweitern, bei denen die Punkte der Basismenge Ω mit einer Funktion gewichtet sind, also etwa in der Art von

$$P(X \leq c) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^c \exp\left(-\frac{x^2}{2}\right) dx = \frac{1}{2} \left(1 + \operatorname{erf}\left(\frac{c}{\sqrt{2}}\right)\right),$$

wenn $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \mathcal{N}(0, 1))$ der Wahrscheinlichkeitsraum der Standard-Normalverteilung ist, wie bei der Berechnung des Volumens der dreidimensionalen Einheitskugel⁴ $K_1 = \{x \in \mathbf{R}^3: |x|^2 \leq 1\} \subset \mathbf{R}^3$,

$$\lambda^3(K_1) = \int_{K_1} d\lambda = \frac{4\pi}{3}$$

oder schließlich bei der Berechnung von Erwartungswerten einer Zufallsvariable $X: \Omega \rightarrow \mathbf{R}$,

$$\mathbf{E}[X] = \int_{\Omega} X(\omega) dP(\omega).$$

⁴ Man mache sich klar, dass $\lambda^4(K_1) = 0$ ist, d.h., die dreidimensionale Einheitskugel $K_1 \subset \mathbf{R}^3$ ist zwar gleichmächtig zum Kontinuum, aber dennoch eine Lebesgue-Nullmenge, wenn man sie als Untermannigfaltigkeit des \mathbf{R}^4 betrachtet.

Messbarkeit, Bildmaß

Die folgende Definition sollte aus der elementaren Stochastik bekannt sein und nicht weiter überraschen – wir identifizieren Zufallsvariable mit *messbaren Funktionen*.

Definition 1.22 (messbare Funktion) Seien $(\Omega_1, \mathcal{F}_1)$ und $(\Omega_2, \mathcal{F}_2)$ zwei Messräume. Die Abbildung $f: \Omega_1 \rightarrow \Omega_2$ heißt \mathcal{F}_1 - \mathcal{F}_2 -messbar, wenn

$$f^{-1}(\mathcal{F}_2) \subset \mathcal{F}_1.$$

Mit anderen Worten: Die Funktion $f: \Omega_1 \rightarrow \Omega_2$ ist \mathcal{F}_1 - \mathcal{F}_2 -messbar, wenn $f^{-1}(B) = \{\omega \in \Omega_1: f(\omega) \in B\}$ für eine beliebige $B \in \mathcal{F}_2$ eine messbare Teilmenge von Ω_1 ist, sprich: ein Element von \mathcal{F}_1 . Sofern keine Unklarheiten bestehen, sprechen wir von f schlicht als *messbarer Funktion*.

Definition 1.23 (messbare Funktion im engeren Sinne, Zufallsvariable) Eine *messbare Funktion* bzw. *Zufallsvariable* ist eine Abbildung $f: \Omega \rightarrow \mathbf{R}$, für die gilt, dass $f^{-1}(B) = \{\omega \in \Omega: f(\omega) \in B\}$ für eine beliebige Borelmenge $B \in \mathcal{B}(\mathbf{R})$ eine messbare Teilmenge von Ω ist.

Sei nun $(\Omega_1, \mathcal{F}_1, \mu)$ ein Maßraum und $f: \Omega_1 \rightarrow \Omega_2$ eine messbare Funktion. Nach der Definition der Messbarkeit ist $f^{-1}(B)$ eine messbare Menge, der ein Maß zugeordnet werden kann. D.h., wir können das transformierte Maß

$$\mu_f: \mathcal{F}_2 \rightarrow \bar{\mathbf{R}}, \quad \mu_f(B) = (\mu \circ f^{-1})(B)$$

bestimmen. Dieses Maß heißt *Bildmaß von μ unter f* .⁵

In der Sprache der Wahrscheinlichkeitstheorie klingt das so: Ist P ein Wahrscheinlichkeitsmaß und $X: \Omega \rightarrow \mathbf{R}$ eine Zufallsvariable, so ist $(\mathbf{R}, \mathcal{B}(\mathbf{R}), P_X)$ der durch X induzierte Wahrscheinlichkeitsraum.

Anmerkung 1.24 Auch bei der Messbarkeit kommt uns das Konzept des Erzeugers zupass. Da nämlich $\mathcal{B}(\mathbf{R})$ eine erzeugte σ -Algebra ist, z.B. von dem System der halboffenen Mengen $I_{a,b}$, reicht es, die Messbarkeitseigenschaft auf dem Erzeugendensystem nachzuweisen, d.h., man zeigt, dass alle $f^{-1}(I_{a,b})$ in \mathcal{F}_1 liegen.

Beispiel 1.25 Wir wollen nun einige Beispiele von messbaren Funktionen aufzählen, die uns beim Entwickeln eines für die Wahrscheinlichkeitstheorie brauchbaren Integralbegriffs begegnen werden. Der Beweis der Messbarkeit der jeweiligen Funktionen bleibt dem Leser als Übungsaufgabe überlassen.⁶

- Für beliebige Mengen $A \in \mathcal{B}(\mathbf{R})$, ist die charakteristische Funktion, auch *Indikatorfunktion* genannt und definiert als

$$\chi_A: \mathcal{B} \rightarrow \{0, 1\}, \quad \chi_A(\omega) = \begin{cases} 0, & \omega \notin A \\ 1, & \omega \in A \end{cases}$$

messbar und beschränkt.

- Ist $\{A_i: i = 1, \dots, m\}$ eine endliche Partition von Ω in messbare Mengen A_i , so ist die Funktion

$$f(\omega) = \sum_{i=1}^m y_i \chi_{A_i}(\omega), \quad y_i \in \mathbf{R}$$

eine messbare Funktion, die als *Treppenfunktion* oder *einfache Funktion* bezeichnet wird; einfache Funktionen sind allgemein messbare Funktionen $f: \Omega \rightarrow \mathbf{R}$ mit endlichem Bild, d.h., $f(\Omega) = \{y_1, \dots, y_m\}$.

- Stetige Funktionen $f: \Omega_1 \rightarrow \Omega_2$ zwischen topologischen Räumen Ω_1, Ω_2 sind messbar.
- Summen, Produkte und Verkettungen von messbaren Funktionen, Grenzwerte messbarer Funktionenfolgen sowie einige elementare Operationen wie “min” oder “max” liefern wieder messbare Funktionen.

⁵ Man spricht bei $f_*\mu(\cdot) = \mu(f^{-1}(\cdot))$ vom *Pushforward* unter f , da f_* das Argument in den Bildraum von f bzw. in die dazugehörige σ -Algebra bewegt; die Umkehrabbildung f^* , d.h., $f^*\mu(\cdot) = \mu(f(\cdot))$ heißt *Pullback*.

⁶ Ansonsten werfe man einen Blick in [10, Kap. 1.3].

Lebesgue-Integral

Wir werden nun die Begriffe *Lebesgue-Integral* und *Lebesgue-Integrierbarkeit* entwickeln. Die Strategie wird ähnlich wie beim Riemann-Integral darin bestehen, zunächst das Integral für eine Folge von Treppenfunktionen zu definieren und dann den Grenzwert zu bilden. Der Einfachheit halber beschränken wir uns auf den Fall

$$\mu(\Omega) < \infty$$

und verfahren wie folgt.

1. Lebesgue-Integral für Treppenfunktionen: Für Treppenfunktionen

$$f_m: \Omega \rightarrow \mathbf{R}, \quad f_m(\omega) = \sum_{i=1}^m y_i \chi_{A_i}(\omega),$$

mit $A_i = f_m^{-1}(y_i)$ und $\Omega = \{A_i: i = 1, \dots, m\}$ definieren wir das Lebesgue-Integral als die Summe

$$\int_{\Omega} f_m d\mu := \sum_{i=1}^m y_i \mu(A_i) \quad (1.11)$$

bzw.

$$\int_B f_m d\mu := \sum_{i=1}^m y_i \mu(A_i \cap B), \quad B \subset \Omega. \quad (1.12)$$

Wie sich leicht zeigen lässt, ist die Lebesgue-Summe von der Darstellung von f_m unabhängig (die y_i müssen i.A. nicht paarweise verschieden sein und die A_i dementsprechend nicht disjunkt). Ebenso zeigt man:

(a) *Linearität*: Sind $f_m, g_n: \Omega \rightarrow \mathbf{R}$ Treppenfunktionen, so gilt

$$\int (\alpha f_m + \beta g_n) d\mu = \alpha \int f_m d\mu + \beta \int g_n d\mu \quad \forall \alpha, \beta \in \mathbf{R}.$$

(b) *Positivität*: Ist $f_n \geq 0$, so ist

$$\int f_n d\mu \geq 0.$$

(c) *Monotonie*: Ist $f_n \leq g_m$, so auch

$$\int f_m d\mu \leq \int g_n d\mu.$$

Wegen der Positivität gilt insbesondere, dass

$$\left| \int f_m d\mu \right| \leq \int |f_m| d\mu \leq \sup_{\omega \in \Omega} |f_m(\omega)| \mu(\Omega).$$

2. **Lebesgue-Integral für beschränkte, messbare Funktionen**: Es sei $f: \Omega \rightarrow \mathbf{R}$ eine beschränkte, messbare Funktion, also $\sup_{\omega} |f(\omega)| \leq M < \infty$. Ferner sei $(f_m)_{m \in \mathbf{N}}$, eine Folge von Treppenfunktionen, die gleichmäßig gegen f konvergiert. Dass eine solche Folge immer existiert, sieht man, indem man z.B. das Intervall $[-M, M]$ in 2^m äquidistante Teilintervalle zerlegt und $f_m = \lfloor 2^m f \rfloor / 2^m$ setzt, wobei $\lfloor x \rfloor$ die größte ganze Zahl ist, die kleiner als x ist. Die Funktion f_m nimmt jeweils nur endlich viele rationale Werte $\{k/2^m: -\lfloor M2^m \rfloor - 1 \leq k \leq \lfloor M2^m \rfloor + 1\}$ an, und die Urbildmengen

$$A_k = \{\omega \in \Omega: k/2^m \leq f(\omega) < (k+1)/2^m\}$$

sind messbar. Das bedeutet, dass

$$f_m(\omega) = \sum_{k=-K}^K \frac{k}{2^m} \chi_{A_k}(\omega), \quad K = \lfloor M2^m \rfloor + 1$$

eine Folge (f_m) von Treppenfunktionen bildet, für die gilt, dass

$$\sup_{\omega \in \Omega} |f_m(\omega) - f(\omega)| \leq 2M/2^m.$$

Sie konvergiert sogar monoton, denn nach Konstruktion ist $f_1(\omega) \leq f_2(\omega) \leq f_3(\omega) \leq \dots$ für alle $\omega \in \Omega$, und wir definieren das Lebesgue-Integral als

$$\int_B f d\mu := \lim_{m \rightarrow \infty} \int_B f_m d\mu, \quad B \subset \Omega. \quad (1.13)$$

Der Grenzwert existiert, denn die Lebesgue-Summe über f_m ist eine Cauchy-Folge.

3. **Verallgemeinerung auf beliebige messbare Funktionen:** Wir nehmen zunächst an, dass f nicht-negativ sei und verzichten dabei auf die Forderung, dass f beschränkt sein soll. Sei also h eine beschränkte, messbare Funktion mit der Eigenschaft $0 \leq h \leq f$. Wir definieren das Lebesgue-Integral als

$$\int f d\mu := \sup_h \left\{ \int h d\mu : h \text{ beschränkt, } 0 \leq h \leq f \right\}. \quad (1.14)$$

Das Lebesgue-Integral als Supremum des Lebesgue-Integrals (1.13) für beschränkte Funktionen ist damit auch für nicht-negative, messbare Funktionen wohldefiniert. (Das Integral kann den Wert unendlich annehmen). Zu guter Letzt lassen wir auch negative Funktionen zu: Seien $f^+ = \max\{0, f\}$ bzw. $f^- = -\min\{0, f\}$ die (messbaren) positiven und negativen Äste von f . Da f^- nicht-negativ ist und sich f durch $f = f^+ - f^-$ darstellen lässt, können wir das Integral beliebiger messbarer Funktionen einfach als

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

definieren.

4. **Eindeutigkeit des Lebesgue-Integrals:** Wir müssen zeigen, dass der Wert des Lebesgue-Integrals für beschränkte und messbare Funktionen unabhängig von der approximierenden Funktionenfolge ist. Dazu betrachten wir zwei monotone Folgen $(f_m)_{m \in \mathbf{N}}$ und $(g_m)_{m \in \mathbf{N}}$ von Treppenfunktionen, die gleichmäßig gegen f konvergieren. Wir müssen zeigen, dass dann

$$\lim_{m \rightarrow \infty} \int f_m d\mu = \lim_{m \rightarrow \infty} \int g_m d\mu.$$

Dabei können wir ohne Einschränkung der Allgemeinheit annehmen, dass

$$f_m(\omega) \leq f(\omega) \leq g_m(\omega) \leq f(\omega) + \frac{L}{2^m}$$

für alle $\omega \in \Omega$ und für eine geeignete Konstante $L < \infty$. (Zum Beispiel können wir f_m wie oben und $g_m = \lceil 2^m f \rceil / 2^m$ wählen, wobei $\lceil x \rceil$ die kleinste ganze Zahl größer x ist, womit $L = 2M < \infty$ ist.) Aus der Monotonie der Lebesgue-Summe und der Positivität folgt sogleich, dass

$$\int f_m d\mu \leq \int g_m d\mu \leq \int f_m d\mu + \frac{L}{2^m} \mu(\Omega) \quad \forall m \in \mathbf{N}$$

und damit

$$\lim_{m \rightarrow \infty} \int f_m d\mu = \lim_{m \rightarrow \infty} \int g_m d\mu.$$

Wir fassen die letzten Schritte mit einer Definition und einem Satz über die wichtigsten Eigenschaften des Lebesgue-Integrals zusammen.

Definition 1.26 (Lebesgue-Integrierbarkeit, Lebesgue-Integral) Sei $f: \Omega \rightarrow \mathbf{R}$ eine messbare Funktion mit Positiv- und Negativteil f^+ und f^- . Die Funktion f heißt *Lebesgue-integrierbar* oder *μ -integrierbar*, wenn

$$\int f^+ d\mu < \infty \quad \text{und} \quad \int f^- d\mu < \infty.$$

Das Lebesgue-Integral ist durch

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

definiert.

Lemma 1.27 (Eigenschaften des Lebesgue-Integrals) Für μ -integrierbare Funktionen $f, g: \Omega \rightarrow \mathbf{R}$ gilt:

1. Linearität: Für beliebige $\alpha, \beta \in \mathbf{R}$ ist

$$\int (\alpha f + \beta g) d\mu = \alpha \int f d\mu + \beta \int g d\mu.$$

2. Positivität: Ist $f \geq 0$, so ist

$$\int f d\mu \geq 0.$$

3. Monotonie: Ist $f \leq g$, so auch

$$\int f d\mu \leq \int g d\mu.$$

Wegen der Positivität gilt insbesondere, dass

$$\left| \int f d\mu \right| \leq \int |f| d\mu.$$

Beweis. Die Eigenschaften des Lebesgue-Integrals ergeben sich direkt aus den Eigenschaften der Lebesgue-Summe (ÜA). \square

Vergleich mit dem Riemann-Integral

Der wesentliche Unterschied zwischen dem Lebesgue- und dem Riemann-Integral besteht darin, dass bei der Konstruktion des Ersteren der Bildbereich, während bei Letzterem der Definitionsbereich der zu integrierenden Funktion diskretisiert wird. Aus der Konstruktion ergibt sich insbesondere gilt, dass, wenn das Lebesgue-Integral von $f = f^+ - f^-$ endlich ist, es dann auch für $|f| = f^+ + f^-$ endlich ist (vergl. Definition 1.26).

Auch wenn sich Riemann- und Lebesgue-Integral in ihrer Konstruktion unterscheiden, und nicht alle Lebesgue-integrierbaren Funktionen auch Riemann-integrierbar sind, so hilft das Riemann-Integral dennoch beim Berechnen des Lebesgue-Integrals. Es lässt sich zeigen (siehe [10, Satz 2.17]), dass jede Funktion $f: \mathbf{R} \rightarrow \mathbf{R}$, die auf einem Intervall $[a, b] \subset \mathbf{R}$ Riemann-integrierbar ist, ebenfalls Lebesgue-integrierbar ist, wobei gilt:

$$\int_{[a,b]} f d\mu = \int_a^b f(x) dx.$$

Der Ausdruck auf der rechten Seite der Gleichung ist dabei das Riemann-Integral über die Funktion f , das sich über eine Stammfunktion von f ausrechnen lässt.

Beispiel 1.28 Dass das Lebesgue-Integral eine größere Klasse von Funktionen zu integrieren erlaubt als das Riemann-Integral, verdeutlicht dieses Beispiel: Es sei $\chi_{\mathbf{Q}}: \mathbf{R} \rightarrow \{0, 1\}$ die Dirichlet-Funktion, d.h., die charakteristische Funktion der rationalen Zahlen,

$$\chi_{\mathbf{Q}}(\omega) = \begin{cases} 0, & \omega \notin \mathbf{Q} \\ 1, & \omega \in \mathbf{Q}. \end{cases}$$

Für das Lebesgue-Integral ist

$$\int_{[0,1]} \chi_{\mathbf{Q}} d\lambda = \int_{\mathbf{R}} \chi_{\mathbf{Q} \cap [0,1]} d\lambda = \lambda(\mathbf{Q} \cap [0, 1]) = 0,$$

da das Lebesgue-Maß einer abzählbaren Punktmenge – und das ist $\mathbf{Q} \cap [0, 1]$ – gleich null ist. Das Riemann-Integral ist jedoch für diese Funktion nicht definiert, da die Obersumme stets den Wert 1 hat, die Untersumme jedoch 0 ist, da für jede noch so feine Zerlegung $\cup_k I_k = [0, 1]$ von $[0, 1]$ in jedem der Teilintervalle I_k stets beliebig viele rationale und irrationale Zahlen liegen.

1.4 Konvergenzsätze der Maßtheorie (starke Konvergenz)

Wir werden nun Aussagen über die Vertauschbarkeit von Integral und Limes bei der Integration über Funktionenfolgen untersuchen. Wie zuvor beschränken wir uns wieder auf den Fall

$$\mu(\Omega) < \infty.$$

Obwohl bis auf die Ausnahme von Lemma 1.31 und Satz 1.32 alle der folgenden Konvergenzsätze auch für unendliche Maße gelten, nehmen wir der Einfachheit halber an, dass $\mu(\Omega) < \infty$ sei. Das ist für unsere Zwecke ausreichend und macht die Beweise eleganter. Als erstes werden wir uns als die verschiedenen Varianten der Aussage f_m konvergiert gegen f für Zufallsvariablen (sprich: messbare Funktionen) ins Gedächtnis rufen.

Definition 1.29 (Fast sichere Konvergenz) Eine Folge von messbaren Funktionen f_n konvergiert fast sicher gegen f (symbolisch: $f_n \xrightarrow{\text{f.s.}} f$), wenn es eine μ -Nullmenge $N \subset \Omega$ gibt und

$$\lim_{n \rightarrow \infty} f_n(\omega) = f(\omega) \quad \forall \omega \in \Omega \setminus N.$$

Definition 1.30 (Konvergenz dem Maße nach bzw. Konvergenz in Wahrscheinlichkeit) Eine Folge von messbaren Funktionen f_n konvergiert dem Maße nach gegen f (symbolisch: $f_n \xrightarrow{\mu} f$), wenn

$$\lim_{n \rightarrow \infty} \mu(\{\omega \in \Omega: |f_n(\omega) - f(\omega)| \geq \epsilon\}) = 0 \quad \forall \epsilon > 0.$$

Lemma 1.31 Fast sichere Konvergenz impliziert Konvergenz dem Maße nach, die Umkehrung gilt nicht.

Beweis. Sei $\mu(N) = 0$ für eine Menge $N \subset \Omega$. Konvergenz $f_n(\omega) \rightarrow f(\omega)$ für alle $\omega \in \Omega \setminus N$ bedeutet, dass

$$\bigcup_{n=m}^{\infty} \{\omega \in \Omega: |f_n(\omega) - f(\omega)| \geq \epsilon\} \downarrow \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{\omega \in \Omega: |f_n(\omega) - f(\omega)| \geq \epsilon\} \subset N$$

für alle $\epsilon > 0$. Wegen der Stetigkeit des Maßes von oben gilt, dass

$$\begin{aligned} \mu(\{\omega \in \Omega: |f_n(\omega) - f(\omega)| \geq \epsilon\}) &\leq \mu\left(\bigcup_{n=m}^{\infty} \{\omega \in \Omega: |f_n(\omega) - f(\omega)| \geq \epsilon\}\right) \\ &\rightarrow \mu\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{\omega \in \Omega: |f_n(\omega) - f(\omega)| \geq \epsilon\}\right). \end{aligned}$$

Der letzte Ausdruck ist wegen der Monotonie des Maßes durch $\mu(N)$ nach oben beschränkt. Nach Voraussetzung ist jedoch $\mu(N) = 0$, somit folgt $f_n \xrightarrow{\mu} f$. Für die Umkehrung reicht ein Gegenbeispiel: Die Folge $f_n(\omega) = \chi_{I_n}(\omega)$ auf $\Omega = [0, 1]$ mit $I_1 = [0, 1/2]$, $I_2 = (1/2, 1]$, $I_3 = [0, 1/3]$, $I_4 = (1/3, 2/3]$, $I_5 = (2/3, 1]$, $I_6 = [0, 1/4]$ etc. konvergiert dem Maße nach gegen 0, nicht aber fast sicher. \square

Satz von der beschränkten Konvergenz

Der folgende Satz gibt Auskunft über die Vertauschbarkeit von Integral und Grenzwert im Falle gleichmäßig beschränkter Folgen und ist der einzige Konvergenzsatz, der nicht auf Maßräumen mit unbeschränktem Maß gilt.

Satz 1.32 (Lebesgue, 1902) Für eine Folge $(f_n)_{n \in \mathbf{N}}$ messbarer, gleichmäßig beschränkter Funktionen (d.h., für alle $n \in \mathbf{N}$ gilt $\|f_n(\omega)\|_{\infty} \leq M < \infty$), die dem Maße nach gegen f konvergiert, gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Beweis. Es gilt

$$\left| \int f_n d\mu - \int f d\mu \right| = \left| \int (f_n - f) d\mu \right| \leq \int |f_n - f| d\mu,$$

und es reicht zu zeigen, dass aus $f_n \xrightarrow{\mu} 0$ auch $\int |f_n| d\mu \rightarrow 0$ folgt. Nun ist

$$\begin{aligned} \int |f_n| d\mu &= \int_{|f_n| < \epsilon} |f_n| d\mu + \int_{|f_n| \geq \epsilon} |f_n| d\mu \\ &\leq \epsilon \mu(\{\omega \in \Omega : |f_n| < \epsilon\}) + \sup_{\omega \in \Omega} |f_n(\omega)| \mu(\{\omega \in \Omega : |f_n| \geq \epsilon\}). \end{aligned}$$

Für $n \rightarrow \infty$ strebt der zweite Summand wegen der gleichmäßigen Beschränktheit der f_n , d.h. $\sup_{\omega} |f_n(\omega)| \leq C_1 < \infty$ für alle $n \in \mathbf{N}$, und der Eigenschaft $f_n \xrightarrow{\mu} 0$ gegen 0, der erste Summand ist wegen der Endlichkeit von μ durch $C_2\epsilon$ mit $0 \leq C_2 < \infty$ beschränkt. Demnach existiert eine Konstante $0 \leq C_3 < \infty$, so dass

$$\limsup_{n \rightarrow \infty} \int |f_n| d\mu \leq C_3 \epsilon \quad \forall \epsilon > 0.$$

Da ϵ beliebig klein gewählt werden kann, folgt die Behauptung. \square

Anmerkung 1.33 Ohne die Annahme $\mu(\Omega) < \infty$ ist der Satz falsch, wie man sich leicht anhand des Beispiels $f_n(\omega) = n^{-1} \chi_{(0,n]}(\omega)$ für das Lebesgue-Maß $\mu = \lambda$ klarmachen kann. Offenbar ist $|f_n| \leq 1$ für alle $\omega \in \Omega$ und $n \in \mathbf{N}$, und f_n konvergiert auf ganz \mathbf{R} punktweise gegen 0. Somit erhalten wir

$$0 = \int_{\mathbf{R}} \lim_{n \rightarrow \infty} f_n d\lambda = \int_{\mathbf{R}} 0 d\lambda \neq \lim_{n \rightarrow \infty} \int_{\mathbf{R}} f_n d\lambda = \lim_{n \rightarrow \infty} \frac{\lambda((0, n])}{n} = 1.$$

Auch ohne die gleichmäßige Beschränktheit geht es schief: Die Folge $g_n(\omega) = n\omega^n$ auf $\Omega = [0, 1]$ konvergiert fast sicher gegen 0, denn $g_n \rightarrow 0$ außer auf der Lebesgue-Nullmenge $N = \{1\}$. Damit gilt erst recht $g_n \xrightarrow{\lambda} 0$, aber

$$\int_{[0,1]} g_n d\lambda = \frac{n}{n+1} \rightarrow 1.$$

Lemma von Fatou und Satz von der monotonen Konvergenz

Wie der letzte Satz ohne die gleichmäßige Beschränktheit aussähe, sagt das berühmte Lemma von Fatou.⁷

Satz 1.34 (Fatou, 1907) Sei $(f_n)_{n \in \mathbf{N}}$ eine Folge nicht-negativer, messbarer Funktionen mit $f_n \xrightarrow{\mu} f$. Dann gilt

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

Beweis. Sei g eine beschränkte und messbare Funktion mit der Eigenschaft $0 \leq g \leq f$. Dann ist die Folge $h_n = \min\{f_n, g\}$ gleichmäßig beschränkt und konvergiert dem Maße nach gegen g :

$$h_n \xrightarrow{\mu} \min\{f, g\} = g.$$

Nach dem eben bewiesenen Satz 1.32 gilt damit:

$$\lim_{n \rightarrow \infty} \int h_n d\mu = \int g d\mu.$$

Nun ist aber $\int h_n d\mu \leq \int f_n d\mu$ für alle $n \in \mathbf{N}$, woraus wir folgern können, dass

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

Die Behauptung folgt nun aus der Tatsache, dass $0 \leq g \leq f$ beliebig war und sich das Lebesgue-Integral über $f \geq 0$ als Supremum über Integrale von beschränkten Funktionen gemäß (1.14) schreiben lässt. \square

Als Korollar aus dem Fatou-Lemma erhalten wir einen der wichtigsten Konvergenzsätze der Maßtheorie, der auf Beppo Levi zurückgeht.⁸

⁷ Pierre Fatou (1878–1929), frz. Mathematiker

⁸ Beppo Levi (1875–1961), ital. Mathematiker

Satz 1.35 (Satz von der monotonen Konvergenz, 1906) Für jede monotone Folge $(f_n)_{n \in \mathbf{N}}$ nicht-negativer, messbarer Funktionen mit $f_n \uparrow f$ (punktweise) gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Beweis. Wegen der Monotonie des Integrals gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu \leq \int f d\mu \quad \forall n \in \mathbf{N}.$$

Umgekehrt folgt aus dem Lemma von Fatou, dass

$$\liminf_{n \rightarrow \infty} \int f_n d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu \geq \int f d\mu$$

und damit die Behauptung. \square

Anmerkung 1.36 Der Satz gilt auch unter der schwächeren Voraussetzung, dass $f_n \uparrow f$ (fast sicher).

Korollar 1.37 Ist $g_k \geq 0$ messbar für alle $k \in \mathbf{N}$, so gilt nach Satz 1.35, dass

$$\int \left(\sum_{k=1}^{\infty} g_k \right) d\mu = \sum_{k=1}^{\infty} \left(\int g_k d\mu \right).$$

Anmerkung 1.38 Der Satz der monotonen Konvergenz ist eine Aussage über die Vertauschbarkeit von punktweisen Grenzwerten mit einem Integral. Bekanntermaßen benötigt das Riemann-Integral dafür die gleichmäßige Konvergenz, und tatsächlich gilt die Aussage des Satzes 1.35 nicht für das Riemann-Integral. Als Beispiel betrachten wir die Darstellung der rationalen Zahlen als $\mathbf{Q} = \{q_1, q_2, q_3, \dots\}$ und die Folge messbarer Funktionen

$$f_n: [0, 1] \rightarrow \mathbf{R}, \quad f_n(\omega) = \begin{cases} 0, & \omega \notin \{q_1, \dots, q_n\} \\ 1, & \omega \in \{q_1, \dots, q_n\}. \end{cases}$$

Dann gilt $f_n \uparrow \chi_{\mathbf{Q} \cap [0,1]}$, d.h., f_n konvergiert punktweise monoton gegen die Dirichlet-Funktion $\chi_{\mathbf{Q}}$ auf $[0, 1]$. Da die f_n für jedes n bis auf endlich viele Punkte 0 ist, sind die f_n Riemann-integrierbar, und somit gilt

$$\int_0^1 f_n(x) dx = 0, \quad n \in \mathbf{N}.$$

Der punktweise Grenzwert $\chi_{\mathbf{Q} \cap [0,1]}$ der f_n ist jedoch nicht einmal eine Riemann-integrierte Funktion.

Anmerkung 1.39 Typischerweise wird der Satz von der monotonen Konvergenz ohne die Zuhilfenahme des Fatou-Lemmas bewiesen, da er zum einen älter ist (isb. in seiner ursprünglichen Formulierung durch Lebesgue) und da er zum anderen für bedeutender erachtet wird. Welcher Satz bedeutsamer ist, soll an dieser Stelle nicht diskutiert werden, wir werden aber sehen, dass auch der nächste wichtige Satz, der *Satz von der majorisierten Konvergenz* von Lebesgue, sich mit Hilfe des Lemmas von Fatou beweisen lässt.

Satz von der majorisierten Konvergenz

Der folgende Satz kursiert in der Literatur unter vielen verschiedenen Namen, u.a., als *Satz von der dominierten Konvergenz* oder *Satz von Lebesgue*.

Satz 1.40 (Lebesgue, 1910) Es sei $(f_n)_{n \in \mathbf{N}}$ eine Folge messbarer Funktionen mit $f_n \xrightarrow{\mu} f$. Ferner sei $g: \Omega \rightarrow \mathbf{R}$ eine μ -integrierte Funktion mit $|f_n(\omega)| \leq g(\omega)$ für alle $n \in \mathbf{N}$. Dann gilt

$$\lim_{n \rightarrow \infty} \int f_n d\mu = \int f d\mu.$$

Beweis. Nach Voraussetzung sind die Funktionenfolgen $g \pm f_n$ messbar und nicht-negativ mit $g \pm f_n \xrightarrow{\mu} g \pm f$. Auf $g + f_n$ angewandt liefert Satz 1.34

$$\liminf_{n \rightarrow \infty} \int (g + f_n) d\mu \geq \int (g + f) d\mu.$$

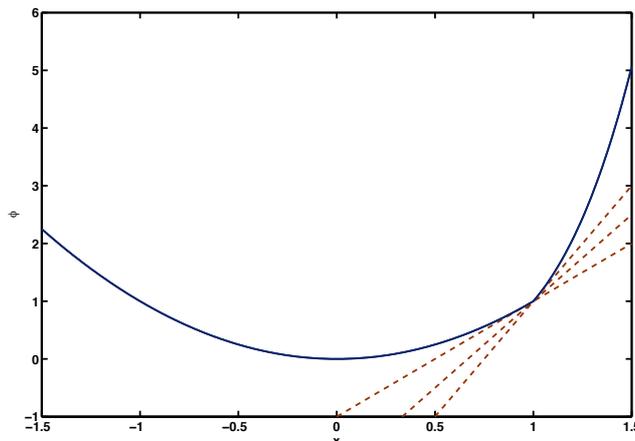


Abb. 2 Konvexe Funktion $\phi(x)$ (blau) und ihre Subgradienten an der Stelle $x = 1$ (rot). Der Subgradient $\partial\phi \in \mathbf{R}^n$ einer konvexen Funktion $\phi: \mathbf{R}^n \rightarrow \mathbf{R}$ an einer Stelle $y \in \mathbf{R}^n$ (hier: $n = 1$) ist dadurch definiert, dass $\phi(x) \geq \phi(y) + \partial\phi^T(x - y)$ für alle $x \in \mathbf{R}^n$ gilt. Mit anderen Worten: Die Tangenten einer konvexen Funktion liegen immer unterhalb ihres Graphen.

Da g integabel ist, können wir $\int g d\mu < \infty$ abziehen, woraus folgt, dass

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

Umgekehrt liefert Satz 1.34 für $g - f_n$ die Ungleichung

$$\limsup_{n \rightarrow \infty} \int (g - f_n) d\mu \geq \liminf_{n \rightarrow \infty} \int (g - f_n) d\mu \geq \int (g - f) d\mu,$$

von der wir wieder $\int g d\mu < \infty$ abziehen dürfen. Zusammen mit der Ungleichung für $g + f_n$ erhalten wir

$$\liminf_{n \rightarrow \infty} \int f_n d\mu \geq \int f d\mu \geq \limsup_{n \rightarrow \infty} \int f_n d\mu,$$

also Gleichheit und damit die gewünschte Aussage. □

Anmerkung 1.41 Auch dieser Satz wird falsch, wenn man die Bedingung, dass f_n durch eine integrabale Funktion g majorisiert wird, weglässt. Als Beispiel betrachten wir

$$f_n: [0, 1] \rightarrow \mathbf{R}, \quad f_n(\omega) = \begin{cases} n, & \omega \in (0, 1/n] \\ 0, & \text{sonst.} \end{cases}$$

Für die Folge $(f_n)_{n \in \mathbf{N}}$ auf $\Omega = [0, 1]$ lässt sich keine Majorante finden, denn, obwohl $f_n \rightarrow 0$, müsste jede solche Funktion g auch eine integrabale Majorante des punktweisen Supremums $h(\omega) = \sup_n f_n(\omega)$ sein. Allerdings ist

$$\int_{[0,1]} h d\lambda \geq \int_{(\frac{1}{m}, 1]} h d\lambda = \sum_{n=1}^{m-1} \int_{(\frac{1}{n+1}, \frac{1}{n}]} n d\lambda = \sum_{n=1}^{m-1} \frac{1}{n+1},$$

und der Ausdruck rechts divergiert für $m \rightarrow \infty$. Tatsächlich sieht man sofort, dass

$$1 = \lim_{n \rightarrow \infty} \int_{[0,1]} f_n d\lambda \neq \int f d\lambda = 0.$$

Die Jensensche Ungleichung

Wir betrachten nun konkret Maße mit der Eigenschaft $\mu(\Omega) = 1$. Es sei also $\mu = P$ ein Wahrscheinlichkeitsmaß auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{B}, P) und $X: \Omega \rightarrow \mathbf{R}$ eine Zufallsvariable (d.h. eine messbare Funktion).

Definition 1.42 (Erwartungswert einer Zufallsvariable) Der Erwartungswert von X ist definiert als

$$\mathbf{E}[X] = \int_{\Omega} X(\omega) dP(\omega) = \int_{\mathbf{R}} x dP_X(x), \quad (1.15)$$

wobei $P_X = P \circ X^{-1}$ das durch X induzierte Bildmaß ist.

Definition 1.43 (Konvexe Funktion) Eine Funktion $\phi: \mathbf{R} \rightarrow \mathbf{R}$ heißt *konvex*, wenn

$$\phi(\lambda y + (1 - \lambda)x) \leq \lambda\phi(y) + (1 - \lambda)\phi(x)$$

für alle $x, y \in \mathbf{R}$ und $\lambda \in [0, 1]$. Ist $-\phi$ konvex, so heißt ϕ *konkav*.

Anmerkung 1.44 Konvexe Funktionen müssen weder stetig noch differenzierbar sein.

Satz 1.45 (Jensen-Ungleichung) Sei $\phi: \mathbf{R} \rightarrow \mathbf{R}$ eine konvexe Funktion. Sind ϕ und $\phi \circ X$ integrierbar, so gilt

$$\phi(\mathbf{E}[X]) \leq \mathbf{E}[\phi(X)].$$

Beweis. Aus Lemma 1.27 wissen wir, dass die Ungleichung stimmt, wenn $\phi(x) = |x|$ die Betragsfunktion ist. Allgemein gibt es wegen der Eigenschaft konvexer Funktionen, subdifferenzierbar zu sein, stets Konstanten $a, b \in \mathbf{R}$, so dass $ax + b \leq \phi(x)$ für alle $x \in \mathbf{R}$ (siehe Abb. 2). Insbesondere gilt also

$$aX(\omega) + b \leq \phi(X(\omega)) \quad \forall \omega \in \Omega.$$

Die Konstanten a, b lassen sich so wählen, dass $am + b = \phi(m)$ für ein beliebiges $m \in \mathbf{R}$ gilt. Mit $m = \mathbf{E}[X]$ folgt aus der Linearität des Erwartungswertes, dass

$$\mathbf{E}[\phi(X)] \geq am + b = \phi(m) = \phi(\mathbf{E}[X])$$

und damit die Behauptung. □

Beispiel 1.46 Die Jensensche Ungleichung impliziert die Ungleichung von arithmetischem und geometrischem Mittel. Sei X eine endliche Zufallsvariable (d.h., X nimmt nur endlich viele, sagen wir N Werte an) und P_X das zugehörige Bildmaß $P_X(x_i) = P(\{\omega \in \Omega: X(\omega) = x_i\})$ mit $i = 1, \dots, N$. Dann ist

$$\mathbf{E}[X] = \sum_{i=1}^N p_i x_i, \quad p_i = P_X(x_i),$$

und wegen der Konkavität der Logarithmusfunktion (d.h. $-\log(\cdot)$ ist konvex) gilt

$$\log\left(\sum_{i=1}^N p_i x_i\right) \geq \sum_{i=1}^N p_i \log(x_i).$$

Exponentieren beider Seiten liefert

$$\sum_{i=1}^N p_i x_i \geq \exp\left(\sum_{i=1}^N p_i \log(x_i)\right) = \prod_{i=1}^N x_i^{p_i},$$

woraus für $p_i = 1/N$ die Ungleichung vom arithmetischen und geometrischen Mittel folgt:

$$\frac{1}{N} \sum_{i=1}^N x_i \geq \left(\prod_{i=1}^N x_i\right)^{1/N}.$$

Eine weitere wichtige Folgerung aus den Eigenschaften von konvexen Funktionen ist die Höldersche Ungleichung, die eine Verallgemeinerung der Cauchy-Schwarz-Ungleichung darstellt.

Korollar 1.47 (Hölder-Ungleichung) Seien $1 \leq p, q \leq \infty$ fest mit $1/p + 1/q = 1$ und $X, Y: \Omega \rightarrow \mathbf{R}$ zwei Zufallsvariable. Dann gilt

$$\mathbf{E}[|XY|] \leq (\mathbf{E}[|X|^p])^{1/p} (\mathbf{E}[|Y|^q])^{1/q}.$$

Beweis. Der Beweis mit Hilfe der Youngschen Ungleichung

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}, \quad x, y \geq 0,$$

bleibt dem Leser als einfache Übungsaufgabe überlassen. □

1.5 Produktmaße

Bei den wesentlichen Überlegungen zum Lebesgue-Integral haben wir uns bislang auf den eindimensionalen Fall beschränkt. Das soll sich nun ändern, indem wir – aufbauend auf den bisherigen Resultaten – die Integration über mehrere Variablen einführen. Dazu betrachten wir zwei Maßräume $(\Omega_1, \mathcal{B}_1, \mu_1)$ sowie $(\Omega_2, \mathcal{B}_2, \mu_2)$, von denen wir der Einfachheit halber annehmen, dass sie endlich sind. Ziel ist es, ein Maß μ über dem kartesischen Produkt

$$\Omega_1 \times \Omega_2 = \{(\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\} \quad (1.16)$$

zu definieren.

Definition 1.48 (Produkt- σ -Algebra) Als *Produkt- σ -Algebra* bezeichnen wir die von den Rechtecksmengen der Form $\{A_1 \times A_2 : A_1 \in \mathcal{B}_1, A_2 \in \mathcal{B}_2\}$ erzeugte σ -Algebra

$$\mathcal{B}_1 \otimes \mathcal{B}_2 := \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{B}_1, A_2 \in \mathcal{B}_2\}). \quad (1.17)$$

Man mache sich klar, dass $\{A_1 \times A_2\}$ zwar ein durchschnittstabiler Erzeuger, aber selbst keine σ -Algebra ist; vgl. die Konstruktion der n -dimensionalen Borelmengen und Satz 1.7.)

Ein geeigneter Kandidat für ein Produktmaß über dem Messraum $(\Omega, \mathcal{B}) := (\Omega_1 \times \Omega_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ ist das Maß, das ähnlich wie das Lebesguemaß den Rechtecksmengen das Produkt der Maße der Grundmengen zuordnet, d.h.,

$$\tilde{\mu}(A_1 \times A_2) := \mu_1(A_1)\mu_2(A_2) \quad (1.18)$$

bzw.

$$\tilde{\mu}\left(\bigcup_{i=1}^N A_1^i \times A_2^i\right) = \sum_{i=1}^N \mu_1(A_1^i)\mu_2(A_2^i) \quad (1.19)$$

für disjunkte endliche Vereinigungen. Für den weiteren Verlauf nehmen wir an, dass μ_1 und μ_2 beide endlich seien. Das folgende Lemma zeigt, dass dadurch bereits ein Maß bzw. Prämaß definiert ist.

Lemma 1.49 (Produktmaß auf Ring) *Sei \mathcal{E} der Mengenalgebra, der die Mengen $\{A_1 \times A_2 : A_1 \in \mathcal{B}_1, A_2 \in \mathcal{B}_2\}$ sowie ihre endlichen, disjunkten Vereinigungen enthält. Dann ist $\tilde{\mu}$ ein endliches Maß auf \mathcal{E} .*

Beweis. Da sich die meisten Eigenschaften wie z.B. die Endlichkeit direkt von μ_1, μ_2 auf $\mu_1 \otimes \mu_2$ vererben, besteht die wesentliche Schwierigkeit wie schon beim Beweis von Satz 1.18 darin, die σ -Additivität des Maßes, d.h., seine Stetigkeit von oben zu zeigen. Wir verfahren in 3 Schritten:

1. Sei $E \in \mathcal{E}$. Als *Schnitt von E* bezeichnen wir die Menge⁹

$$E_{\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in E\}.$$

Nach Konstruktion des Schnitts ist $E_{\omega_2} \in \mathcal{B}_1$ eine messbare Menge. Gleichermäßen ist die Funktion f , die durch $f(\omega_2) = \mu_1(E_{\omega_2})$ erklärt ist, als Funktion von ω_2 messbar. (Für Mengen $E = E_1 \times E_2$ ist $\mu_1(E_{\omega_2}) = \mu_1(E_1)\chi_{E_2}(\omega_2)$ eine Treppenfunktion, damit ist f messbar.) Nach Definition von $\tilde{\mu}$ gilt dann

$$\tilde{\mu}(E) = \int_{\Omega_2} \mu_1(E_{\omega_2}) d\mu_2(\omega_2).$$

2. Sei $E_1 \supseteq E_2 \supseteq \dots, E_n \in \mathcal{E}$ eine absteigende Kette mit $E_n \downarrow \emptyset$. Dann gilt auch, dass die Kette der

$$E_{n,\omega_2} := \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in E_n\}.$$

für alle $\omega_2 \in \Omega_2$ absteigend ist mit $E_{n,\omega_2} \downarrow \emptyset$. Nun ist μ_1 ein Maß, mithin stetig. Daher gilt

$$\lim_{n \rightarrow \infty} \mu_1(E_{n,\omega_2}) = 0.$$

⁹ Der Schnitt ist das Urbild der kanonischen Bündelprojektion $\pi_2 : \Omega_1 \times \Omega_2 \rightarrow \Omega_2, (\omega_1, \omega_2) \mapsto \omega_2$, eingeschränkt auf die Menge E (das ist eine stetige Abbildung). D.h., E_{ω_2} enthält alle Punkte (ω_1, ω_2) , die in E liegen und deren Projektion auf die zweite Komponente ω_2 ergibt; die obige Definition gilt analog für den Schnitt E_{ω_1} über Ω_1 .

3. Nach Satz 1.18 reicht es für die σ -Additivität, zu zeigen, dass $\tilde{\mu}$ stetig ist. Gemäß unserer Definition ist

$$\tilde{\mu}(E_n) = \int_{\Omega_2} \mu_1(E_{n,\omega_2}) d\mu_2(\omega_2),$$

und wir müssen zeigen, dass $\tilde{\mu}(E_n) \rightarrow 0$ für $n \rightarrow \infty$. Wegen der Endlichkeit von μ_1 und der Monotonie der Folge $(E_n)_{n \in \mathbb{N}}$ ist $(\mu_1(E_{n,\omega_2}))_{n \in \mathbb{N}}$ gleichmäßig beschränkt ist, d.h., es gibt eine Konstante $M < \infty$, so dass

$$0 \leq \mu_1(E_{n,\omega_2}) \leq M$$

für alle $n \in \mathbb{N}$ und $\omega_2 \in \Omega_2$. Aus dem Satz von der beschränkten Konvergenz, Satz 1.32, folgt somit

$$\lim_{n \rightarrow \infty} \tilde{\mu}(E_n) = \int_{\Omega_2} \lim_{n \rightarrow \infty} \mu_1(E_{n,\omega_2}) d\mu_2(\omega_2) = 0,$$

was zu beweisen war. □

Anmerkung 1.50 Die Aussage von Lemma 1.49 gilt gleichermaßen für σ -endliche Maße.

Nach dem Satz von Carathéodory für Halbringe (vgl. Satz 1.13) hat $\tilde{\mu}$ eine eindeutige, σ -additive Fortsetzung auf $\sigma(\mathcal{E})$ – die von \mathcal{E} erzeugte σ -Algebra. Die Fortsetzung wird *Produktmass* genannt und mit $\mu = \mu_1 \otimes \mu_2$ bezeichnet. Das n -dimensionale Produktmaß wird analog definiert:

Definition 1.51 (Produktmaß) Das Produktmass $\mu = \bigotimes_{i=1}^n \mu_i$ von n σ -endlichen Maßen μ_1, \dots, μ_n ist das eindeutige Maß auf der Produkt- σ -Algebra $\mathcal{B} = \mathcal{B}_1 \otimes \dots \otimes \mathcal{B}_n$, das auf dem System der Rechtecksmengen mit $\tilde{\mu}$ übereinstimmt, wobei $\tilde{\mu}$ die n -dimensionale Verallgemeinerung von (1.18)–(1.19) ist.

Integration in mehreren Variablen

Wie integriert man also bezüglich $\mu = \mu_1 \otimes \mu_2$? Vom Lebesgue-Maß wissen wir, dass

$$\lambda^2((a_1, b_1] \times (a_2, b_2]) = \int_{\mathbf{R}} \lambda^1((a_1, b_1] \chi_{(a_2, b_2]}(\omega_2)) d\lambda^1(\omega_2) = (b_1 - a_1)(b_2 - a_2).$$

Allgemein gilt die folgende Aussage.

Korollar 1.52 Für alle $B \in \mathcal{B} = \mathcal{B}_1 \otimes \mathcal{B}_2$ ist

$$\begin{aligned} \mu(B) &= \int_{\Omega_2} \left(\int_{\Omega_1} \chi_B(\omega_1, \omega_2) d\mu_1(\omega_1) \right) d\mu_2(\omega_2) \\ &= \int_{\Omega_1} \left(\int_{\Omega_2} \chi_B(\omega_1, \omega_2) d\mu_2(\omega_2) \right) d\mu_1(\omega_1). \end{aligned}$$

Beweis. Es reicht, die Behauptung für $\tilde{\mu} = \mu|_{\mathcal{E}}$ und die Rechtecksmengen $B = A_1 \times A_2$ nachzurechnen. □

Satz von Fubini

Unter praktischen Gesichtspunkten sagt Korollar 1.52 im wesentlichen alles, was wir beim Rechnen mit Mehrfachintegralen über die Vertauschbarkeit der Integrationsreihenfolge wissen müssen. Der berühmte Satz von Fubini präzisiert die Bedingungen, die eine zu integrierende Funktion erfüllen muss.¹⁰

Satz 1.53 (Fubini, 1907) Sei $f: \Omega_1 \times \Omega_2 \rightarrow \mathbf{R}$ messbar und mit f die Funktionen $g_{\omega_1}: \Omega_2 \rightarrow \mathbf{R}$ und $h_{\omega_2}: \Omega_1 \rightarrow \mathbf{R}$, die durch $g_{\omega_1}(\omega_2) = h_{\omega_2}(\omega_1) = f(\omega_1, \omega_2)$ definiert sind. Dann gilt:

1. Ist f integrierbar, so sind es auch g_{ω_1} und h_{ω_2} für fast alle ω_1 bzw. ω_2 ; ihre Integrale

$$G(\omega_1) = \int_{\Omega_2} g_{\omega_1} d\mu_2, \quad H(\omega_2) = \int_{\Omega_1} h_{\omega_2} d\mu_1$$

als Funktionen von ω_1 bzw. ω_2 sind messbar, fast überall endlich und integrierbar, und es gilt

$$\int_{\Omega} f d\mu = \int_{\Omega_1} G d\mu_1 = \int_{\Omega_2} H d\mu_2.$$

¹⁰ Guido Fubini (1879–1943), ital. Mathematiker

2. Ist $f \geq 0$ messbar und entweder H oder G integrierbar, so ist es auch die jeweils andere Funktion und

$$\int_{\Omega_1} G d\mu_1 = \int_{\Omega_2} H d\mu_2 = \int_{\Omega} f d\mu,$$

d.h., f ist integrierbar.

Beweis. Wir skizzieren den Beweis, der analog zur Konstruktion des Lebesgue-Integrals verläuft: Sei zunächst $B \in \mathcal{B}$ messbar und $f(\omega_1, \omega_2) = \chi_B(\omega_1, \omega_2)$. Für diesen Fall folgt die Aussage des Satzes aus Korollar 1.52. Wegen der Linearität des Lebesgue-Integrals gilt das gleichermaßen für Treppenfunktionen, und nach dem Satz von der beschränkten Konvergenz, Satz 1.32, auch für beschränkte, messbare Funktionen, die, wie wir gesehen haben, sich gleichmäßig durch Treppenfunktionen approximieren lassen. Für nicht-negative, messbare Funktionen, die sich als punktweise monotone Limites von Treppenfunktionen darstellen lassen, folgt die Aussage des Satzes sodann aus dem Satz von der monotonen Konvergenz, Satz 1.35, und schließlich für beliebige integrierbare Funktionen mit Hilfe der Zerlegung in negativen und positiven Teil, $f = f^+ - f^-$. \square

Das folgende Beispiel stammt aus [10].

Beispiel 1.54 (Integral als Fläche unter dem Graphen) Sei $f: \Omega \rightarrow [0, \infty)$ messbar auf $(\Omega, \mathcal{E}, \mu)$ und $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \lambda)$ der Lebesgue-Maßraum. Wir setzen $\nu = \mu \otimes \lambda$ und definieren

$$A := \{(\omega, y) \in \Omega \times [0, \infty) : 0 \leq y \leq f(\omega)\}$$

als die Menge aller Punkte unter dem Graphen von f . Wie man leicht klarmacht, sind

$$f(\omega) = \int_{[0, \infty)} \chi_A(\omega, y) d\lambda(y)$$

und

$$\mu(\{\omega \in \Omega : f(\omega) \geq y\}) = \int_{\Omega} \chi_A(\omega, y) d\mu(\omega).$$

Damit folgt aus dem Satz von Fubini, dass die Fläche unter dem Graphen das Integral über die Funktion ist:

$$\nu(A) = \int_{\Omega} f d\mu = \int_{[0, \infty)} \mu(\{f \geq y\}) d\lambda(y).$$

2 Summen von unabhängigen Zufallsvariablen

Es sei (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbf{R}$ eine reellwertige Zufallsvariable (d.h., eine messbare Funktion). Der durch X induzierte Wahrscheinlichkeitsraum ist durch $(\mathbf{R}, \mathcal{B}(\mathbf{R}), P_X)$ gegeben, wobei $P_X = P \circ X^{-1}$ das Bildmaß von P unter der Abbildung X bezeichnet. Für Wahrscheinlichkeiten bezüglich P bzw. P_X werden wir je nach Kontext die folgenden Schreibweisen verwenden:

$$P_X(A) = P(X \in A) = P(\{\omega \in \Omega: X(\omega) \in A\}), \quad A \in \mathcal{B}(\mathbf{R}).$$

Wegen der Messbarkeit von X ist $X^{-1}(A)$ für alle Borelmengen $A \in \mathcal{B}(\mathbf{R})$ ein Element aus \mathcal{E} und damit wieder eine messbare Menge, der sich eindeutig eine Wahrscheinlichkeit $P(X \in A)$ zuordnen lässt.

2.1 Wahrscheinlichkeitsverteilungen und -dichten

Die folgende Definition haben wir in ähnlicher Form bereits in Satz 1.18 kennengelernt.

Definition 2.1 (Verteilungsfunktion) Eine Funktion $F: \mathbf{R} \rightarrow [0, 1]$ mit der Eigenschaft

$$F(x) = P(\{\omega \in \Omega: X(\omega) \leq x\}) = P_X((-\infty, x]) \quad (2.1)$$

heißt *Verteilungsfunktion von X* .

Aus der Monotonie und Stetigkeit von P bzw. P_X von oben folgt sofort, dass F monoton wachsend und rechtsseitig stetig ist. Wegen der Stetigkeit von P bzw. P_X von unten gilt außerdem, dass

$$\lim_{x \rightarrow \infty} F(x) = \lim_{x \rightarrow \infty} P_X((-\infty, x]) = 1.$$

Definition 2.1 ist folglich äquivalent mit der Definition der maßerzeugenden Funktion aus Satz 1.18 auf S. 10. Umgekehrt besagt Satz 1.18, dass es zu jeder Verteilungsfunktion genau ein Wahrscheinlichkeitsmaß P gibt, das von F erzeugt wird. Ein Spezialfall sind die stetigen Verteilungsfunktionen:

Lemma 2.2 *Es sei (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und $X: \Omega \rightarrow \mathbf{R}$ messbar. Dann gilt*

$$P_X(\{x\}) = 0, \quad \forall x \in \mathbf{R}$$

genau dann, wenn die Verteilungsfunktion $F(x) = P_X((-\infty, x])$ stetig ist.

Beweis. Sei $A_n = (x_n, x]$ mit $x_n \uparrow x$. Dann gilt $A_n \downarrow \{x\}$ und wegen der Stetigkeit von oben ist

$$P_X(\{x\}) = \lim_{n \rightarrow \infty} P_X(A_n) = \lim_{n \rightarrow \infty} |F(x) - F(x_n)|,$$

womit die Behauptung bewiesen ist. □

Definition 2.3 (Randverteilungen) Sei $X: \Omega \rightarrow \mathbf{R}^n$, $X = (X_1, \dots, X_n)$ eine vektorwertige Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) .

1. $P_X = P \circ X^{-1}$ heißt *gemeinsame Verteilung von X_1, \dots, X_n* .
2. Sei $\pi_i: \mathbf{R}^n \rightarrow \mathbf{R}$, $(x_1, \dots, x_n) \mapsto x_i$ die Projektion auf die i -te Komponente von X . die $P_X \circ \pi_i^{-1} = P \circ X_i^{-1}$, $i = 1, \dots, n$ werden *Randverteilungen von X* genannt.

In Anwendungen taucht oft das Problem auf, dass man Zufallszahlen zu einer vorgegebenen Verteilung erzeugen möchte (z.B. auf einem Computer, der nur uniform verteilte Zufallszahlen erzeugen kann).

Beispiel 2.4 (Inversionsmethode I) Sei $\Omega = [0, 1]$ und $P = \chi_{[0,1]} \lambda$ das Wahrscheinlichkeitsmaß der stetigen Gleichverteilung auf Ω (d.h., das auf Ω eingeschränkte Lebesgue-Maß). Wir wollen zu einer gegebenen Verteilungsfunktion $F: \mathbf{R} \rightarrow [0, 1]$ eine Zufallsvariable $X: [0, 1] \rightarrow \mathbf{R}$ finden, die F als Verteilungsfunktion hat.

- Zunächst definieren wir eine verallgemeinerte Umkehrfunktion F^{-1} durch

$$F^{-1}(u) = \inf\{x \in \mathbf{R}: F(x) \geq u\}.$$

Ist F stetig und streng monoton steigend, so entspricht $F^{-1}(\cdot)$ dem lokalen Auflösen nach x . (Eine "klassische" Umkehrfunktion existiert i.A. nicht, denn F muss weder stetig noch streng monoton sein.)

- Als Zufallsvariable definieren wir $X(\omega) = F^{-1}(\omega)$. Somit gilt

$$\begin{aligned} P_X((-\infty, x]) &= P(\{\omega \in \Omega: X(\omega) \leq x\}) \\ &= P(\{\omega \in \Omega: F^{-1}(\omega) \leq x\}) \\ &= P(\{\omega \in \Omega: \omega \leq F(x)\}) \\ &= F(x), \end{aligned}$$

wobei wir in der vorletzten Zeile die Monotonie von F und in der letzten Zeile die Tatsache, dass $\omega \in [0, 1]$ gleichverteilt ist, ausgenutzt haben. Also ist X nach F verteilt.

- Zum Erzeugen F -verteilter, unabhängiger Zufallszahlen x_1, \dots, x_n , erzeugt man sich auf dem Intervall $[0, 1]$ uniform verteilte, unabhängige Zufallszahlen u_1, \dots, u_n und setzt $x_k = F^{-1}u_k$, $k = 1, \dots, n$. Die Unabhängigkeit der x_k folgt dabei aus der Messbarkeit der Verteilungsfunktion F .

Das folgende Beispiel ist (traurige) Geschichte: Stanislaw Ulam und John von Neumann, zwei der Pioniere der Monte-Carlo-Methode, verwendeten das Verfahren bei der Entwicklung der Atombombe am *Los Alamos National Laboratory* zur Simulation der Neutronenstreuung (einen historischen Abriss findet man in [12]).

Beispiel 2.5 (Inversionsmethode II) Sei $([0, 1], \mathcal{B}[0, 1], \chi_{[0,1]}\lambda)$ der Wahrscheinlichkeitsraum der auf $[0, 1]$ uniform verteilten Zufallsvariablen $U \sim \mathcal{U}(0, 1)$. Dann ist die Zufallsvariable

$$X = -\lambda^{-1} \log(1 - U)$$

exponentialverteilt mit Parameter $\lambda > 0$ (gedächtnislose Wartezeit); die zugehörige Verteilungsfunktion ist

$$F: [0, \infty) \rightarrow [0, 1], \quad F(x) = 1 - \exp(-\lambda x).$$

Wahrscheinlichkeitsdichten

Satz 2.6 (Dichtefunktion) Ist $(\Omega, \mathcal{E}, \mu)$ ein Maßraum und $f: \Omega \rightarrow [0, \infty)$ eine messbare Funktion, so definiert

$$\nu: \mathcal{E} \rightarrow [0, \infty], \quad E \mapsto \int_{\Omega} f \chi_E d\mu = \int_E f d\mu$$

ein Maß auf \mathcal{E} ; ν ist genau dann ein Wahrscheinlichkeitsmaß, wenn $\int_{\Omega} f d\mu = 1$ ist.

Beweis. ÜA. □

Lässt sich ein Maß ν durch eine messbare Funktion $f \geq 0$ darstellen, so sprechen wir von f als *Dichtefunktion* von ν bezüglich μ , symbolisch $f = d\nu/d\mu$.¹¹

Beispiel 2.7 Es folgen eine Reihe von Beispielen zu Dichtefunktion und Randverteilungen.

1. *Zähldichte*: Sei Ω abzählbar und $\mathcal{E} \subset 2^{\Omega}$. Ist mit $(f_{\omega})_{\omega \in \Omega}$ eine Familie nicht-negativer Zahlen gegeben, die sich zu 1 aufsummieren, d.h. $\sum_{\omega} f_{\omega} = 1$, so wird durch

$$P: \mathcal{E} \rightarrow [0, \infty], \quad E \mapsto \sum_{\omega \in E} f_{\omega}$$

ein diskretes Wahrscheinlichkeitsmaß auf \mathcal{E} definiert. Die Familie f_{ω} heißt *Zähldichte* von P ; sie ist eine Dichte bezüglich des Zählmaßes μ , das durch $\mu(A) = |A|$ für alle $A \subset \Omega$ erklärt ist. Ferner gilt

$$P(E) = \int_{\Omega} f_{\omega} \chi_E(\omega) d\mu(\omega) = \sum_{\omega \in E} f_{\omega}.$$

2. *Stetige Gleichverteilung*: Sei $\Omega = [a, b] \subset \mathbf{R}$ und $\mathcal{E} = \mathcal{B}([a, b])$. Dann ist $f = (b - a)^{-1} \chi_{[a, b]}$ die Dichte der Gleichverteilung bezüglich Lebesgue-Maß λ .

¹¹ Der Ausdruck $d\nu/d\mu$ wird Radon-Nikodym-Ableitung genannt und in Kapitel 3 präzise definiert.

3. *Randverteilung:* Sei (Ω, \mathcal{E}, P) ein Wahrscheinlichkeitsraum und P_X die gemeinsame Verteilung der Zufallsvariablen $X_1(\omega), X_2(\omega)$ mit Dichtefunktion $f(x_1, x_2)$ bezüglich $\lambda = \lambda_1 \otimes \lambda_2$. Dann ist

$$f_1(x_1) = \int_{\mathbf{R}} f \, d\lambda_2 = \int_{\mathbf{R}} f(x_1, x_2) \, dx_2$$

die Dichte der Randverteilung

$$(P_X \circ \pi_1^{-1})((-\infty, x_1]) = \lim_{x_2 \rightarrow \infty} P(X_1 \leq x_1, X_2 \leq x_2)$$

in x_1 bezüglich λ_1 .

4. *Diracmaß:* Das Punktmaß δ_x über $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ hat keine Lebesgue-Dichte; für alle $B \in \mathcal{B}(\mathbf{R})$ gilt vielmehr:

$$\int_B f \, d\delta_x = f(x)\delta_x(B) = \begin{cases} f(x), & x \in B \\ 0 & \text{sonst.} \end{cases}$$

Satz 2.8 Ist P_X ein Wahrscheinlichkeitsmaß auf $\mathcal{B}(\mathbf{R})$ mit der Dichte $f = dP_X/d\lambda$, so gilt

$$F(x) = P_X((-\infty, x]) = \int_{(-\infty, x]} f \, d\lambda = \int_{-\infty}^x f(z) \, dz.$$

Ist umgekehrt P_X ein Wahrscheinlichkeitsmaß mit differenzierbarer Verteilungsfunktion F , so ist $f = F'$ Dichte von P_X .

Beweis. Die erste Aussage folgt direkt aus der Definition der Wahrscheinlichkeitsdichte, die zweite aus dem Hauptsatz der Differential- und Integralrechnung. \square

2.2 Verteilungskonvergenz (schwache Konvergenz)

In diesem Abschnitt werden wir die (schwache) Konvergenz von Zufallsvariablen einführen, die nur von ihren Verteilungen abhängt, und die unter anderem dem Zentralen Grenzwertsatz zugrundeliegt. Ziel ist es, die schwache Konvergenz von Zufallsvariablen auf die punktweise Konvergenz ihrer Verteilungsfunktion bzw. ihrer durch die Verteilungsfunktion eindeutig bestimmten charakteristischen Funktion zurückzuführen.

Erwartungswert und Momente

Definition 2.9 (Erwartungswert) Die Abbildung $X \mapsto \mathbf{E}(X)$, die durch

$$\mathbf{E}[X] = \int_{\Omega} X(\omega) \, dP(\omega) = \int_{\mathbf{R}} x \, dP_X(x). \quad (2.2)$$

definiert ist, wird *Erwartungswert von X* genannt.

Man kann nun das Spiel mit Maß und Bildmaß wiederholen und sich leicht klarmachen, dass der Erwartungswert für eine neue Zufallsvariable $Y = g(X)$ gerade durch

$$\mathbf{E}[Y] = \mathbf{E}[g(X)] = \int_{\mathbf{R}} g \, dP_X.$$

gegeben sein muss ($P_X \circ g^{-1}$ ist die Verteilung von Y). Das motiviert die nächste Definition:

Definition 2.10 (Momente, zentrierte Momente) Das k -te Moment von X ist durch

$$M_k(X) = \mathbf{E}[X^k] = \int_{\mathbf{R}} x^k \, dP_X(x). \quad (2.3)$$

definiert; als *zentriertes k -tes Moment von X* bezeichnen wir

$$C_k(X) = \mathbf{E}[(X - \mathbf{E}(X))^k] = \int_{\mathbf{R}} (x - \mathbf{E}[X])^k \, dP_X(x). \quad (2.4)$$

Definition 2.11 (Varianz) Das zweite zentrierte Moment $\mathbf{V}[X] := C_2(X)$ wird *Varianz* genannt.

Eine unmittelbare Folgerung aus der Definition des Erwartungswertes ist seine Linearität. (Den Beweis überlassen wir der LeserIn als Übungsaufgabe.)

Korollar 2.12 (Linearität des Erwartungswertes) Seien X_1, \dots, X_n reelle Zufallsvariable auf einem gemeinsamen Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) und $\alpha_1, \dots, \alpha_n$ reelle Zahlen. Dann gilt

$$\mathbf{E}[\alpha_1 X_1 + \dots + \alpha_n X_n] = \alpha_1 \mathbf{E}[X_1] + \dots + \alpha_n \mathbf{E}[X_n].$$

Charakteristische Funktion

Definition 2.13 (Charakteristische Funktion) Es sei $X: \Omega \rightarrow \mathbf{R}$ eine reellwertige Zufallsvariable mit Verteilung P_X . Die Funktion $\varphi_X: \mathbf{R} \rightarrow \mathbf{C}$, die durch

$$\varphi_X(s) = \mathbf{E}[e^{isX}] = \int_{\mathbf{R}} e^{isx} dP_X(x) \quad (2.5)$$

erklärt ist, heißt *charakteristische Funktion von X*.

Satz 2.14 φ_X ist gleichmäßig stetig, beschränkt und positiv semi-definit, d.h., für beliebige komplexe Zahlen ξ_1, \dots, ξ_n und reelle Zahlen s_1, \dots, s_n ($n \in \mathbf{N}$ beliebig) gilt¹²

$$\sum_{i,j=1}^n \xi_i \bar{\xi}_j \varphi_X(s_j - s_i) \geq 0.$$

Beweis. Aus der Identität $e^{iz} = \cos(z) + i \sin(z)$ folgt sofort, dass $|e^{iz}| = 1$ für alle $z \in \mathbf{R}$, so dass wir mit der Monotonie des Erwartungswertes

$$|\varphi_X(s)| = |\mathbf{E}[e^{isX}]| \leq \mathbf{E}[|e^{isX}|] = 1$$

erhalten. Dass φ_X gleichmäßig stetig ist, sieht man wie folgt: Es gilt

$$\begin{aligned} |\varphi_X(t) - \varphi_X(s)| &\leq \mathbf{E}[|e^{itX} - e^{isX}|] \\ &= \mathbf{E}[|e^{i(t-s)X} e^{isX} - e^{isX}|] \\ &= \mathbf{E}[|e^{i(t-s)X} - 1|]. \end{aligned}$$

Da der Ausdruck unter dem Erwartungswert in der letzten Zeile beschränkt ist, folgt aus dem Satz von der beschränkten Konvergenz, Satz 1.32, dass

$$\lim_{s \rightarrow t} |\varphi_X(t) - \varphi_X(s)| = \mathbf{E}\left[\lim_{s \rightarrow t} |e^{i(t-s)X} - 1|\right] = 0,$$

wobei der Ausdruck hinter dem Limes auf der rechten Seite für alle $s, t \in \mathbf{R}$ mit $|t - s| < \epsilon$ gleichmäßig durch $M\epsilon$ mit $0 < M < \infty$ beschränkt ist. Die Positivität folgt schließlich aus der Linearität des Erwartungswertes:

$$\sum_{i,j=1}^n \xi_i \bar{\xi}_j \varphi_X(s_j - s_i) = \sum_{i,j=1}^n \xi_i \bar{\xi}_j \mathbf{E}[e^{i(s_j - s_i)X}] = \mathbf{E}\left[\left|\sum_{i=1}^n \xi_i e^{is_i X}\right|^2\right] \geq 0.$$

□

Die Bedeutung der charakteristischen Funktion φ_X erklärt sich dadurch, dass sie alle Informationen über P_X enthält, so dass sich viele Aussagen über P_X auf die charakteristische Funktion zurückführen lassen (dazu später mehr). Sind alle Momente von X endlich, so lässt sich φ_X als Potenzreihe darstellen

$$\varphi_X(s) = \sum_{n=1}^{\infty} \frac{(is)^n}{n!} \mathbf{E}[X^n] = 1 + is\mathbf{E}[X] - \frac{s^2}{2}\mathbf{E}[X^2] + \dots$$

¹² Mit $\bar{\xi}$ bezeichnen wir die komplex Konjugierte zu $\xi \in \mathbf{C}$, d.h., $\bar{\xi} = \Re(\xi) - i\Im(\xi)$.

Beispiel 2.15 Für die meisten gebräuchlichen Wahrscheinlichkeitsverteilungen lässt sich φ_X direkt ausrechnen: Hier sind einige Beispiele:

1. Stetige Gleichverteilung auf $[a, b]$:

$$\varphi_X(s) = \frac{e^{isb} - e^{isa}}{is(b-a)}.$$

2. Normalverteilung $\mathcal{N}(\mu, \sigma^2)$:

$$\varphi_X(s) = e^{is\mu - \sigma^2 s^2 / 2}.$$

3. Poissonverteilung zum Parameter $\lambda > 0$:

$$\varphi_X(s) = e^{\lambda(e^{is} - 1)}.$$

Anmerkung 2.16 Nicht immer existieren alle Momente einer Zufallsvariable X . Ist jedoch $\mathbf{E}[|X|] < \infty$, so ist φ_X stetig differenzierbar und es gilt, dass $\mathbf{E}[X] = -i\varphi_X'(0)$. Umgekehrt kann φ_X durchaus stetig differenzierbar sein, ohne dass $\mathbf{E}[|X|]$ und damit der Erwartungswert existieren. Allgemein lässt sich zeigen, dass aus der Existenz des k -ten Moments folgt, dass die Funktion φ_X k -mal stetig differenzierbar ist und

$$\mathbf{E}[X^k] = -i^k \varphi_X^{(k)}(0).$$

Die Umkehrung der Aussage gilt nur, wenn k gerade ist; für ungerade k ist die Umkehrung falsch.¹³

Zentrale Fragen

Wir werden nun untersuchen,

1. ob die Verteilungsfunktion F eindeutig durch die charakteristische Funktion φ_X bestimmt ist und
2. wie die Konvergenz einer Folge $(X_n)_{n \in \mathbf{N}}$ mit der punktweisen Konvergenz von φ_{X_n} zusammenhängt.

Im nächsten Abschnitt, Abschnitt 2.3, werden wir die Überlegungen auf Summen S_n von unabhängigen Zufallsvariablen X_1, \dots, X_n übertragen, woraus sich dann leicht Aussagen wie der Zentrale Grenzwertsatz oder das schwache Gesetz der großen Zahlen beweisen lassen. Das genaue Verständnis des Zusammenhangs zwischen Verteilungsfunktion und charakteristischer Funktion ist der erste Schritt in diese Richtung.

1. Frage: Zusammenhang von Verteilungsfunktion und charakteristischer Funktion

Um $F(x) = P_X((-\infty, x])$ aus φ_X zu berechnen, nehmen wir zunächst an, dass P_X eine Dichtefunktion $f = dP_X/dx$ hat. In diesem Fall ist

$$\varphi_X(s) = \int_{-\infty}^{\infty} e^{isx} f(x) dx$$

die Fouriertransformierte von f . Für integrierbare Funktionen f ist die Fouriertransformierte eine stetige Funktion, die im Unendlichen verschwindet.¹⁴ Die Umkehrung der Fouriertransformation liefert eine Darstellung von f :

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-isx} \varphi_X(s) ds$$

Auch dieses Integral lässt sich berechnen, indem man Real- und Imaginärteil getrennt integriert. Wir können nun versuchen, einen allgemeinen Zusammenhang zwischen Verteilungsfunktion und charakteristischer Funktion zu erraten. Da $f(x) = F'(x)$ ist, drängt sich die folgende Darstellung auf:

$$F(b) - F(a) = \frac{1}{2\pi} \int_a^b \int_{-\infty}^{\infty} e^{-isx} \varphi_X(s) ds dx.$$

¹³ ÜA.

¹⁴ Für die genaue Definition der Fouriertransformation und ihre Eigenschaften sei auf [17, Kap. V.2] verwiesen.

Nach dem Satz von Fubini können wir die Integrationsreihenfolge vertauschen, so dass

$$\begin{aligned} F(b) - F(a) &= \frac{1}{2\pi} \int_a^b \left[\int_{-\infty}^{\infty} e^{-isx} \varphi_X(s) ds \right] dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(s) \left[\int_a^b e^{-isx} dx \right] ds \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(s) \frac{e^{-isb} - e^{-isa}}{-is} ds \\ &= \lim_{S \rightarrow \infty} \frac{1}{2\pi} \int_{-S}^S \varphi_X(s) \frac{e^{-isa} - e^{-isb}}{is} ds. \end{aligned}$$

Wir behaupten nun, dass die Annahme, dass P_X eine Dichtefunktion hat, für den soeben hergeleiteten Zusammenhang zwischen charakteristischer Funktion und Verteilungsfunktion unerheblich ist.

Behauptung 2.17 Für alle $a, b \in \mathbf{R}$, an denen F stetig ist, gilt

$$F(b) - F(a) = \lim_{S \rightarrow \infty} \frac{1}{2\pi} \int_{-S}^S \varphi_X(s) \frac{e^{-isa} - e^{-isb}}{is} ds.$$

Beweis. Nach Definition der charakteristischen Funktion und dem Satz von Fubini gilt

$$\begin{aligned} \lim_{S \rightarrow \infty} \frac{1}{2\pi} \int_{-S}^S \varphi_X(s) \frac{e^{-isa} - e^{-isb}}{is} ds &= \lim_{S \rightarrow \infty} \frac{1}{2\pi} \int_{-S}^S \left[\int_{\mathbf{R}} e^{isx} dP_X \right] \frac{e^{-isa} - e^{-isb}}{is} ds \\ &= \lim_{S \rightarrow \infty} \frac{1}{2\pi} \int_{\mathbf{R}} \left[\int_{-S}^S \frac{e^{is(x-a)} - e^{is(x-b)}}{is} ds \right] dP_X. \end{aligned}$$

Das innere Integral in der letzten Zeile ist ein Cauchyscher Hauptwert und lässt sich getrennt nach Real- und Imaginärteil integrieren. Da der Hauptwert beschränkt ist, können wir den Limes $S \rightarrow \infty$ nach innen ziehen (das ist gewissermaßen der Satz von der beschränkten Konvergenz rückwärts gelesen); damit erhalten wir

$$\begin{aligned} \lim_{S \rightarrow \infty} \frac{1}{2\pi} \int_{\mathbf{R}} \left[\int_{-S}^S \frac{e^{is(x-a)} - e^{is(x-b)}}{is} ds \right] dP_X &= \lim_{S \rightarrow \infty} \frac{1}{2\pi} \int_{\mathbf{R}} \left[\int_{-S}^S \frac{\sin(s(x-a)) - \sin(s(x-b))}{s} ds \right] dP_X \\ &= \frac{1}{2} \int_{\mathbf{R}} (\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b)) dP_X, \end{aligned}$$

wobei die zweite Gleichung aus der Definition des Integralsinus folgt:

$$\lim_{S \rightarrow \infty} \int_0^S \frac{\sin(sx)}{s} ds = \begin{cases} -\pi/2, & x < 0 \\ 0, & x = 0 \\ \pi/2, & x > 0. \end{cases}$$

Da a und b nach Voraussetzung Stetigkeitsstellen von F sind, ist

$$\frac{1}{2} \int_{\mathbf{R}} (\operatorname{sgn}(x-a) - \operatorname{sgn}(x-b)) dP_X = P_X((-\infty, b]) - P_X((-\infty, a]) = F(b) - F(a),$$

womit die Behauptung bewiesen wäre. \square

2. Frage: Impliziert Konvergenz von φ_X die Konvergenz in Verteilung?

Wir betrachten eine Folge $(X_n)_{n \in \mathbf{N}}$ von reellwertigen Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) , zu denen eine Folge von induzierten Verteilungen μ_n auf $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ gehört. Um zu verstehen, was *Konvergenz in Verteilung* bedeutet, müssen wir uns zunächst klarmachen, was es für zwei Maße μ, ν heißt, dass $\mu \approx \nu$.

Definition 2.18 (TV-Norm) Für zwei Wahrscheinlichkeitsmaße μ, ν auf $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ heißt

$$\|\mu - \nu\|_{\text{TV}} = \sup_{B \in \mathcal{B}(\mathbf{R})} |\mu(B) - \nu(B)|$$

der *totale Variationsabstand* zwischen μ und ν .

Die starke Konvergenz $\mu_n \rightarrow \mu$ lässt sich nun ganz einfach als Konvergenz in der Norm definieren. Ein typisches Beispiel für starke Konvergenz auf einem abzählbaren Messraum $(\mathbf{N}, 2^{\mathbf{N}})$ ist die Konvergenz der Binomialverteilung $\mu_n = B(k; n, p)$ gegen die Poissonverteilung $\mu = P(k; \lambda)$ für $n \rightarrow \infty$, $p \rightarrow 0$ und $np \rightarrow \lambda$.

Offensichtlich ist $0 \leq \|\mu - \nu\|_{\text{TV}} \leq 1$, und es ist gewissermaßen Folklore, dass die TV-Norm fast immer 1 ist; zum Beispiel ist $\|\delta_x - \delta_y\|_{\text{TV}} = 1$ für Diracmaße δ_x, δ_y mit $x \neq y$ – unabhängig davon wie klein $|x - y|$ ist. Damit ist die TV-Norm für unsere Zwecke ungeeignet. Da wir insbesondere an Folgen von stetigen Verteilungen interessiert sind, die sich an einzelnen Punkten konzentrieren, sind selbst “punktweise” Konvergenzbegriffe á la “ $\mu_n(B) \rightarrow \mu(B)$ für alle Borelmengen $B \in \mathcal{B}(\mathbf{R})$ ” ungeeignet (siehe Beispiel 2.20 unten).

Andererseits haben wir mit dem Erwartungswert, Definition 2.9, ein lineares Funktional gegeben, das für Punktmaße und Maße mit Dichtefunktion auf einheitliche Art und Weise stetigen und beschränkten Funktionen reelle Zahlen zuordnet. Das motiviert die folgende Definition (vgl. [2, Kap. 4]).

Definition 2.19 (Schwache Konvergenz, Konvergenz in Verteilung) Sei $(\mu_n)_{n \in \mathbf{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Wir sagen, μ_n konvergiere schwach gegen μ (symbolisch: $\mu_n \rightharpoonup \mu$), wenn

$$\int_{\mathbf{R}} f d\mu_n \rightarrow \int_{\mathbf{R}} f d\mu \quad \forall f \in C_b(\mathbf{R}),$$

wobei $C_b(\mathbf{R})$ der Raum der beschränkten und stetigen Funktionen auf \mathbf{R} ist. Eine Folge $(X_n)_{n \in \mathbf{N}}$ von Zufallsvariablen auf einer Familie von Wahrscheinlichkeitsräumen $(\Omega_n, \mathcal{E}_n, P_n)$ heißt *konvergent in Verteilung* (symbolisch: $X_n \xrightarrow{i.V.} X$), wenn die Folge ihrer Verteilungen P_{X_n} schwach gegen P_X konvergiert, d.h. wenn

$$\mathbf{E}[f(X_n)] \rightarrow \mathbf{E}[f(X)] \quad \forall f \in C_b(\mathbf{R}).$$

Beispiel 2.20 Wir betrachten zwei typische Fälle von Wahrscheinlichkeitsmaßen mit Dichte, deren schwache Grenzwerte Punktmaße sind.

1. Es sei $X_n \sim \mathcal{N}(a, n^{-2})$ normalverteilt mit $a = \mathbf{E}[X_n]$ und $n^{-2} = \mathbf{E}[(X_n - a)^2]$ und der Lebesguedichte

$$f_n(x) = \frac{n}{\sqrt{2\pi}} e^{-\frac{n^2}{2}(x-a)^2}.$$

Da f_n für $n \rightarrow \infty$ nur an der Stelle $x = a$ von 0 verschieden ist, konzentriert sich dort die gesamte Wahrscheinlichkeit, d.h.,

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} h(x) f_n(x) dx = h(a)$$

für alle stetigen und beschränkten Funktionen h . Folglich konvergiert $\mu_n = f_n \lambda$ schwach gegen das Diracmaß bei a , d.h. $\mu_n \rightharpoonup \delta_a$. Dass μ_n nicht auch stark konvergiert, sieht man, indem man sich klarmacht, dass $\mu_n(\{a\}) = 0$ für alle $n \in \mathbf{N}$ ist, da Punktmengen stets Lebesgue-Maß 0 haben.

2. Wir betrachten eine Folge von Wahrscheinlichkeitsmaßen auf $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, die durch $\mu_n(B) = \chi_B(1/n)$ für alle $B \in \mathcal{B}(\mathbf{R})$ erklärt sind. Nun gilt $1/n \rightarrow 0$, und man sieht leicht ein, dass $\mu_n \rightharpoonup \delta_0$. Andererseits konvergiert μ_n nicht stark, sprich: nicht für alle Borelmengen $B \in \mathcal{B}(\mathbf{R})$; das sieht man, indem man beispielsweise $B = (-\infty, 0]$ wählt – in diesem Fall ist $\mu_n(B) = 0$ für alle $n \in \mathbf{N}$, aber $\delta_0(B) = 1$.

Satz 2.21 (Lévy-Cramérscher Stetigkeitssatz¹⁵) Sei $(\mu_n)_{n \in \mathbf{N}}$ eine Folge von Wahrscheinlichkeitsmaßen auf dem Messraum $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Dann sind folgende Aussagen äquivalent:

1. $F_n(x) \rightarrow F(x)$ an allen Stetigkeitsstellen von F ,
2. $\mu_n \rightharpoonup \mu$,
3. $\varphi_{X_n} \rightarrow \varphi_X$ punktweise auf ganz \mathbf{R} .

Beweis. Es gilt $F_n(x) = \mu_n((-\infty, x])$ und $\varphi_{X_n}(s) = \int_{\mathbf{R}} e^{ixs} d\mu_n$ und analog für $F(x)$ sowie $\varphi_X(s)$. Wir beweisen die Implikationen der Reihe nach.

¹⁵ Paul Lévy (1886–1971), frz. Mathematiker; Harald Cramér (1893–1985), schwed. Mathematiker

1. \Rightarrow 2. Sei $\epsilon > 0$. Ferner seien $a, b \in \mathbf{R}$ Stetigkeitsstellen mit $a < b$ und $F(a) \leq \epsilon$ und $1 - F(b) \leq \epsilon$. Da $F_n(x) \rightarrow F(x)$ für alle x , an denen F stetig ist, existiert ein $m = m(\epsilon) \in \mathbf{N}$ mit $F_m(a) \leq 2\epsilon$ und $1 - F_m(b) \leq 2\epsilon$.

Sei nun $f \in C_b(\mathbf{R})$ eine Testfunktion. Für ein beliebiges $\delta > 0$ zerlegen wir das Intervall $(a, b]$ in N disjunkte Teilintervalle $I_j = (a_j, a_{j+1}]$ mit $a = a_1 < a_2 \dots < a_{N+1} = b$, wobei die a_j allesamt Stetigkeitsstellen sind und $N = N(\delta)$ so gewählt ist, dass die Differenz zwischen dem größten und kleinsten Wert, den f auf jedem der I_j annimmt, höchstens δ ist; letzteres ist immer möglich, da f nach Voraussetzung auf ganz \mathbf{R} stetig und damit auf allen beschränkten Intervallen $(a, b]$ gleichmäßig stetig ist.

Mit der Treppenfunktion

$$g(x) = \sum_{j=1}^N f(a_j) \chi_{I_j}(x)$$

definieren wir eine gleichmäßige Approximation von f auf $(a, b]$, für die gilt, dass $\sup_{x \in (a, b]} |f(x) - g(x)| \leq \delta$, und die auf $(a, b]^c = \mathbf{R} \setminus (a, b]$ verschwindet (d.h. $g|_{(a, b]^c} = 0$). Da f beschränkt ist, existiert eine Konstante $M < \infty$ mit $\sup_{x \in \mathbf{R}} |f(x)| \leq M$. Mit Hilfe der Dreiecksungleichung erhalten wir die Abschätzungen

$$\begin{aligned} \left| \int_{\mathbf{R}} f d\mu_m - \sum_{j=1}^N f(a_j)(F_m(a_{j+1}) - F_m(a_j)) \right| &= \left| \int_{(a, b]} (f - g) d\mu_m + \int_{(-\infty, a]} f d\mu_m + \int_{(b, \infty)} f d\mu_m \right| \\ &\leq \delta + MF_m(a) + M(1 - F_m(b)) \\ &\leq \delta + 4M\epsilon. \end{aligned}$$

bzw.

$$\left| \int_{\mathbf{R}} f d\mu - \sum_{j=1}^N f(a_j)(F(a_{j+1}) - F(a_j)) \right| \leq \delta + 2M\epsilon.$$

Wegen $F_n(a_j) \rightarrow F(a_j)$ für alle $j = 1, \dots, N$ ergibt sich nun mit der Dreiecksungleichung:

$$\limsup_{n \rightarrow \infty} \left| \int_{\mathbf{R}} f d\mu_n - \int_{\mathbf{R}} f d\mu \right| \leq \delta + 6M\epsilon.$$

Da $\epsilon, \delta > 0$ beliebig sind, folgt die Behauptung.

2. \Rightarrow 3. Es gilt $\varphi_{X_n}(s) = \int_{\mathbf{R}} e^{isx} d\mu_n$ mit $|e^{isx}| = 1$. Damit ist die Behauptung trivial.

3. \Rightarrow 1. Dieser Schritt ist am schwierigsten; wir unterteilen den Beweis in mehrere Schritte:

(a) Es bezeichne $q_1, q_2, \dots \in \mathbf{Q}$ eine Nummerierung der rationalen Zahlen. Für festes $j \in \mathbf{N}$ betrachten wir die Folge $(F_n(q_j))_{n \in \mathbf{N}} = \{F_n(q_j) : n \in \mathbf{N}\}$, wobei

$$F_n(q_j) = \frac{1}{2\pi} \int_{-\infty}^{q_j} \int_{-\infty}^{\infty} e^{-isx} \varphi_{X_n}(s) ds dx.$$

Da $0 \leq F_n(q_j) \leq 1$ hat $(F_n(q_j))_{n \in \mathbf{N}}$ mindestens eine konvergente Teilfolge, von der wir wiederum eine Teilfolge $(G_k)_{k \in \mathbf{N}} = \{F_{n_k}(q) : q \in \mathbf{Q}, k \in \mathbf{N}\}$ mit der Eigenschaft $G_k(q) \rightarrow g_q$ für alle $q \in \mathbf{Q}$ wählen können, wobei aufgrund der Monotonie von F_n gilt, dass

$$q < p \quad \Rightarrow \quad g_q \leq g_p.$$

(b) Aus den Grenzwerten g_q der Teilfolgen wird durch

$$G(x) := \inf_{q > x} g_q. \tag{2.6}$$

eine rechtsseitig stetige Funktion definiert; wegen der Monotonie der g_q ist G monoton wachsend und nach Konstruktion gilt für jede Folge $(x_n)_{n \in \mathbf{N}}$ mit $x_n \downarrow x$, dass $G(x) = \inf_{n \in \mathbf{N}} G(x_n)$, denn $q > x$ impliziert wegen der Konvergenz der Folge, dass auch $q > x_n$ für alle hinreichend großen n . Also ist die Funktion G rechtsseitig stetig. Es lässt sich zeigen, dass sogar

$$\lim_{n \rightarrow \infty} G_n(x) = G(x)$$

gilt, wenn $x \in \mathbf{R}$ eine Stetigkeitsstelle von G ist. Dazu betrachten wir eine rationale Zahl $q > x$, für die nach dem oben Gesagten gilt, dass $G_n(x) \leq G_n(q) \rightarrow g_q$. Folglich ist

$$\limsup_{n \rightarrow \infty} G_n(x) \leq g_q.$$

Da das für alle $q \in \mathbf{Q}$ mit $q > x$ gilt, ist nach (2.6)

$$\limsup_{n \rightarrow \infty} G_n(x) \leq G(x).$$

Umgekehrt erhalten wir für $y < p < x$ mit $p \in \mathbf{Q}$ und $x, y \in \mathbf{R}$ die Ungleichungen

$$\liminf_{n \rightarrow \infty} G_n(x) \geq \liminf_{n \rightarrow \infty} G_n(p) = g_p = G(p) \geq G(y),$$

die für alle reellen Zahlen $y < x$ gelten. Da wir angenommen haben, dass G an der Stelle x stetig sei, folgt mit der obigen Abschätzung sofort, dass

$$\liminf_{n \rightarrow \infty} G_n(x) \geq \sup_{y < x} G(y) = \lim_{\epsilon \downarrow 0} G(x - \epsilon) = G(x) \geq \limsup_{n \rightarrow \infty} G_n(x),$$

also mit anderen Worten:

$$\lim_{n \rightarrow \infty} G_n(x) = G(x).$$

- (c) Aus der Stetigkeit von φ_X folgt nun, dass G tatsächlich eine Verteilungsfunktion ist: Sei φ_X die charakteristische Funktion von X mit der Verteilung μ , so dass

$$\begin{aligned} \frac{1}{2S} \int_{-S}^S \varphi_X(s) ds &= \int_{\mathbf{R}} \left[\frac{1}{2S} \int_{-S}^S e^{isx} ds \right] d\mu(x) \\ &= \int_{\mathbf{R}} \frac{\sin(Sx)}{Sx} d\mu(x) \\ &\leq \int_{\mathbf{R}} \left| \frac{\sin(Sx)}{Sx} \right| d\mu(x) \\ &= \int_{|x| < l} \left| \frac{\sin(Sx)}{Sx} \right| d\mu(x) + \int_{|x| \geq l} \left| \frac{\sin(Sx)}{Sx} \right| d\mu(x) \\ &\leq \mu((-l, l)) + \frac{1}{Sl} \mu((-l, l)^c), \end{aligned}$$

wobei $(-l, l)^c = (-\infty, -l] \cup [l, \infty)$ und wir in der letzten Ungleichung die Abschätzungen $|\sin z| \leq |z|$ sowie $|\sin y| \leq 1$ verwendet haben. Damit haben wir

$$\begin{aligned} 1 - \frac{1}{2S} \int_{-S}^S \varphi_X(s) ds &\geq 1 - \mu((-l, l)) - \frac{1}{Sl} \mu((-l, l)^c) \\ &= \left(1 - \frac{1}{Sl}\right) \mu((-l, l)^c) \\ &= \left(1 - \frac{1}{Sl}\right) (1 - F(l) + F(-l)). \end{aligned}$$

Insbesondere gilt also für $l = 2/S$, dass

$$1 - F(2/S) + F(-2/S) \leq 2 - \frac{1}{S} \int_{-S}^S \varphi_X(s) ds$$

bzw. analog

$$1 - F_{n_k}(2/S) + F_{n_k}(-2/S) \leq 2 - \frac{1}{S} \int_{-S}^S \varphi_{X_{n_k}}(s) ds \quad \forall k \in \mathbf{N}.$$

Wir wählen S so, dass $\pm 2/S$ Stetigkeitsstellen von G sind. Aus der letzten Ungleichung folgt dann mit dem Satz von der beschränkten Konvergenz, Satz 1.32, dass für $k \rightarrow \infty$

$$1 - G(2/S) + G(-2/S) \leq 2 - \frac{1}{S} \int_{-S}^S \varphi_X(s) ds.$$

Um zu sehen, dass G tatsächlich eine Verteilungsfunktion ist, lassen wir $S \rightarrow 0$ gehen, so dass $\pm 2/S$ Stetigkeitsstellen bleiben. Wegen der Stetigkeit von φ_X und $\varphi_X(0) = 1$ folgt aus der Monotonie von G , dass die rechte Seite der letzten Ungleichung gegen 0 strebt, so dass

$$\lim_{x \rightarrow -\infty} G(x) = 0, \quad \lim_{x \rightarrow \infty} G(x) = 1.$$

Folglich ist G eine Verteilungsfunktion.

- (d) Der Rest des Beweises besteht darin, zu zeigen, dass $F_n \rightarrow F$ an allen Stellen, an denen F stetig ist. Dazu verwenden wir, dass wir bereits gezeigt haben, dass mit $\varphi_{X_{n_k}} \rightarrow \varphi_X$ auch $G_k = F_{n_k} \rightarrow G$ an allen Stetigkeitsstellen von G konvergiert. *Per definitionem* ist F die eindeutige Verteilungsfunktion zu φ_X , folglich muss $F = G$ sein. Und da das Argument nicht von der Wahl der Teilfolge $(F_{n_k})_{k \in \mathbb{N}} \subset (F_n)_{n \in \mathbb{N}}$ abhängt, muss es andere Teilfolgen von $(F_n)_{n \in \mathbb{N}}$ geben, deren Teilfolgen alle den Grenzwert G haben. Daraus folgt, dass F_n an allen Stetigkeitsstellen gegen F konvergiert, und wir sind fertig. □

Fazit: Punktweise Konvergenz der charakteristischen Funktion ist gleichbedeutend mit Konvergenz in Verteilung bzw. schwacher Konvergenz der Maße. Allgemein gelten folgende Implikationen (vgl. Lemma 1.31).

Lemma 2.22 (Konvergenz von Zufallsvariablen) *Fast sichere Konvergenz impliziert Konvergenz dem Maße nach; letztere impliziert Konvergenz in Verteilung; die Umkehrungen gelten nicht.*

Beweis. Siehe [10, Satz 7.2]. □

Anmerkung 2.23 Die schwache Konvergenz von Wahrscheinlichkeitsmaßen entspricht der schwach- $*$ -Konvergenz von fast überall beschränkten Funktionen, wie sie in der Funktionalanalysis verstanden wird. Wir definieren

$$L^q(\mathbf{R}, \nu) = \left\{ f: \mathbf{R} \rightarrow \mathbf{R} : f \text{ messbar und } \left(\int_{\mathbf{R}} |f(x)|^q d\nu(x) \right)^{1/q} < \infty \right\}$$

für $1 \leq q < \infty$. Dabei ist ν irgendein endliches Maß auf $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Im Fall $q = \infty$ setzen wir

$$L^\infty(\mathbf{R}, \nu) = \{ f: \mathbf{R} \rightarrow \mathbf{R} : f \text{ messbar und } |f(x)| < \infty \text{ bis auf } \nu\text{-Nullmengen} \}.$$

Unter der *schwachen Konvergenz* einer Folge f_n in $L^q(\mathbf{R}, \nu)$ für $q < \infty$ versteht man die Konvergenz aller linearen Funktionale. Das heißt, $f_n \rightharpoonup f$ in $L^q(\mathbf{R}, \nu)$ genau dann, wenn

$$\int_{\mathbf{R}} f_n \phi d\nu \rightarrow \int_{\mathbf{R}} f \phi d\nu, \quad \forall \phi \in L^p(\mathbf{R}, \nu)$$

mit

$$\frac{1}{q} + \frac{1}{p} = 1.$$

Schwache Konvergenz einer Folge in $L^q(\mathbf{R}, \nu)$ bedeutet also Konvergenz im Dualraum $(L^q(\mathbf{R}, \nu))^*$ —und das ist gerade $L^p(\mathbf{R}, \nu)$ unter der Bedingung $1/q + 1/p = 1$. Der Fall $q = \infty$ ist wieder gesondert zu betrachten: Eine Folge ϕ_n in $L^\infty(\mathbf{R}, \nu)$ heißt *schwach- $*$ -konvergent* mit Grenzwert $\phi \in L^\infty(\mathbf{R}, \nu)$ genau dann, wenn¹⁶

$$\int_{\mathbf{R}} f \phi_n d\nu \rightarrow \int_{\mathbf{R}} f \phi d\nu, \quad \forall f \in L^1(\mathbf{R}, \nu).$$

¹⁶ Symbolisch: $\phi_n \overset{*}{\rightharpoonup} \phi$.

Im Fall, dass die ϕ_n eine Folge von Dichtefunktionen bilden, die durch $\mu_n = \phi_n d\nu$ ein Wahrscheinlichkeitsmaß definieren, entspricht $L^1(\mathbf{R}, \nu)$ gerade dem Raum $C_b(\mathbf{R})$ der beschränkten stetigen Funktionen.

Es bleibt zu klären, warum der Fall $q = \infty$ separat behandelt wird. Dazu mache man sich zunächst klar, dass für $1 < q < \infty$ die L^q -Räume reflexiv sind, das bedeutet $(L^q(\mathbf{R}, \nu))^{**} = L^q(\mathbf{R}, \nu)$. Das Paar L^1, L^∞ ist aber eine Ausnahme, denn hier gilt

$$L^1(\mathbf{R}, \nu)^* = L^\infty(\mathbf{R}, \nu),$$

aber

$$(L^\infty(\mathbf{R}, \nu))^* \supsetneq L^1(\mathbf{R}, \nu).$$

Das bedeutet, dass $L^1(\mathbf{R}, \nu)$ kleiner als der Dualraum von $L^\infty(\mathbf{R}, \nu)$ ist; der "doppelte" Dualraum $(L^1(\mathbf{R}, \nu))^{**}$ enthält also mehr Funktionen als $L^1(\mathbf{R}, \nu)$ bzw. $C_b(\mathbf{R})$ im Fall, dass die ϕ_n Dichtefunktionen sind.

Anmerkung 2.24 Aus der Konvergenz einer Folge in L^1 folgt *nicht* ihre fast sichere Konvergenz. Zum Beispiel gilt für die Folge von Zufallsvariablen

$$X_n(\omega) = \chi_{[j2^{-k}, (j+1)2^{-k}]}(\omega), \quad n = j + 2^k, \quad 0 \leq j \leq 2^k, \quad k \in \mathbf{N}_0$$

auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, P) = ([0, 1], \mathcal{B}([0, 1]), \lambda|_{[0, 1]})$, dass

$$\mathbf{E}[|X_n - X|] = \int_{[0, 1]} |X_n(\omega)| d\lambda(\omega) = 2^{-k}, \quad k = k(n),$$

d.h. $X_n \rightarrow 0$ in $L^1(\Omega, P)$, aber die Folge der X_n konvergiert für kein $\omega \in [0, 1]$ punktweise.

2.3 Unabhängigkeit

Wir rufen uns zunächst die Definition von unabhängigen Zufallsvariablen ins Gedächtnis und zählen einige ihrer elementaren Eigenschaften auf.

Definition 2.25 (Unabhängigkeit I) Zwei Zufallsvariablen $X_1, X_2: \Omega \rightarrow \mathbf{R}$ heißen *unabhängig*, wenn die Ereignisse $\{\omega \in \Omega: X_1(\omega) \in A\}$ und $\{\omega \in \Omega: X_2(\omega) \in B\}$ für alle $A, B \in \mathcal{B}(\mathbf{R})$ unabhängig sind, d.h. wenn

$$P(\{X_1 \in A\} \cap \{X_2 \in B\}) = P(X_1 \in A)P(X_2 \in B)$$

Definition 2.26 (Unabhängigkeit II) Eine abzählbare Familie X_1, X_2, X_3, \dots von Zufallsvariablen $X_i: \Omega \rightarrow \mathbf{R}$ heißt unabhängig, wenn für jede endliche Teilfamilie X_{j_1}, \dots, X_{j_n} , $n \in \mathbf{N}$ gilt, dass

$$P\left(\bigcap_{k=1}^n \{X_{j_k} \in A_k\}\right) = \prod_{k=1}^n P(X_{j_k} \in A_k).$$

Lemma 2.27 Zwei Zufallsvariable X_1, X_2 auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) sind genau dann unabhängig, wenn das Bildmaß $P_X = P \circ X^{-1}$, $X = (X_1, X_2)$ auf $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2))$, sprich: die gemeinsame Verteilung von X_1 und X_2 , das Produktmaß $\mu_1 \otimes \mu_2$ der Randverteilungen $\mu_i = P \circ X_i^{-1}$ ist.

Beweis. ÜA. □

Beispiel 2.28 (Mehrdimensionale Normalverteilung) Die mehrdimensionale (auch: multivariate) Normalverteilung stellt einen Sonderfall dar, denn bei ihr fallen Unabhängigkeit und Unkorreliertheit zusammen.

Es sei $X \sim \mathcal{N}(\mu, \Sigma)$, $X = (X_1, \dots, X_n)$, wobei mit $\mathcal{N}(\mu, \Sigma)$ die n -dimensionale Normalverteilung mit Dichtefunktion (bezüglich des Produkt-Lebesgue-Maßes $dx = dx_1 \dots dx_n$)

$$f(x) = (\det(2\pi\Sigma))^{-1/2} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad x = (x_1, \dots, x_n) \in \mathbf{R}^n$$

bezeichnet werde. Dabei ist $\mu = (\mu_1, \dots, \mu_n)$ der n -dimensionale Vektor der Erwartungswerte $\mu_i = \mathbf{E}[X_i]$ und $\Sigma \in \mathbf{R}^{n \times n}$, $\Sigma = (\Sigma_{ij})_{i,j=1, \dots, n}$ eine symmetrische und positiv definite Matrix mit den Einträgen

$$\Sigma_{ij} = \mathbf{E}[(X_i - \mu_i)(X_j - \mu_j)],$$

Kovarianzmatrix genannt. Offensichtlich ist $f(x) = f_1(x_1) \dots f_n(x_n)$ genau dann, wenn $\Sigma = \text{diag}(\sigma_1^2, \dots, \sigma_n^2)$ Diagonalgestalt hat; das ist gerade dann der Fall, wenn die X_1, \dots, X_n unabhängig sind (siehe Abb. 3).

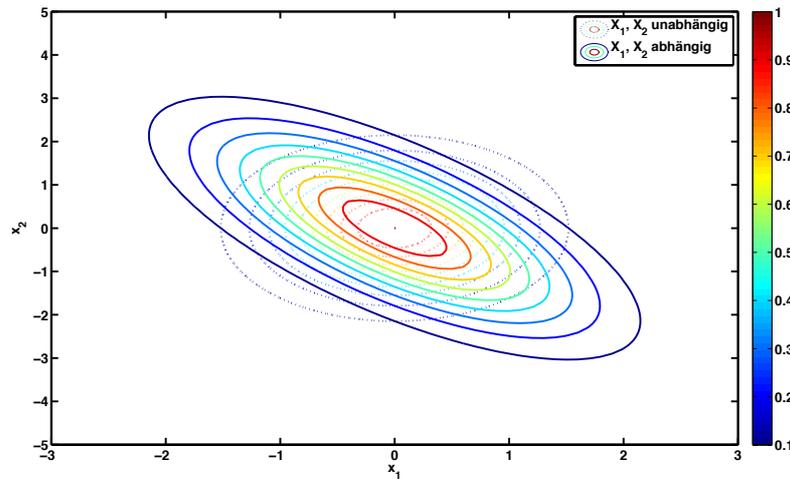


Abb. 3 Dichten abhängiger und unabhängiger normalverteilter Zufallsvariabler X_1, X_2 (Kontourplot zeigt die Isolinien der unnormierten Dichtefunktionen $f = f(x_1, x_2)$).

Beispiel 2.29 (Fallstricke beim Hantieren mit Normalverteilungen) Dass man vorsichtig sein muss, wenn man von der Unkorreliertheit auf die Unabhängigkeit schließen will, zeigt das folgende Beispiel, das aus [11] stammt.

Sei $X \sim \mathcal{N}(0, 1)$ und W eine von X unabhängige, diskrete Zufallsvariable, die mit Wahrscheinlichkeit $P(W = \pm 1) = 1/2$ die Werte $W = \pm 1$ annimmt. Die Zufallsvariable $Y = WX$ ist nach Konstruktion genau wie X standardnormalverteilt, $Y \sim \mathcal{N}(0, 1)$ und abhängig von X . Dennoch sind X und Y unkorreliert, denn

$$\mathbf{E}[XY] = \mathbf{E}[X^2]P(W = 1) - \mathbf{E}[X^2]P(W = -1) = \frac{1}{2} - \frac{1}{2} = 0.$$

Also gilt $P_{XY} \neq P_X \otimes P_Y$. Man mache sich klar, dass das kein Widerspruch zum oben Gesagten ist, denn nur die Randverteilungen P_X, P_Y von X und Y sind Gaußsch, nicht aber ihre gemeinsame Verteilung P_{XY} . (Wäre sie es, so müssten X und Y auch unabhängig sein.)

Summen von unabhängigen Zufallsvariablen

Lemma 2.30 (Summen unabhängiger Zufallsvariabler I) *Seien $X, Y: \Omega \rightarrow \mathbf{R}$ zwei unabhängige Zufallsvariable und $Z = X + Y$ ihre Summe. Dann ist*

$$P_Z(B) = \int_{\mathbf{R}} P_X(B - s) dP_Y(s) = \int_{\mathbf{R}} P_Y(B - s) dP_X(s) \quad \forall B \in \mathcal{B}(\mathbf{R}).$$

Haben P_X, P_Y Lebesguegedichten f_X, f_Y , so hat die Verteilung der Summe die Dichte

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z - s)f_Y(s) ds = \int_{-\infty}^{\infty} f_Y(z - s)f_X(s) ds \quad \forall z \in \mathbf{R}.$$

Beweis. Wegen der Unabhängigkeit ist die gemeinsame Verteilung von X und Y das Produktmaß $P_{XY} = P_X \otimes P_Y$. Berechnen des Bildmaßes von P_{XY} unter der Abbildung $(x, y) \mapsto x + y$ liefert die Aussage. \square

Anmerkung 2.31 Die Operation

$$(f * g)(y) = \int_{-\infty}^{\infty} f(y - x)g(x) dx = \int_{-\infty}^{\infty} g(y - x)f(x) dx$$

wird *Faltung* von f und g genannt (Integrabilität von f und g vorausgesetzt). Die Faltung ist symmetrisch: Es gilt $f * g = g * f$.

Lemma 2.32 (Summen unabhängiger Zufallsvariabler II) *Die Summe $Z = X + Y$ zweier unabhängiger Zufallsvariabler X, Y hat die charakteristische Funktion*

$$\varphi_Z(s) = \varphi_X(s)\varphi_Y(s).$$

Dabei sind $\varphi_X(s) = \mathbf{E}[e^{isX}]$ und $\varphi_Y = \mathbf{E}[e^{isY}]$ die charakteristischen Funktionen von X und Y .

Beweis. Die Fouriertransformation einer Faltung zweier Funktionen ist das Produkt der Fouriertransformierten Funktionen. \square

Wir zählen einige unmittelbare Folgerungen aus den vorstehenden Überlegungen auf, deren Beweis wir den Lesern als Übungsaufgabe überlassen (ebenso wie die Verallgemeinerung auf abzählbar viele Zufallsvariable).

Korollar 2.33 (Rechenregeln für Summen von Zufallsvariablen) *Für unabhängige Zufallsvariable X, Y gilt:*

1. $\mathbf{V}[X + Y] = \mathbf{V}[X] + \mathbf{V}[Y]$,
2. $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$,
3. $\mathbf{E}[f(X)g(Y)] = \mathbf{E}[f(X)]\mathbf{E}[g(Y)]$ für messbare Funktionen $f, g: \mathbf{R} \rightarrow \mathbf{R}$.

Beispiel 2.34 (Ändern der Verteilung beim Summieren) Im Allgemeinen ändert sich der Verteilungstyp beim Summieren, auch wenn die Summanden die gleiche Verteilung haben, sprich: identisch verteilt sind. Zum Beispiel ist die Summe von zwei unabhängigen Zufallsvariablen $X, Y \sim \mathcal{U}(0, 1)$ nicht mehr uniform verteilt, denn für $f_X = f_Y = \chi_{[0,1]}$ rechnet man leicht nach, dass

$$f_{X+Y}(z) = \begin{cases} z, & z \in [0, 1] \\ 2 - z, & z \in (1, 2] \\ 0 & \text{sonst.} \end{cases}$$

Eine der wenigen Ausnahmen ist die Normalverteilung.¹⁷ Sind $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ unabhängig normalverteilt (aber nicht notwendigerweise identisch verteilt), so ist

$$X + Y \sim \mathcal{N}(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

2.4 Grenzwertsätze

Ein in jeder Hinsicht nützliches Werkzeug sind die Ungleichungen vom Typ der Tschebyscheff-Ungleichung: Sei $X: \Omega \rightarrow \mathbf{R}$ eine Zufallsvariable mit endlichem Erwartungswert $\mathbf{E}[X] < \infty$. Dann gilt für alle $\delta > 0$:

$$\begin{aligned} P(|X - \mathbf{E}[X]| \geq \delta) &= \int_{|X - \mathbf{E}[X]| \geq \delta} dP \\ &\leq \delta^{-k} \int_{|X - \mathbf{E}[X]| \geq \delta} |X - \mathbf{E}[X]|^k dP \\ &\leq \delta^{-k} \int_{\Omega} |X - \mathbf{E}[X]|^k dP. \end{aligned} \tag{2.7}$$

wobei im Falle $k = 2m$ das Integral gerade $C_{2m}(X)$, d.h. das $2m$ -te zentrierte Moment ist. Der Spezialfall $k = 2$ ist unter dem Namen *Tschebyscheff-Ungleichung* geläufig:

$$P(|X - \mathbf{E}[X]| \geq \delta) \leq \frac{\mathbf{V}[X]}{\delta^2} \quad \forall \delta > 0. \tag{2.8}$$

¹⁷ Tatsächlich stellt die Normalverteilung einen Sonderfall dar, denn sie bildet einen Fixpunkt beim Summieren – nach dem Zentralen Grenzwertsatz (s.u.) konvergiert die Verteilung einer Summe von unabhängigen und identisch verteilten (IID) Zufallsvariablen mit endlicher Varianz für $n \rightarrow \infty$ gegen die Normalverteilung.

Beispiel 2.35 (Abschätzen von Wahrscheinlichkeiten) Wir betrachten das wiederholte Werfen eines möglicherweise unfairen Würfels. Dazu sei $(\Omega, 2^\Omega, P)$ der Raum der diskreten Gleichverteilung (Laplaceraum) über $\Omega = \{1, \dots, 6\}$. Wir fragen nach der Häufigkeit des Auftretens der Zahl 6 und definieren $X(\omega) = \chi_{\{6\}}(\omega)$ als die Indikatorfunktion des Ereignisses $\omega = \{6\}$; entsprechend steht X_i für die Zahl 6 beim i -ten Wurf. Dann gilt

$$\mathbf{E}[X] = p, \quad \mathbf{V}[X] = p(1-p),$$

mit $p \in [0, 1]$ und die Tschebyscheff-Ungleichung liefert für die relative Häufigkeit

$$Y_n = \frac{X_1 + \dots + X_n}{n} \quad (X_1, \dots, X_n \text{ unabhängig})$$

die Abschätzung

$$\begin{aligned} P(|Y_n - p| \geq \delta) &\leq \frac{\mathbf{V}[Y_n]}{\delta^2} \\ &= \frac{p(1-p)}{n\delta^2} \\ &\leq \frac{1}{4n\delta^2}. \end{aligned}$$

Dabei haben wir ausgenutzt, dass $\mathbf{E}[Y_n] = \mathbf{E}[X_1]$ und wegen der Unabhängigkeit $\mathbf{V}[Y_n] = \mathbf{V}[X_1]/n$ ist.

Schwaches Gesetz der großen Zahlen

Wir formulieren das schwache Gesetz der großen Zahlen in zwei Varianten, deren erste nichts anderes als eine elementare Umformulierung der Tschebyscheff-Ungleichung (2.8) ist.

Satz 2.36 (Schwaches Gesetz der großen Zahlen I) *Sei X_1, X_2, \dots eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit endlicher Varianz $\mathbf{V}[X_1] < \infty$. Dann gilt*

$$\lim_{n \rightarrow \infty} P\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbf{E}[X_1]\right| \geq \epsilon\right) = 0 \quad \forall \epsilon > 0,$$

d.h.,

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mathbf{E}[X_1].$$

Mit anderen Worten: Der Mittelwert konvergiert in Wahrscheinlichkeit gegen den Erwartungswert.

Beweis. ÜA. □

Satz 2.37 (Schwaches Gesetz der großen Zahlen II) *Sei X_1, X_2, \dots eine Folge von unabhängigen, identisch verteilten Zufallsvariablen, deren Erwartungswert existiert. Dann gilt wieder*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{P} \mathbf{E}[X_1].$$

Beweis. Wir beweisen den Satz mit Hilfe von charakteristischen Funktionen. Dazu sei $\varphi_X(s) = \mathbf{E}[e^{isX_1}]$. Nach Lemma 2.32 hat der Mittelwert Y_n die charakteristische Funktion

$$\varphi_{Y_n}(s) = \mathbf{E}\left[e^{\frac{is}{n}(X_1 + \dots + X_n)}\right] = \left(\varphi_X\left(\frac{s}{n}\right)\right)^n.$$

Nun ist $\varphi_X(s)$ wegen $\mathbf{E}[X_1] < \infty$ an der Stelle $s = 0$ stetig differenzierbar, so dass $\varphi_X'(0) = i\mathbf{E}[X_1]$. Wir setzen $m = \mathbf{E}[X_1]$. Die Taylorentwicklung von φ_X bis zur ersten Ordnung liefert dann

$$\varphi_X\left(\frac{s}{n}\right) = 1 + \frac{ism}{n} + o\left(\frac{1}{n}\right),$$

wobei das Landau-Symbol $o(n^\alpha)$ mit $\alpha \in \mathbf{R}$ alle Terme bezeichnet, die asymptotisch gegenüber n^α vernachlässigbar sind. Damit ergibt sich für den Mittelwert Y_n :

$$\begin{aligned} \varphi_{Y_n}(s) &= \left(1 + \frac{ism}{n}\right)^n + o(1) \\ &\rightarrow e^{ims} \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Letztere ist die charakteristische Funktion des Diracmaßes δ_m und nach Satz 2.21 ist dies gleichbedeutend mit der Aussage $P_{Y_n} \rightarrow \delta_m$ bzw. $Y_n \xrightarrow{i.V.} Y$ mit $Y = m$ (fast sicher).

Wir müssen noch zeigen, dass aus $Y_n \xrightarrow{i.V.} m$ sogar die stärkere Aussage $Y_n \xrightarrow{P} m$ folgt. Das ist aber klar, denn wegen der schwachen Maßkonvergenz $P_{Y_n} \rightarrow \delta_m$ gibt es eine stetige und beschränkte Funktion $f: \mathbf{R} \rightarrow [0, 1]$ mit den Eigenschaften $f(m) = 1$ und $f \leq \chi_{M_\epsilon}$, $M_\epsilon = (m - \epsilon, m + \epsilon)$, so dass

$$\begin{aligned} P(|Y_n - m| < \epsilon) &= \int_{\Omega} \chi_{M_\epsilon}(Y_n(\omega)) dP(\omega) \\ &\geq \int_{\mathbf{R}} f dP_{Y_n} \\ &\rightarrow \int_{\mathbf{R}} f d\delta_m \\ &= 1. \end{aligned}$$

Also haben wir

$$\lim_{n \rightarrow \infty} P(|Y_n - m| \geq \epsilon) = 1 - \lim_{n \rightarrow \infty} P(|Y_n - m| < \epsilon) = 0 \quad \forall \epsilon > 0,$$

d.h., $Y_n \xrightarrow{P} m$. □

Was folgt, ist weniger ein Beispiel, sondern vielmehr ein Gegenbeispiel zum Gesetz der großen Zahlen.

Beispiel 2.38 (Cauchy-Verteilung) Sei $X: \Omega \rightarrow \mathbf{R}$ eine Cauchy-verteilte Zufallsvariable mit Dichtefunktion $f(x) = (\pi(1+x^2))^{-1}$ und charakteristischer Funktion $\varphi_X(s) = e^{-|s|}$. Dann gilt $\mathbf{E}[|X|] = \infty$ und

$$\varphi_{Y_n}(s) = \varphi_X(s),$$

d.h., Y_n ist für alle n Cauchy-verteilt (vgl. Anmerkung 2.16).

Starkes Gesetz der großen Zahlen

Wir werden nun beweisen, dass der Mittelwert von unabhängigen und identisch verteilten Zufallsvariablen X_1, \dots, X_n unter bestimmten Voraussetzungen sogar fast sicher gegen Erwartungswert konvergiert. Dazu müssen wir zeigen, dass Ausreißer bei der Konvergenz fast sicher nur endlich oft vorkommen. Grob gesprochen betrachten wir dazu Funktionenfolgen der Form $f_n = \chi_{A_n}$ für Ereignisse $A_n \in \mathcal{E}$. Es lässt sich leicht zeigen, dass f_n genau dann in Wahrscheinlichkeit gegen 0 konvergiert, wenn die $P(A_n)$ eine Nullfolge bilden. Zum Beweis stärkerer Aussagen ist der folgende Satz hilfreich.

Lemma 2.39 (Borel-Cantelli-Lemma) *Ist $\sum_{n=1}^{\infty} P(A_n) < \infty$, so gilt*

$$P\left(\left\{\omega \in \Omega: \lim_{n \rightarrow \infty} \chi_{A_n}(\omega) = 0\right\}\right) = 1.$$

Die Umkehrung gilt, wenn A_1, A_2, \dots unabhängig sind, was äquivalent ist zu der Aussage:

$$\sum_{n=1}^{\infty} P(A_n) = \infty \quad \implies \quad P\left(\left\{\omega \in \Omega: \lim_{n \rightarrow \infty} \chi_{A_n}(\omega) = 0\right\}^c\right) = 1.$$

Beweis. Es sei $S(\omega) = \sum_n \chi_{A_n}(\omega)$. Wegen $\sum_n P(A_n) < \infty$ ist $\mathbf{E}[S] = \sum_n P(A_n) < \infty$, woraus folgt, dass $S(\omega)$ fast sicher endlich ist. Damit konvergiert χ_{A_n} fast sicher gegen 0 und der erste Teil des Lemmas ist bewiesen. Für die Umkehrung reicht es zu zeigen, dass folgende Implikation gilt, wenn die A_1, A_2, \dots unabhängig sind:

$$\sum_n P(A_n) = \infty \quad \implies \quad \lim_{m \rightarrow \infty} P(\cup_{n \geq m} A_n) = 1,$$

denn dann ist

$$P\left(\bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j^c\right) \leq P\left(\bigcap_{n=m}^{\infty} A_n^c\right) = 1 - P\left(\bigcup_{n=m}^{\infty} A_n\right) = 0.$$

Wegen der Stetigkeit von P gilt

$$\begin{aligned} P\left(\bigcup_{n=m}^{\infty} A_n\right) &= 1 - P\left(\bigcap_{n=m}^{\infty} A_n^c\right) \\ &= 1 - \prod_{n=m}^{\infty} P(A_n^c) \\ &= 1 - \prod_{n=m}^{\infty} (1 - P(A_n)) \\ &\geq 1 - \exp\left(-\sum_{n=m}^{\infty} P(A_n)\right), \end{aligned}$$

und der letzte Ausdruck ist gleich 1, wenn $\sum_{n \geq m} P(A_n) = \infty$. Dabei haben wir im zweiten Schritt die Unabhängigkeit von A_1, A_2, \dots ausgenutzt und am Ende die Ungleichung $1 - x \leq e^{-x}$. \square

Anmerkung 2.40 Es gilt

$$\{\omega: \lim_{n \rightarrow \infty} \chi_{A_n}(\omega) = 0\}^c = \{\omega: \limsup_{n \rightarrow \infty} \chi_{A_n}(\omega) = 1\} = \{\omega: \omega \in A_n \text{ unendlich oft}\},$$

was gerade $\limsup_{n \rightarrow \infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{j=n}^{\infty} A_j$ entspricht. Das Borel-Cantelli-Lemma sagt also, dass $\omega \in A_n$ fast sicher nur endlich oft vorkommt, sofern die $P(A_n)$ genug abfallen, so dass $\sum_n P(A_n) < \infty$.

Satz 2.41 (Starkes Gesetz der großen Zahlen) *Ist X_1, X_2, \dots eine Folge unabhängiger und identisch verteilter Zufallsvariablen mit der Eigenschaft $\mathbf{E}[X_1^4] = M < \infty$, dann gilt*

$$\frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{f.s.}} \mathbf{E}[X_1].$$

Beweis. O.B.d.A. setzen wir wieder $\mathbf{E}[X_1] = 0$. Für die Summe $S_n = X_1 + \dots + X_n$ gilt dann

$$\mathbf{E}[S_n^4] = n\mathbf{E}[X_1^4] + 3n(n-1)(\mathbf{E}[X_1^2])^2 \leq nM + 3n^2\sigma^4$$

wobei $\mathbf{E}[X_1^2] = \sigma^2 < \infty$ nach Voraussetzung. Aus der Tschebyscheff-Ungleichung für 4. Momente,

$$P\left(\left|\frac{S_n}{n}\right| \geq \epsilon\right) = P(|S_n| \geq n\epsilon) \leq \frac{\mathbf{E}[S_n^4]}{(n\epsilon)^4} \leq \frac{nM + 3n^2\sigma^4}{n^4\epsilon^4},$$

folgt, dass $P(|S_n/n| \geq \epsilon) \rightarrow 0$ mit $\sum_n P(|S_n/n| \geq \epsilon) < \infty$. Nach dem Borel-Cantelli-Lemma gilt damit

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0\right) = 1,$$

was zu beweisen war. \square

Anmerkung 2.42 Die Wahrscheinlichkeiten $P(|S_n/n| \geq \epsilon)$ lassen sich mit etwas Mühe schärfer abschätzen. Damit kann das starke Gesetz sogar unter der schwächeren Voraussetzung $\mathbf{E}[|X_1|] < \infty$ bewiesen werden.

Beispiel 2.43 (Monte-Carlo-Integration) Sei $X: \Omega \rightarrow \mathbf{R}^d$ eine mehrdimensionale Zufallsvariable mit Verteilung μ und $(X_n)_{n \in \mathbf{N}}$ eine Folge unabhängiger, identischer Kopien von X . Dann gilt

$$\frac{1}{N} \sum_{i=1}^N f(X_i) \xrightarrow{\text{f.s.}} \int_{\mathbf{R}^d} f(x) d\mu(x)$$

für jede bezüglich μ integrierbare Funktion $f: \mathbf{R}^d \rightarrow \mathbf{R}$. Anders als bei gewöhnlichen Quadraturverfahren zur Approximation hochdimensionaler Integrale hängt der Approximationsfehler nicht von der Dimension d ab; z.B. ist der numerische Aufwand, um mit der Trapezregel einen Fehler $\epsilon > 0$ zu erreichen, von der Ordnung $\epsilon^{-d/2}$, während eine grobe Fehlerabschätzung für die Monte-Carlo-Integration durch die Tschebyscheff-Ungleichung,

$$P\left(\left|\frac{1}{N} \sum_{i=1}^N f(X_i) - \int_{\mathbf{R}^d} f d\mu\right| \geq \epsilon\right) \leq \frac{\mathbf{V}[f(X)]}{N\epsilon^2},$$

gegeben ist. Obwohl die Abschätzung grob ist, zeigt sie die typische Konvergenzrate von Monte-Carlo-Verfahren: der Fehler wird nur mit $N^{-1/2}$ kleiner (vorausgesetzt, dass die Varianz der zu integrierenden Funktion endlich ist). Das ist zwar langsamer als bei jedem gitterbasierten Quadraturverfahren, aber in hohen Dimensionen unschlagbar.

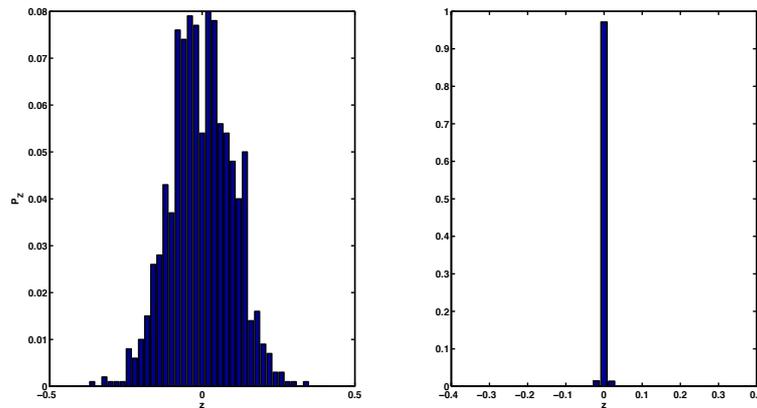


Abb. 4 Empirische Verteilungen des Monte-Carlo-Schätzers Z_N für $N = 100$ (links) und $N = 10000$ (rechts).

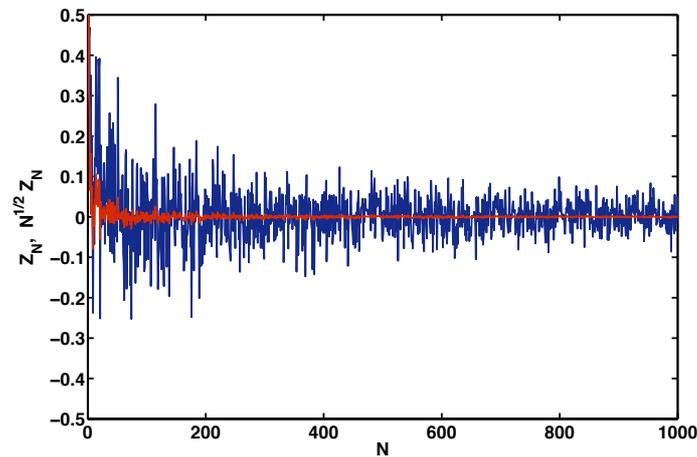


Abb. 5 Realisierung des Monte-Carlo-Schätzers $Z_N(\omega)$ (rote Kurve) und des skalierten Schätzers $\sqrt{N}Z_N(\omega)$ (blau).

Der Zentrale Grenzwertsatz

Wir betrachten noch einmal das letzte Beispiel. Dazu setzen wir $Y = f(X)$ und bezeichnen mit $S_N = (Y_1 + \dots + Y_N)$ die Summe von Y_1 bis Y_N . Dann ist $Z_N = S_N/N$ ein erwartungstreuer Schätzer von $\mathbf{E}[Y]$, der für $N \rightarrow \infty$ fast sicher gegen $\mathbf{E}[Y]$ konvergiert. Für festes N jedoch ist Z_N eine Zufallsvariable, deren Verteilung durch die Verteilung der Summanden $Y_i = f(X_i)$ induziert ist (siehe Abb. 4).

Für eine einzelne Realisierung $Z_N(\omega)$, mit anderen Worten: für ein festes $\omega \in \Omega$, zeigt Abbildung 5 die Konvergenz des Schätzers gegen den Erwartungswert $\mathbf{E}[Y] = 0$. Andererseits können wir nach der Tschebyscheff-Ungleichung erwarten, dass der skalierte Schätzer $\sqrt{N}Z_N(\omega)$ asymptotisch die Streuung um den Erwartungswert angibt. Exemplarisch zeigt das die blaue Kurve in Abb. 5, und mit Hilfe des Zentralen Grenzwertsatzes werden wir dieses Verhalten nun genauer unter die Lupe nehmen.

Satz 2.44 (Zentraler Grenzwertsatz für IID Zufallsvariable¹⁸) Sei $X: \Omega \rightarrow \mathbf{R}$ eine Zufallsvariable mit der Eigenschaft $\mathbf{E}[X] = 0$ und $\mathbf{V}[X] = \sigma^2 < \infty$. Für unabhängige Kopien X_1, X_2, \dots von X gilt

$$\frac{X_1 + \dots + X_n}{\sqrt{n}} \xrightarrow{i.V.} \mathcal{N}(0, \sigma^2).$$

¹⁸ IID = independent and identically distributed (unabhängig und identisch verteilt).

Beweis. Wir führen den Beweis mit Hilfe charakteristischer Funktionen. Sei $\varphi_X(s) = \mathbf{E}[e^{isX}]$ und $S_n = X_1 + \dots + X_n$. Wegen der Unabhängigkeit der X_i hat S_n/\sqrt{n} die charakteristische Funktion

$$\psi_n(s) = \mathbf{E}\left[e^{\frac{is}{\sqrt{n}}(X_1 + \dots + X_n)}\right] = \left(\varphi_X\left(\frac{s}{\sqrt{n}}\right)\right)^n.$$

Da nach Voraussetzung $\mathbf{E}[|X|]$ und $\mathbf{E}[|X|^2]$ existieren, können wir φ_X in eine Taylorreihe bis zur zweiten Ordnung entwickeln:

$$\varphi_X(s) = 1 - \frac{\sigma^2 s^2}{2} + o(s^2).$$

Daraus folgt, dass

$$\psi_n(s) = \left(1 - \frac{\sigma^2 s^2}{2n}\right)^n + o(1) \rightarrow e^{-\frac{\sigma^2 s^2}{2}},$$

und das ist gerade die charakteristische Funktion von $\mathcal{N}(0, \sigma^2)$.¹⁹ Der Lévy-Cramérsche Stetigkeitssatz, Satz 2.21 liefert damit die gewünschte Aussage. \square

Die Formulierung des Zentralen Grenzwertsatzes für unabhängige und identisch verteilte Zufallsvariable geht auf Lévy und Feller²⁰ zurück. Wir werden weiter unten seine Verallgemeinerung diskutieren.

Beispiel 2.45 (Brownsche Bewegung) Eine interessante Anwendung des Zentralen Grenzwertsatzes ist die Konstruktion der Brownschen Molekularbewegung. Im Jahr 1827 beobachtete der schottische Botaniker Robert Brown unter dem Mikroskop die erratischen Bewegungen von Blütenpollen in wässriger Lösung, was ihn zunächst zu der Annahme verleitete, die Blütenpollen seien "lebendig". Heutzutage bezeichnet *Brownsche Bewegung* allgemein die Wärmebewegung von Teilchen in Flüssigkeiten und Gasen, und es war Einstein, der 1905 erstmals eine befriedigende mathematische Erklärung für Browns Beobachtungen lieferte.

Einstein nahm an, dass das beobachtete Teilchen bei Raumtemperatur von einer sehr großen Zahl von Wassermolekülen umgeben ist, die zufällig mit dem Teilchen kollidieren und dadurch seine "Zitterbewegung" verursachen. Sei nun N die Anzahl der Wassermoleküle, die pro Zeiteinheit mit dem beobachteten Teilchen zusammenstoßen und dabei eine zufällige Verschiebung um X_i verursachen, wobei X_1, X_2, \dots unabhängige Kopien einer Zufallsvariable X mit $\mathbf{E}[X] = 0$ seien. Die Position des Teilchens zum Zeitpunkt $t > 0$ ist durch

$$S_t^N = S_0 + \sum_{n=1}^{\lfloor Nt \rfloor} X_n$$

gegeben. Realistischerweise ist N gigantisch ($N \sim 10^{23}$), und es bietet sich an, den Grenzfall $N \rightarrow \infty$ zu betrachten. Damit die mittlere quadratische Verschiebung²¹ des Teilchens endlich bleibt, nehmen wir an, dass $\mathbf{V}[X] = \gamma/N$ für ein $\gamma > 0$ ist, so dass jedes Teilchen nur einen Beitrag zur zufälligen Verschiebung von der Größenordnung $N^{-1/2}$ liefert. Durch umskalieren der Zufallsvariable X_n in $\xi_n = X_n \sqrt{N/\gamma}$ erhalten wir

$$S_t^N = S_0 + \sqrt{\gamma t} \frac{\sum_{n=1}^{\lfloor Nt \rfloor} \xi_n}{\sqrt{Nt}}$$

mit normierten Zufallsvariablen ξ_n , d.h., es gilt $\mathbf{E}[\xi_n] = 0$ und $\mathbf{V}[\xi_n] = 1$ für alle $n = 1, 2, \dots$. Aus dem Zentralen Grenzwertsatz folgt damit die Verteilungskonvergenz von S_t^N für festes t und $N \rightarrow \infty$:

$$S_t^N \xrightarrow{i.V.} \mathcal{N}(S_0, \gamma t).$$

Abbildung 6 zeigt Realisierungen von S_t^N für verschiedene Werte von N und $X_n = \pm N^{-1/2}$ mit $P(X_n = -N^{-1/2}) = P(X_n = N^{-1/2}) = 1/2$. Die pfadweise Konvergenz der Abbildung $t \mapsto S_t^N$ zu einem messbaren stochastischen Prozess $(S_t)_{t \geq 0}$ ist schwieriger zu beweisen und soll an dieser Stelle nicht behandelt werden.²²

Wir erwähnen kurz zwei nützliche Verallgemeinerungen des Zentralen Grenzwertsatzes. Für die Beweise verweisen wir auf [10, S. 203 ff.]

¹⁹ ÜA.

²⁰ William Feller (1906–1970), kroat.-amer. Mathematiker

²¹ Physikerjargon für die Varianz.

²² So ist S_t eine Hölder-stetige Funktion mit dem Exponenten $1/2$, d.h., S_t ist stetig, aber an keiner Stelle differenzierbar. Wer es genauer wissen möchte, sei auf die einschlägige Literatur zum Thema *Diffusionsprozesse* verwiesen, z.B. [1, Kap. 2 und 3].

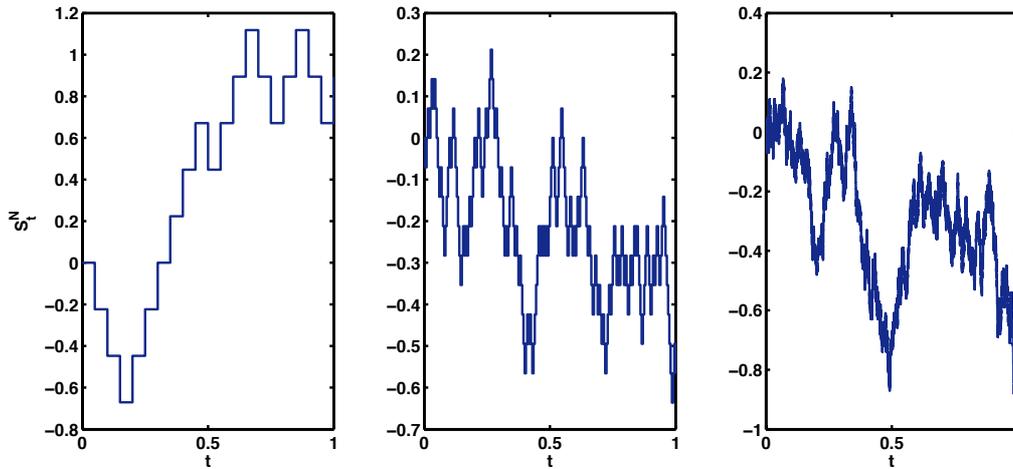


Abb. 6 Verschiedene Realisierungen des Zufallsspaziergangs S_t^N für $N = 20, 200, 10000$ und $X_n = \pm N^{-1/2}$.

1. **Ungleich verteilte Zufallsvariable:** Seien X_1, \dots, X_n unabhängig mit $\mathbf{E}[X_i] = 0$, $\mathbf{V}[X_i] = \sigma_i^2 < \infty$ und Verteilungen $\mu_i = P_{X_i}$, $i = 1, \dots, n$. Es lässt sich leicht zeigen, dass die Zufallsvariable $Y_n = S_n/s_n$ mit $S_n = X_1 + \dots + X_n$ und $s_n^2 = \sigma_1^2 + \dots + \sigma_n^2$ normiert ist, d.h., es gilt $\mathbf{E}[Y_n] = 0$ und $\mathbf{V}[Y_n] = 1$.

Satz 2.46 (Lindeberg-Theorem I) *Seien X_1, \dots, X_n Zufallsvariable wie oben, für die die Lindeberg-Bedingung*

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \int_{|x| \geq \epsilon s_n} x^2 d\mu_i(x) = 0 \quad \forall \epsilon > 0$$

gelte. Dann konvergiert Y_n in Verteilung gegen die Standard-Normalverteilung, d.h. $Y_n \xrightarrow{i.V.} \mathcal{N}(0, 1)$.

Im IID-Fall ist die Lindeberg-Bedingung automatisch erfüllt, d.h., Satz 2.46 impliziert den gewöhnlichen Zentralen Grenzwertsatz, Satz 2.44.²³ Dass die Lindeberg-Bedingung erfüllt ist, sieht man, indem man $s_n^2 = n\sigma^2$ setzt und mit Hilfe von Cauchy-Schwarz- und Tschebyscheff-Ungleichung folgert, dass

$$\begin{aligned} \frac{1}{\sigma^2} \int_{|x| \geq \epsilon \sqrt{n} \sigma} x^2 d\mu &= \frac{1}{\sigma^2} \int_{\mathbf{R}} x^2 \chi_{\{|x| \geq \epsilon \sigma \sqrt{n}\}} d\mu \\ &\leq \frac{1}{\sigma^2} (\mathbf{V}[X])^{1/2} \left(\int_{\mathbf{R}} \chi_{\{|x| \geq \epsilon \sigma \sqrt{n}\}} d\mu \right)^{1/2} \\ &= \frac{1}{\sigma} (P(|X| \geq \epsilon \sigma \sqrt{n}))^{1/2} \\ &\leq \frac{1}{\epsilon \sigma \sqrt{n}}. \end{aligned}$$

2. **Schwach unabhängige Zufallsvariable:** Eine doppelt indizierte Folge von Zufallsvariablen

$$\{X_{n,i} : i = 1, \dots, k_n, n \in \mathbf{N}\}$$

wir *Dreiecksschema* genannt. Auch für diese lässt sich der Zentrale Grenzwertsatz beweisen, wenn die $X_{n,i}$ nur jeweils zeilenweise unabhängig sind. In diesem Fall betrachtet man die Partialsummen $\tilde{S}_n = X_{n,1} + \dots + X_{n,k_n}$, die durch die Wurzel aus den Zeilenvarianzen $\tilde{s}_n^2 = \sigma_{n,1}^2 + \dots + \sigma_{n,k_n}^2$ geteilt werden.

Satz 2.47 (Lindeberg-Theorem II) *Sind die Zeilen $X_{n,1}, \dots, X_{n,k_n}$ eines Dreiecksschemas für alle $n \in \mathbf{N}$ unabhängig mit zeilenweiser Lindeberg-Bedingung*

$$\lim_{n \rightarrow \infty} \frac{1}{\tilde{s}_n^2} \sum_{i=1}^{k_n} \int_{|x| \geq \epsilon \tilde{s}_n} x^2 d\mu_{n,i}(x) = 0 \quad \forall \epsilon > 0,$$

²³ Jarl Waldemar Lindeberg (1876–1932), finn. Mathematiker

so gilt der Zentrale Grenzwertsatz:

$$\frac{\tilde{S}_n}{\tilde{s}_n} \xrightarrow{i.V.} \mathcal{N}(0, 1).$$

Es ist leicht einzusehen, dass Satz 2.47 im Falle unabhängiger Zufallsvariablen das obige Lindeberg-Theorem, Satz 2.46 impliziert und damit auch den Zentralen Grenzwertsatz für IID Zufallsvariable, Satz 2.44.

Konvergenzgeschwindigkeit beim Zentralen Grenzwertsatz

Der Zentrale Grenzwertsatz ist nützlich, um die Verteilung von Summen von IID Zufallsvariablen zu approximieren. Daher stellt sich die Frage, ab welchem n die Approximation durch die Normalverteilung gerechtfertigt ist. Der folgende Satz beantwortet die Frage nach der Konvergenzgeschwindigkeit.

Satz 2.48 (Berry-Esseen-Theorem) *Seien $(X_n)_{n \in \mathbf{N}}$ IID Zufallsvariable mit $\mathbf{E}[X_1] = 0$ und $\mathbf{V}[X_1] = 1$. Zudem sei $\mathbf{E}[|X_1|^{2+\alpha}] < \infty$ für ein $\alpha > 0$. Dann existieren ein $\delta > 0$ und $0 < C < \infty$, so dass*

$$\sup_{a \in \mathbf{R}} \left| P\left(\frac{S_n}{\sqrt{n}} \leq a\right) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a e^{-\frac{x^2}{2}} \right| \leq Cn^{-\delta}.$$

Beweis. Siehe [16, Thm. 3.26]. □

Beispiel 2.49 (Jungen- und Mädchengeburt) Das folgende Beispiel stammt aus [10]. Zwischen 1996 und 1999 wurden in Baden-Württemberg 231 432 Jungen und 218 674 Mädchen geboren. Wie wahrscheinlich ist es, dass diese Diskrepanz innerhalb der "normalen" statistischen Schwankungen liegt?

Wir nehmen zunächst an, dass Jungen- und Mädchengeburt gleichwahrscheinlich sind und wählen als Wahrscheinlichkeitsmodell für jede einzelne Geburt die Bernoulliverteilung mit $p = 1/2$. Sei X die Zufallsvariable *Anzahl der Jungen*; bei $n = 450\,106$ Geburten insgesamt ist die erwartete Anzahl der Jungen $\mathbf{E}[X] = np = 225\,053$, die zugehörige Standardabweichung ist $\sqrt{\mathbf{V}[X]} = \sqrt{np(1-p)} = \sqrt{n/4} \approx 335$. Die relative Differenz zwischen der erwarteten Anzahl von Jungengeburt und dem realen Datum $S_n = 231\,432$ beträgt damit

$$D_n = \frac{S_n - n/2}{\sqrt{n/4}} \approx 19.04$$

Standardabweichungen. Nach dem Zentralen Grenzwertsatz und Satz 2.48 sollte D_n annähernd standard-normalverteilt sein, d.h., $D_n \sim \mathcal{N}(0, 1)$. Da jedoch $P(D_n \geq 19) < 10^{-60}$ ist, wobei $P(\cdot)$ das Maß der Standard-Normalverteilung bezeichnet, ist es nahezu ausgeschlossen, dass Jungen- und Mädchengeburt gleichwahrscheinlich sind und die beobachtete Zahl von Geburten innerhalb der zu erwartenden statistischen Schwankungen liegt.

2.5 Mittlere und große Abweichungen

Wir wenden uns nun wieder dem Standardfall unabhängiger und identisch verteilter Zufallsvariablen X_1, X_2, \dots zu, und nehmen an, dass $\mathbf{E}[X_i] = 0$ und $\mathbf{V}[X_i] = 1$ für alle $i = 1, 2, \dots$. Nach dem starken Gesetz der großen Zahlen und dem Zentralen Grenzwertsatz wissen wir, dass für die Summen $S_n = X_1 + \dots + X_n$ gilt

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{n} = 0\right) = 1$$

bzw.

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}} \geq a\right) = \frac{1}{\sqrt{2\pi}} \int_a^{\infty} e^{-\frac{x^2}{2}} dx.$$

In Anwendungen, z.B. bei der Monte-Carlo-Integration, stellt sich oft die Frage nach der Konvergenzrate – wie schnell konvergieren S_n/n bzw. S_n/\sqrt{n} ? Das Berry-Esseen-Theorem, Satz 2.48, gibt eine Antwort auf diese Frage im Zusammenhang mit dem Zentralen Grenzwertsatz. Darüber hinaus möchte man abschätzen, wie wahrscheinlich Ausreißer für sehr große, aber endliche n sind. Die Frage nach der Konvergenzrate betrifft das Verschwinden der mittleren oder normalen Abweichungen vom Grenzfall $n \rightarrow \infty$, die zweite Frage die sogenannten großen Abweichungen, auch *seltene Ereignisse* genannt. In beiden Fällen interessiert man sich für das Grenzverhalten der Verteilungen von S_n/n bzw. S_n/\sqrt{n} an ihren Rändern (engl.: *tails*).

Die vorstehenden Überlegungen zum starken Gesetz der großen Zahlen und zum Zentralen Grenzwertsatz legen nahe, dass

$$P\left(\frac{S_n}{n} \geq a\right) = e^{-I(a)+o(n)}$$

für eine geeignete Funktion $I \geq 0$, die an der Stelle $a = \mathbf{E}[X_1]$ ein eindeutiges Minimum hat.

Zunächst wollen wir uns überlegen, wie ein für unsere Zwecke geeigneter Wahrscheinlichkeitsraum aussehen könnte. Dazu sei $\Omega = \mathbf{R}^\infty$ die Menge aller reellwertigen Folgen $\{X_n : n \geq 1\}$ von Zufallsvariablen. Zu Ω gibt es eine natürliche σ -Algebra \mathcal{F} , die die kleinste σ -Algebra ist, die alle Zylindermengen der Form

$$F_{E,k} := \{\text{alle Folgen } \omega : (X_1, \dots, X_k) \in E\}$$

enthält, wobei $E \in \mathcal{B}(\mathbf{R}^k)$ irgendeine Borelmenge ist und k eine ganze Zahl. Die $F_{E,k}$ bilden zusammen einen durchschnittstabilen Erzeuger von

$$\mathcal{F} = \sigma(\{F_{E,k} : E \in \mathcal{B}(\mathbf{R}^k), k \geq 1\}).$$

Die Randverteilung P_n der ersten n Folgenglieder ist entsprechend ein Maß auf dem Messraum $(\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n))$, das aus der gemeinsamen Verteilung P auf (Ω, \mathcal{F}) durch die kanonische Projektion $\pi_n(\omega) = (X_1, \dots, X_n)$ mittels $P_n = P \circ \pi_n^{-1}$ induziert wird. Dass ein solches Maß P überhaupt existiert und sich sogar eindeutig aus der (konsistenten) Familie der endlich-dimensionalen Randverteilungen P_n ergibt, garantiert der Konsistenzsatz von Kolmogorov, auf den wir an dieser Stelle nur verweisen (siehe [14, Thm. 2.1.5]).

Definition 2.50 (Terminale σ -Algebra) Ist $(X_n)_{n \in \mathbf{N}}$ eine Folge von Zufallsvariablen und $\mathcal{G}_n = \sigma(\{X_k : k \geq n\})$ die von $(X_n, X_{n+1}, X_{n+2}, \dots)$ erzeugte σ -Algebra²⁴, so heißt

$$\mathcal{G}_\infty = \bigcap_{n=1}^{\infty} \mathcal{G}_n$$

terminale σ -Algebra (engl.: *tail σ -field*); ihre Elemente werden *terminale Ereignisse* genannt (engl.: *tail events*).

Beispiel 2.51 Ist $(X_n)_{n \in \mathbf{N}}$ eine Folge von Zufallsvariablen, so ist

$$A = \left\{ \omega \in \Omega : \sum_{n=1}^{\infty} X_n \text{ konvergiert} \right\}$$

ein terminales Ereignis. Mit anderen Worten: Terminale Ereignisse sind gerade solche Ereignisse, die nicht vom Verhalten endlich vieler Folgenglieder X_n abhängen.

Satz 2.52 (Kolmogorovs 0-1-Gesetz) *Ist $(X_n)_{n \in \mathbf{N}}$ eine Folge von unabhängigen Zufallsvariablen, so ist $P(A) = 0$ oder $P(A) = 1$ für jedes $A \in \mathcal{G}_\infty$.*

Beweis. Wir werden zeigen, dass jedes $A \in \mathcal{G}_\infty$ von sich selbst unabhängig ist. In diesem Fall ist

$$P(A) = P(A \cap A) = (P(A))^2,$$

was nur erfüllt sein kann, wenn $P(A) = 0$ oder $P(A) = 1$ ist. Dazu sei (Ω, \mathcal{F}, P) der oben definierte Wahrscheinlichkeitsraum mit der Grundmenge $\Omega = \mathbf{R}^\infty$ als der Menge aller reellen Folgen $\{X_n : n \geq 1\}$ von Zufallsvariablen X_n . Sei ferner $\mathcal{F}_n = \sigma(\{X_k : k \leq n\})$ die von X_1, \dots, X_n erzeugte σ -Algebra. Wegen $A \in \mathcal{G}_\infty \subset \mathcal{G}_{n+1}$ ist A für alle $n \in \mathbf{N}$ unabhängig von \mathcal{F}_n (sprich: von allen $F \in \mathcal{F}_n$) und wegen der Stetigkeit von P folglich auch von \mathcal{F}_∞ .

Mit \mathcal{M} bezeichnen wir die Familie von Mengen, die von A unabhängig sind. Da mit $M_1 \subset M_2 \subset \dots \in \mathcal{M}$, auch ihre abzählbare Vereinigung $\cup_n M_n$ in \mathcal{M} liegt und mit $M_1 \supset M_2 \supset \dots \in \mathcal{M}$ auch ihr Durchschnitt $\cap_n M_n$, ist \mathcal{M} eine monotone Klasse; nach dem Satz über monotone Klassen (siehe [7, Satz I.6.2]) enthält \mathcal{M} mit \mathcal{F}_∞ insbesondere die von \mathcal{F}_∞ erzeugte σ -Algebra $\sigma(\mathcal{F}_\infty)$ – das ist aber gerade \mathcal{F} und da $A \in \mathcal{F}$, ist A von sich selbst unabhängig. \square

²⁴ Gemäß der obenstehenden Definition von \mathcal{F} wird \mathcal{G}_n von dem durchschnittstabilen System der Zylindermengen $G_{E,k} := \{\omega : (X_n, X_{n+1}, \dots, X_{n+k}) \in E\} \in \mathcal{F}$ erzeugt, wobei E die Borelmengen aus $\mathcal{B}(\mathbf{R}^k)$ sind und k über alle ganzen Zahlen läuft.

Bevor wir zur Frage der Konvergenzgeschwindigkeit kommen, stellen wir zunächst einige generelle Überlegungen zur Konvergenz von Summen von Zufallsvariablen an; insbesondere interessiert uns die Frage, ob S_n/\sqrt{n} für $n \rightarrow \infty$ gegen eine "echte" Zufallsvariable Y mit Verteilung $\mathcal{N}(0, 1)$ konvergiert.²⁵ Die schlichte Antwort lautet "Nein" und hat mit dem Abfallverhalten der Ränder der Verteilung der Summe S_n/\sqrt{n} zu tun.

Lemma 2.53 Für jede Folge $(n_k)_{k \in \mathbf{N}}$ mit $n_k \uparrow \infty$ gilt

$$P\left(\limsup_{k \rightarrow \infty} \frac{S_{n_k}}{\sqrt{n_k}} = \infty\right) = 1$$

bzw.

$$P\left(\liminf_{k \rightarrow \infty} \frac{S_{n_k}}{\sqrt{n_k}} = -\infty\right) = 1.$$

Beweis. Wir beweisen nur den ersten Teil der Behauptung. Sei $Z = \limsup_{k \rightarrow \infty} S_{n_k}/\sqrt{n_k}$. Da

$$P\left(\lim_{n \rightarrow \infty} \frac{S_n}{\sqrt{n}} \geq a\right) = \frac{1}{\sqrt{2\pi}} \int_a^\infty e^{-\frac{x^2}{2}} dx > 0 \quad \forall a \in \mathbf{R}$$

muss erst recht $P(Z \geq a) > 0$ für beliebige $a \in \mathbf{R}$ gelten. Da aber Z nicht von X_1, \dots, X_j für endliches $j \in \mathbf{N}$ abhängt, ist das Ereignis $A = \{\omega : Z(\omega) \geq a\}$ terminal, d.h., $P(A) = 0$ oder $P(A) = 1$. Wegen $P(A) > 0$ muss also $P(A) = 1$ sein, womit die Behauptung bewiesen ist. (Den zweiten Teil der Aussage beweist man analog.) \square

Beispiel 2.54 (Brownsche Bewegung, Fortsetzung) Es sei

$$S_t^N = \sqrt{t} \frac{\sum_{n=1}^{\lfloor Nt \rfloor} \xi_n}{\sqrt{Nt}}$$

die diskrete Version der Brownschen Bewegung aus Beispiel 2.45 mit $S_0 = 0$ und $\gamma = 1$. Dabei sind ξ_n IID Zufallsvariable mit $\mathbf{E}[\xi_1] = 0$ und $\mathbf{V}[\xi_1] = 1$ und nach dem Zentralen Grenzwertsatz gilt

$$S_t^N \xrightarrow{i.V.} \mathcal{N}(0, t).$$

Nach Lemma 2.53 ist

$$P\left(\limsup_{N \rightarrow \infty} S_t^N = \infty\right) = P\left(\liminf_{N \rightarrow \infty} S_t^N = -\infty\right) = 1,$$

d.h., mit Wahrscheinlichkeit 1 über- bzw. unterschreitet S_t^N für $N \rightarrow \infty$ jede beliebige Schranke $\pm a$. (Das ist umso bemerkenswerter, als dass für alle $t > 0$ und $N \in \mathbf{N}$ gilt: $\mathbf{E}[S_t^N] = 0$ und $\mathbf{V}[S_t^N] < \infty$.)

Konvergenzgeschwindigkeit beim starken Gesetz der großen Zahlen

Beim Beweis des schwachen Gesetzes der großen Zahlen haben wir gesehen, dass die Wahrscheinlichkeit für Abweichungen vom Erwartungswert von der Größenordnung ϵ mit $N^{-1/2}$ gegen 0 geht (vgl. Beispiel 2.43). Die Beobachtung deckt sich mit der Tatsache, dass nach dem Zentralen Grenzwertsatz die Standardabweichung des Mittelwertschätzers mit der Rate $N^{-1/2}$ kleiner wird. Für die fast sichere Konvergenz sind die Abschätzungen, die auf dem Zentralen Grenzwertsatz oder der Tschebyscheff-Ungleichung basieren, zu grob. Der nachfolgende Satz sagt aus, dass die Rate beim starken Gesetz der großen Zahlen geringfügig kleiner ist, nämlich:

$$\left| \frac{S_n}{n} - \mathbf{E}[X_1] \right| = \mathcal{O}\left(\sqrt{\frac{\log \log n}{n}}\right) \quad \text{f.s.}$$

Satz 2.55 (Gesetz vom iterierten Logarithmus) Für jede Folge von unabhängigen, identisch verteilten Zufallsvariablen X_1, X_2, \dots mit $\mathbf{E}[X_1] = 0$ und $\mathbf{V}[X_1] = 1$ gilt

$$P\left(\limsup_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{\sqrt{2n \log \log n}} = 1\right) = 1$$

bzw.

$$P\left(\liminf_{n \rightarrow \infty} \frac{X_1 + \dots + X_n}{\sqrt{2n \log \log n}} = -1\right) = 1.$$

²⁵ Man spricht von Y als *uneigentlicher Zufallsvariable* – das ist eine Zufallsvariable, die mit positiver Wahrscheinlichkeit die Werte $\pm\infty$ annimmt. Ähnliche Überlegungen gelten für die Konvergenz des Mittelwerts S_n/n für $n \rightarrow \infty$ (ÜA).

Beweis. Wir skizzieren den Beweis für den ersten Teil der Aussage: In diesem Fall ist zu zeigen, dass $P(S_n \geq \sqrt{\lambda n \log \log n} \text{ unendlich oft}) = 0$ ist, falls $\lambda > 2$. Die Strategie dabei wird sein, das Borel-Cantelli-Lemma auf die Ränder der Verteilung von S_n/\sqrt{n} anzuwenden. Hinreichend nach Lemma 2.39 wäre die Bedingung

$$\sum_{n=1}^{\infty} P\left(S_n \geq \sqrt{\lambda n \log \log n}\right) < \infty,$$

falls $\lambda > 2$. Weil die Partialsummen S_n nicht unabhängig sind, ist das Kriterium jedoch nicht notwendig, und wir kommen mit einer etwas schwächeren Bedingung aus. Dazu verfährt man wie folgt.

- Wir setzen $\phi(n) = \sqrt{n \log \log n}$; für $n \geq 3$ ist $\phi(n)$ monoton wachsend und es reicht, eine Folge $(k_n)_{n \in \mathbf{N}}$ zu finden, so dass

$$\sum_{n=1}^{\infty} P\left(\sup_{k_{n-1} \leq j \leq k_n} S_j \geq \sqrt{\lambda} \phi(k_{n-1})\right) < \infty,$$

für $\lambda > 2$. Dann gilt nämlich nach dem Borel-Cantelli-Lemma, Lemma 2.39, dass

$$P\left(\limsup_{n \rightarrow \infty} \frac{\sup_{k_{n-1} \leq j \leq k_n} S_j}{\sqrt{\lambda} \phi(k_{n-1})} \leq 1\right) = 1.$$

- Wegen der Monotonie von ϕ gilt dann auch

$$P\left(\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{\lambda} \phi(n)} \leq 1\right) = 1.$$

Die Behauptung des Satzes folgt dann aus der Tatsache, dass $\lambda > 2$ beliebig ist.

- Der (durchaus anspruchsvolle) Rest des Beweises besteht darin, mit Hilfe von Satz 2.48 die Terme $P(\sup \dots)$ abzuschätzen; das macht man zunächst für standard-normalverteilte X_1, X_2, \dots und sodann für beliebige Verteilungen unter der Voraussetzung, dass $\mathbf{E}[|X_1|^{2+\alpha}]$ für ein $\alpha > 0$ existiert (vgl. [16, Thm. 3.25]).

□

Große Abweichungen vom Mittel (seltene Ereignisse)

Sei $X: \Omega \rightarrow \mathbf{R}$ eine Zufallsvariable mit Erwartungswert $\mathbf{E}[X] = 0$, für die wir die Wahrscheinlichkeit $p = P(X \geq l)$ für ein $l \geq 0$ berechnen wollen. Sind X_1, \dots, X_n unabhängigen Kopien von X , so ist

$$\hat{p}_n = \frac{1}{n} \sum_{i=1}^n \chi_{\{X_i \geq l\}} = \frac{1}{n} \sum_{i=1}^n \chi_{[l, \infty)} \quad (2.9)$$

ein erwartungstreuer und konsistenter Schätzer für

$$p = P(X \geq l) = \mathbf{E}[\chi_{\{X \geq l\}}].$$

Nach dem starken Gesetz der großen Zahlen konvergiert \hat{p}_n für alle $l \in \mathbf{R}$ sogar fast sicher gegen p , und der Zentrale Grenzwertsatz erlaubt es uns, für hinreichend große n durch

$$P\left(|\hat{p}_n - p| \geq \frac{a}{\sqrt{n}}\right) \approx \sqrt{\frac{2}{\pi}} \int_{-\infty}^{-a/\sigma} e^{-\frac{x^2}{2}} dx$$

entsprechende Konfidenzintervalle zu berechnen. So weit, so gut – wie aber wird sich unser Schätzer verhalten, wenn l von der selben Größenordnung wie n oder \sqrt{n} ist? In der Praxis bedeutet das, dass sowohl l als auch n endlich sind, aber $l \sim \sqrt{n}$. In diesem Fall spricht man von $\{X \geq l\}$ als *seltene Ereignis*. Das Ereignis $\{X \geq l\}$ kann selten sein, weil zum Beispiel l sehr viel größer als $\mathbf{E}[X]$ ist (bei beschränkter Varianz), so dass $p \approx 0$ ist. Typische Anwendungsbeispiele, bei denen seltene Ereignisse eine Rolle spielen, sind das Abschätzen großer Versicherungsschäden, die Prognose von Naturkatastrophen oder Phasenübergänge in physikalischen Systemen.

Beispiel 2.56 (Wahrscheinlichkeit großer Werte bei der Normalverteilung) Wir betrachten eine $\mathcal{N}(0, 1)$ -verteilte Zufallsvariable X mit unabhängigen Kopien X_1, \dots, X_n für hinreichend großes n . Für $l = 5$ ist $p = P(X \geq l) \approx 3 \cdot 10^{-7}$, und wir möchten p durch \hat{p}_n wie in (2.9) schätzen. Dazu lassen wir uns in Matlab $n = 100\,000$ standard-normalverteilte Zufallszahlen x_1, \dots, x_n erzeugen und erhalten die Schätzung

$$\hat{p}_{100000} = \frac{1}{100000} \sum_{i=1}^{100000} \chi_{\{x_i \geq l\}} = 0.$$

Das ist zwar nicht p , aber auch nicht weiter überraschend, denn das Ereignis $\{X \geq 5\}$ ist selten. Leider wird unsere Schätzung nicht besser, wenn wir das Experiment 100 mal ausführen und über die Wiederholungen mitteln:

$$\hat{p}_{100000}^{(100)} = \frac{1}{100} \sum_{j=1}^{100} \left(\frac{1}{100000} \sum_{i=1}^{100000} \chi_{\{x_i^{(j)} \geq l\}} \right) = 1 \cdot 10^{-7}.$$

Das ist zwar etwas besser als vorher, aber immer noch um einen Faktor 3 daneben – von den insgesamt 10^7 erzeugten Zufallszahlen²⁶ war gerade mal eine einzige größer als 5. Um das Problem besser zu verstehen, schauen wir uns zunächst die Varianz des Schätzers als Maß für seine Konvergenz an; da die beiden möglichen Ereignisse $\{X \geq 5\}$ bzw. $\{X < 5\}$ bernoulliverteilt mit Wahrscheinlichkeit p sind, ist

$$\mathbf{V}[\hat{p}_n] = \frac{\mathbf{V}[\chi_{[l, \infty)}]}{n} = \frac{p(1-p)}{n} \leq \frac{1}{4n}.$$

Der Schätzer verhält sich also erwartungsgemäß. Das Problem ist, dass die Varianz nicht schnell genug abfällt, um den relativen Fehler unter Kontrolle zu halten, denn dieser ist für verschwindende p unbeschränkt:

$$\varepsilon_{\text{rel}} = \frac{\text{Standardabweichung von } \hat{p}_n}{\text{Erwartungswert von } \hat{p}_n} = \frac{1}{\sqrt{n}} \frac{\sqrt{p-p^2}}{p}.$$

Wie man sieht, geht ε_{rel} für $p \downarrow 0$ gegen unendlich. Im unseren Fall ist $\varepsilon_{\text{rel}} \approx 2 \cdot 10^3 / \sqrt{n}$ und das ist für $n = 10\,000$ noch immer $2\sqrt{10} \approx 6.32$; bei 100 Wiederholungen verringert sich der relative Fehler um den Faktor 10.

Um das Konvergenzverhalten unseres Schätzers besser zu verstehen, lohnt es sich, die Wahrscheinlichkeit $P(S_n/n \geq x)$ für $n \gg 1$ und $x > \mathbf{E}[X] = 0$ genauer unter die Lupe zu nehmen. Wir wissen, dass

$$P\left(\frac{S_n}{n} \geq x\right) \rightarrow 0 \quad \& \quad P\left(\frac{S_n}{n} \in [0, x)\right) \rightarrow \frac{1}{2} \quad \forall x > 0,$$

wobei die erste Aussage aus dem Gesetz der großen Zahlen, die zweite aus dem Zentralen Grenzwertsatz folgt. Wir wollen nun einen geeigneten asymptotischen Ausdruck für $p_n = P(S_n/n \geq x)$ für $n \rightarrow \infty$ herleiten, der die Konvergenz der beiden Aussagen genauer beschreibt. Dazu benötigen wir einige Werkzeuge.

Definition 2.57 (Kumulantenerzeugende Funktion) Die Funktion

$$\gamma: \mathbf{R} \rightarrow (-\infty, \infty], \quad \gamma(s) = \log \mathbf{E}[e^{sX}]$$

heißt *kumulantenerzeugende Funktion* von X .

Lemma 2.58 (Hölder-Ungleichung²⁷) Seien $X, Y: \Omega \rightarrow \mathbf{R}$ messbar mit $\mathbf{E}[|X|], \mathbf{E}[|Y|] < \infty$. Dann gilt

$$\mathbf{E}[|XY|] \leq (\mathbf{E}[|X|^p])^{1/p} (\mathbf{E}[|Y|^q])^{1/q}$$

für alle $1 \leq p, q \leq \infty$ mit $\frac{1}{p} + \frac{1}{q} = 1$.

Beweis. Siehe [7, Kap. VI.1.2]. □

Korollar 2.59 Die Kumulantenerzeugende γ ist konvex mit $\gamma(0) = 0$.

²⁶ Es ist im übrigen kaum anzunehmen, dass die 10^7 von Matlab erzeugten normalverteilten Zufallszahlen tatsächlich noch unabhängig sind. Das soll an dieser Stelle aber nicht unser Problem sein.

²⁷ Otto Hölder (1859–1937), dt. Mathematiker

Beweis. Die Konvexität ergibt sich aus der Hölderschen Ungleichung: Für $\alpha \in [0, 1]$ ist

$$\begin{aligned}\mathbf{E}[e^{\alpha t X + (1-\alpha)s X}] &= \mathbf{E}[e^{\alpha t X} e^{(1-\alpha)s X}] \\ &\leq \mathbf{E}[(e^{\alpha t X})^{1/\alpha}]^\alpha \mathbf{E}[(e^{(1-\alpha)s X})^{1/(1-\alpha)}]^{1-\alpha} \\ &= \mathbf{E}[e^{t X}]^\alpha \mathbf{E}[e^{s X}]^{1-\alpha},\end{aligned}$$

wobei wir im zweiten Schritt die Hölder-Ungleichung mit $p = 1/\alpha$ und $q = 1/(1-\alpha)$ verwendet haben. Logarithmieren liefert dann die Ungleichung

$$\gamma(\alpha t + (1-\alpha)s) \leq \alpha \gamma(t) + (1-\alpha)\gamma(s),$$

d.h., γ ist konvex. □

Anmerkung 2.60 Ebenso wie die charakteristischen oder die momentenerzeugenden Funktionen enthalten auch die Kumulantenerzeugenden (oft *Log-Momentenerzeugende* genannt) Informationen über die Momente einer Zufallsvariable X . Konkret: Ist $\mathbf{E}[|X|^k] < \infty$ für alle $k \in \mathbf{N}$, so ist

$$\gamma(s) = s\mathbf{E}[X] + \frac{s^2}{2}\mathbf{V}[X] + \dots$$

Wie auch die charakteristische Funktion lässt sich γ für die meisten gerbäuchlichen Verteilungen ausrechnen, so z.B. für die

1. Bernoulliverteilung mit Parameter $p \in [0, 1]$: $\gamma(s) = \log(1 - p + pe^s)$,
2. Poissonverteilung mit Parameter $\lambda > 0$: $\gamma(s) = \lambda(e^s - 1)$,
3. Normalverteilung $\mathcal{N}(\mu, \sigma^2)$: $\gamma(s) = \mu s + \frac{\sigma^2 s^2}{2}$.

Definition 2.61 (Legendre-Fenchel-Transformierte) Sei $f: \mathbf{R} \rightarrow \mathbf{R}$. Die Legendre-Fenchel-Transformierte (auch: *konvex Konjugierte*) von f ist durch

$$f^*(u) = \sup_{x \in \mathbf{R}} \{xu - f(x)\}$$

definiert.

Lemma 2.62 Für die konvex Konjugierte f^* einer Funktion f gilt:

1. f^* ist konvex,
2. f^* hat ein eindeutiges Minimum.

Beweis. ÜA. □

Wir kommen nun zur zentralen Aussage dieses Abschnitts – dem Prinzip großer Abweichungen für IID Zufallsvariable.

Satz 2.63 (Cramérsches Theorem) Für alle $x \geq \mathbf{E}[X]$ gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) = -\gamma^*(x) = -\sup_{s \in \mathbf{R}} \{sx - \gamma(s)\}. \quad (2.10)$$

Beweis. O.B.d.A. sei $\mathbf{E}[X] = 0$. Der Beweise gliedert sich in 2 Teile.

1. Obere Schranke “ \leq ”: Wegen der Unabhängigkeit der X_i in der Summe $S_n = X_1 + \dots + X_n$ ist für alle $s \geq 0$

$$\begin{aligned}P(S_n \geq nx) &= \mathbf{E}[\chi_{\{S_n \geq nx\}}] \\ &\leq \mathbf{E}[e^{s(S_n - nx)}] \\ &= (\mathbf{E}[e^{sX}])^n e^{-snx} \\ &= e^{n\gamma(s) - snx}.\end{aligned}$$

Daraus folgt, dass

$$\frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \leq \gamma(s) - sx \leq -\sup_{s \geq 0} \{sx - \gamma(s)\}.$$

Im Supremum auf der rechten Seite dürfen wir die Bedingung $s \geq 0$ durch $s \in \mathbf{R}$ ersetzen, denn nach der Jensen-Ungleichung, Satz 1.45, gilt

$$\log \mathbf{E}[e^{sX}] \geq \log e^{\mathbf{E}[sX]} = s\mathbf{E}[X] = 0,$$

d.h. $\gamma(s) \geq 0$ mit $\gamma(0) = 0$ als dem globalen Minimum. Für $x \geq 0$ ist $\sup_{s < 0} \{sx - \gamma(s)\} = 0$, so dass

$$\sup_{s \geq 0} \{sx - \gamma(s)\} = \sup_{s \in \mathbf{R}} \{sx - \gamma(s)\}.$$

2. Untere Schranke “ \geq ”: Da $P(S_n/n \geq x) \geq P(S_n/n \in [x, x + \epsilon])$ für alle $\epsilon > 0$, reicht es zu zeigen, dass

$$\lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \in [x, x + \epsilon]\right) \geq -\gamma^*(x).$$

Wir nehmen zunächst an, dass es für die Gleichung $\gamma'(s) = x$ eine Lösung $s^*(x)$ gebe, so dass $\gamma^*(x) = s^*(x)x - \gamma(s^*(x))$ ist, und definieren ein neues Wahrscheinlichkeitsmaß P_s , das bezüglich P die Dichte

$$f(X_1, \dots, X_n) = e^{sS_n - n\gamma(s)}, \quad S_n = X_1 + \dots + X_n \tag{2.11}$$

habe. Der Erwartungswert von X bezüglich $dP_s = f dP$ ist damit

$$\begin{aligned} \int_{\Omega} X dP_s &= \mathbf{E}[X e^{sS_n - n\gamma(s)}] \\ &= \mathbf{E}[X e^{sX - \gamma(s)}] \underbrace{(\mathbf{E}[e^{sX - \gamma(s)}])^{n-1}}_{=1} \\ &= \frac{\mathbf{E}[X e^{sX}]}{\mathbf{E}[e^{sX}]} \\ &= \gamma'(s). \end{aligned}$$

Umgekehrt gilt für alle $\epsilon > 0$:

$$\begin{aligned} P\left(\frac{S_n}{n} \in [x, x + \epsilon]\right) &= \mathbf{E}_{P_s} \left[e^{-sS_n + n\gamma(s)} \chi_{\{S_n/n \in [x, x + \epsilon]\}} \right] \\ &= e^{n\gamma(s) - nsx} \mathbf{E}_{P_s} \left[e^{-sS_n + nsx} \chi_{\{S_n/n \in [x, x + \epsilon]\}} \right] \\ &= e^{n\gamma(s) - nsx} \mathbf{E}_{P_s} \left[e^{-ns(S_n/n - x)} \chi_{\{S_n/n \in [x, x + \epsilon]\}} \right] \\ &\geq e^{n(\gamma(s) - sx)} e^{-ns\epsilon} P_s \left(\frac{S_n}{n} \in [x, x + \epsilon] \right), \end{aligned}$$

wobei $\mathbf{E}_{P_s}[\cdot]$ den Erwartungswert bezüglich P_s bezeichnet und wir im letzten Schritt ausgenutzt haben, dass $S_n/n - x \leq \epsilon$ ist. Damit erhalten wir die Ungleichung

$$\frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \geq \gamma(s) - sx - |s|\epsilon + \frac{1}{n} \log P_s \left(\frac{S_n}{n} \geq x \right).$$

Nach Konstruktion von P_s ist $\mathbf{E}_{P_{s^*}}[X] = \gamma'(s^*(x)) = x$, und aus dem Zentralen Grenzwertsatz folgt, dass

$$\lim_{n \rightarrow \infty} P_{s^*} \left(\frac{S_n}{n} \geq [x, x + \epsilon] \right) = \frac{1}{2} \quad \forall \epsilon > 0.$$

Zusammenfassend erhalten wir

$$\lim_{\epsilon \rightarrow 0} \liminf_{n \rightarrow \infty} \frac{1}{n} \log P\left(\frac{S_n}{n} \geq x\right) \geq -\sup_{s \in \mathbf{R}} \{sx - \gamma(s)\} = -\gamma^*(x)$$

Die beiden Abschätzungen zusammengenommen ergeben schließlich die Behauptung. \square

Anmerkung 2.64 Nach dem eben bewiesenen Prinzip großer Abweichungen sind Ausreißer vom Gesetz der großen Zahlen “exponentiell unwahrscheinlich”. Eine geläufige Schreibweise ist dementsprechend

$$P\left(\frac{S_n}{n} \geq x\right) \asymp e^{-n\gamma^*(x)}.$$

Das Symbol “ \asymp ” deutet an, dass die Gleichung im Sinne von (2.10) zu lesen ist. Die Funktion γ^* , die den exponentiellen Abfall der Wahrscheinlichkeit angibt, wird *Ratenfunktion* genannt. Wegen der Konvexität von γ und der Definition der konvex Konjugierten ist γ^* an der Stelle $\tilde{x} = \mathbf{E}[X]$ minimal mit dem eindeutigen Minimum $\gamma^*(\tilde{x}) = 0$ (hier: $\tilde{x} = 0$), was gerade der Konvergenz gegen den Erwartungswert entspricht.

Beispiel 2.65 (Fortsetzung von Beispiel 2.56) Wir betrachten noch einmal das Problem des Schätzens kleiner Wahrscheinlichkeiten. Sei X eine Zufallsvariable mit $\mathbf{E}[X] = 0$ und $\{X \geq l\}$ selten unter der Wahrscheinlichkeit P . Anstatt $p = P(X \geq l)$ direkt aus unabhängigen Kopien der Zählvariable

$$Y = \chi_{\{X \geq l\}}$$

zu schätzen, definieren wir ein neues Wahrscheinlichkeitsmaß \tilde{P} , unter dem $\{X \geq l\}$ nicht mehr selten ist – idealerweise ist \tilde{P} so gewählt, dass $\mathbf{E}_{\tilde{P}}[X] = l$ ist. Ein geeigneter Kandidat ist $\tilde{P} = P_s$ aus Gleichung (2.11):

$$\frac{dP_s}{dP} = e^{sS_n - n\gamma(s)}, \quad S_n = X_1 + \dots + X_n.$$

Sei also $p_n = \mathbf{E}[\chi_{\{S_n/n \geq l\}}]$, was sich bezüglich P_s wie folgt schreiben lässt:

$$p_n = \mathbf{E}_{P_s}[\chi_{\{S_n/n \geq l\}} e^{n\gamma(s) - sS_n}].$$

Einen erwartungstreuen Schätzer für p_n erhalten wir, indem wir den Mittelwert über unabhängige Kopien von

$$Z = \chi_{\{S_n/n \geq l\}} e^{n\gamma(s) - sS_n}$$

mit $S_n = X_1 + \dots + X_n$ und $X_i \sim P_s$, $i = 1, \dots, n$ bilden. Der Parameter s kann dabei so gewählt werden, dass die Varianz des Schätzers (sprich: sein 2. Moment) minimal wird. Nun ist das 2. Moment von Z gerade durch

$$M_2^n(Z) = \mathbf{E}_{P_s}[\chi_{\{S_n/n \geq l\}} e^{2n\gamma(s) - 2sS_n}] \leq e^{-2n(sl - \gamma(s))}$$

gegeben, und $M_2^n(Z)$ wird minimal für $s = s^*(l)$, d.h. $\sup_s \{sl - \gamma(s)\} = \gamma^*(l)$. In diesem Fall ist

$$M_2^n(Z) \leq e^{-2n\gamma^*(l)}.$$

Andererseits gilt nach der Cauchy-Schwarz-Ungleichung und Satz 2.63

$$\begin{aligned} \mathbf{E}_{P_s}[\chi_{\{S_n/n \geq l\}} e^{2n\gamma(s) - 2sS_n}] &\geq \left(\mathbf{E}_{P_s}[\chi_{\{S_n/n \geq l\}} e^{n\gamma(s) - sS_n}] \right)^2 \\ &= \left(\mathbf{E}[\chi_{\{S_n/n \geq l\}}] \right)^2 \\ &= p_n^2 \\ &\asymp e^{-2n\gamma^*(l)}, \end{aligned}$$

woraus zusammen mit der oberen Schranke für das 2. Moment unseres neuen Schätzers folgt, dass

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log M_2^n = -2\gamma^*(l) = 2 \lim_{n \rightarrow \infty} \frac{1}{n} \log p_n.$$

Mit anderen Worten: $\sqrt{M_2^n} \asymp p_n$, d.h., der relative Schätzfehler bleibt auch im Falle $p_n \rightarrow 0$ beschränkt.

3 Abhängige Zufallsvariable: Markovketten

3.1 Bedingte Wahrscheinlichkeiten und Erwartungswerte

Im Folgenden sei (Ω, \mathcal{E}, P) wieder unser generischer Wahrscheinlichkeitsraum. Wir betrachten zwei Ereignisse $A, B \in \mathcal{E}$, wobei wir $P(A) > 0$ annehmen, und bezeichnen mit $\mathcal{E}_A \subset \mathcal{E}$ die σ -Algebra der Teilmengen von A , die in \mathcal{E} liegen.

Definition 3.1 (Bedingte Wahrscheinlichkeit) Seien $A, B \in \mathcal{E}$ mit $P(A) > 0$. Als bedingte Wahrscheinlichkeit von B unter der Bedingung A bezeichnen wir

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

Die bedingte Wahrscheinlichkeit $P_A(\cdot) := P(\cdot|A)$ definiert ein Maß auf \mathcal{E} , das auf A konzentriert ist, d.h., $P_A(A^c) = 0$. Für alle $B \in \mathcal{E}_A$ ist insbesondere $P_A(B) = P(B)/P(A)$.

Definition 3.2 (Bedingte Erwartung I) Es sei $X: \Omega \rightarrow \mathbf{R}$ eine Zufallsvariable, deren Erwartungswert existiert. Als bedingten Erwartungswert von X unter der Bedingung A definieren wir

$$\mathbf{E}[X|A] = \frac{\mathbf{E}[X\chi_A]}{P(A)}.$$

Nun ist nach obiger Definition

$$\mathbf{E}[X|A] = \frac{1}{P(A)} \int_A X(\omega) dP(\omega) = \int_{\Omega} X(\omega) dP_A(\omega),$$

insbesondere erhalten wir mit $X = \chi_B$ die bedingte Wahrscheinlichkeit $\mathbf{E}[\chi_B|A] = P(B|A)$. Insofern sollte es nicht überraschen, dass für die bedingten Erwartungswerte die für bedingte Wahrscheinlichkeiten üblichen Rechenregeln gelten; beispielsweise entspricht das Gesetz des totalen Erwartungswertes,

$$\mathbf{E}[X] = \mathbf{E}[X\chi_A] + \mathbf{E}[X\chi_{A^c}] = \mathbf{E}[X|A]P(A) + \mathbf{E}[X|A^c]P(A^c),$$

im wesentlichen dem Gesetz von der totalen Wahrscheinlichkeit,

$$P(B) = P(B \cap A) + P(B \cap A^c) = P(B|A)P(A) + P(B|A^c)P(A^c).$$

Wir werden uns weiter unten noch eingehender mit dem bedingten Erwartungswert beschäftigen. Zunächst wollen wir nur einige allgemeine Überlegungen anstellen, die den Fall $P(A) = 0$ betreffen. Konkret werden wir bedingende Ereignisse der Form $A = \{Y = y\}$ betrachten, die im Falle stetig verteilter Zufallsvariablen gerade Wahrscheinlichkeit 0 haben. Der diskrete Fall wird uns dabei als Blaupause dienen.

1. **Diskrete Zufallsvariable:** Wir nehmen zunächst an, dass Y diskret sei und definieren für jede integrierbare Zufallsvariable $X: \Omega \rightarrow \mathbf{R}$ durch

$$g: \mathbf{R} \rightarrow \mathbf{R}, \quad g(y) = \begin{cases} \mathbf{E}[X|Y = y], & P(Y = y) > 0 \\ 0 & \text{sonst.} \end{cases}$$

eine neue Zufallsvariable $Z = g(Y)$. Ferner bezeichnen wir mit \mathcal{F} die von den Bildern von Y erzeugte σ -Algebra. Dann steht die Messbarkeit von Z bezüglich \mathcal{F} stellvertretend für die Aussage, dass Z eine Funktion von y ist. Gleichzeitig gilt nach Konstruktion von Z , dass

$$\mathbf{E}[X\chi_{\{Y \in F\}}] = \mathbf{E}[Z\chi_{\{Y \in F\}}] \quad \forall F \in \mathcal{F}, \quad (3.1)$$

denn

$$\mathbf{E}[X\chi_{\{Y \in F\}}] = \sum_{y \in F, P(Y=y) > 0} \mathbf{E}[X\chi_{\{Y=y\}}] = \sum_{y \in F, P(Y=y) > 0} g(y)P(Y=y) = \mathbf{E}[Z\chi_{\{Y \in F\}}].$$

2. **Stetig verteilte Zufallsvariable:** Für stetig verteilte Zufallsvariable Y ist in der Regel $P(Y = y) = 0$, aber wir können die Messbarkeit der oben definierten Zufallsvariable Z zusammen mit (3.1) als *Definiens* für den bedingten Erwartungswert nehmen. Das geht, weil wir uns in unseren Überlegungen gar nicht auf die bedingende Zufallsvariable Y , sondern nur auf die von Y erzeugte σ -Algebra von bedingenden Ereignissen bezogen haben und motiviert die nachstehende Definition.

Definition 3.3 (Bedingte Erwartung II) Es sei $X: \Omega \rightarrow \mathbf{R}$ eine integrable Zufallsvariable auf (Ω, \mathcal{E}, P) und $\mathcal{F} \subset \mathcal{E}$ eine Sub- σ -Algebra. Dann nennen wir die reelle Zufallsvariable Z *bedingte Erwartung von X unter der Bedingung \mathcal{F}* (symbolisch: $\mathbf{E}[X|\mathcal{F}]$), wenn Z messbar bezüglich \mathcal{F} ist und

$$\mathbf{E}[X\chi_F] = \mathbf{E}[Z\chi_F] \quad \forall F \in \mathcal{F}.$$

Alle wesentlichen Eigenschaften wie die Linearität des Erwartungswertes oder Aussagen wie die verschiedenen Konvergenzsätze für Folgen von Zufallsvariablen gelten auch für die bedingten Erwartungswerte, allerdings gelten die Gleichungen nun, da sie sich auf Zufallsvariable beziehen, nur noch fast sicher. Bedingte Wahrscheinlichkeiten lassen sich wie schon zuvor mittels $P(B|\mathcal{F}) = \mathbf{E}[\chi_B|\mathcal{F}]$ für beliebige $B \in \mathcal{E}$ definieren.

Beispiel 3.4 (Bedingte Normalverteilung) Sei $X = (X_1, X_2) \in \mathbf{R}^n \times \mathbf{R}^m$ eine mehrdimensionale, normalverteilte Zufallsvariable mit Erwartungswert μ und Kovarianzmatrix $\Sigma = \Sigma^T > 0$,

$$\mu = (\mu_1, \mu_2) \in \mathbf{R}^n \times \mathbf{R}^m, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}^T & \Sigma_{22} \end{pmatrix} \in \mathbf{R}^{(n+m) \times (n+m)}.$$

Die gemeinsame Dichte von $X = (X_1, X_2)$ ist durch

$$f(x) = \det(2\pi\Sigma)^{-1/2} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right), \quad x = (x_1, x_2) \in \mathbf{R}^n \times \mathbf{R}^m$$

gegeben. Für festes $X_2 = y$ ist die bedingte Verteilung die Normalverteilung²⁸

$$P(\cdot | X_2 = y) = \mathcal{N}(\bar{\mu}_1, \Sigma_1)$$

mit dem bedingten Erwartungswert

$$\bar{\mu}_1(y) = \mu_1 - \Sigma_{12}\Sigma_{22}^{-1}(y - \mu_2)$$

und der bedingten Kovarianzmatrix

$$\Sigma_1 = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}^T,$$

die durch das Schurkomplement gegeben ist. Letzteres erbt die Definitheit von Σ , wie man durch

$$\begin{aligned} x^T \Sigma x &= x_1^T \Sigma_{11} x_1 + 2x_2 \Sigma_{12}^T x_1 + x_2^T \Sigma_{22} x_2 \\ &= (x_2 - \Sigma_{22}^{-1} \Sigma_{12}^T x_1)^T \Sigma_{22} (x_2 - \Sigma_{22}^{-1} \Sigma_{12}^T x_1) + x_1^T (\Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{12}^T) x_1 \end{aligned}$$

sieht; da $x^T \Sigma x > 0$ für alle $x \neq 0$ ist, gilt dies insbesondere auch für $x_2 = \Sigma_{22}^{-1} \Sigma_{12}^T x_1$, woraus folgt, dass auch der zweite Summand für alle $x_1 \neq 0$ positiv ist. Die bedingte Dichte von $P(\cdot | X_2 = y)$ bezüglich dx_1 ist damit

$$\bar{f}(x_1; y) = \det(2\pi\Sigma_1)^{-1/2} \exp\left(-\frac{1}{2}(x_1 - \bar{\mu}_1(y))^T \Sigma_1^{-1}(x_1 - \bar{\mu}_1(y))\right).$$

Wir belassen es zunächst bei diesem kurzen formalen Exkurs zum Thema *bedingte Erwartung* und werden uns zunächst den Markovketten zuwenden, bei denen bedingte Wahrscheinlichkeiten und Erwartungen eine wichtige Rolle spielen. Weitergehende Fragestellungen wie die nach Existenz und Eindeutigkeit der bedingten Erwartungswertes werden später im Abschnitt 4.1 wieder aufgegriffen werden.

²⁸ ÜA.

3.2 Markovketten auf abzählbarem Zustandsraum

Als einführendes Beispiel betrachten wir die Beobachtung des Wetters. Bekanntermaßen ist die Behauptung, dass das Wetter morgen so wie heute sein werde, eine recht gute Wettervorhersage (die Trefferquote beträgt angeblich über 60%). Wir nehmen, dass wir durch Beobachtungen herausgefunden haben, dass mit Wahrscheinlichkeit $p = 0.7$ auf einen Regentag ein Regentag folgt, während für Sonnentage die Wahrscheinlichkeit immerhin 0.6 beträgt. Da sich die (bedingten) Wahrscheinlichkeiten zu 1 aufsummieren, folgt mit Wahrscheinlichkeit 0.3 auf Regen Sonne und mit Wahrscheinlichkeit 0.4 auf Sonne Regen.

Mathematisch können wir unser Wettermodell modellieren, indem wir eine Zufallsvariable $X: \Omega \rightarrow S \subset \mathbf{R}$ über einem Wahrscheinlichkeitsraum (Ω, \mathcal{E}, P) einführen, deren Werte $X \in \{1, 2\}$ für Sonne ($X = 1$) bzw. Regen ($X = 2$) stehen. Entsprechend stehen die nicht notwendigerweise unabhängigen Kopien X_n mit $n = 0, 1, 2, \dots$ für das Wetter am n -ten Tag (der Tag $n = 0$ ist dabei ein beliebig gewähltes Datum, an dem beispielsweise die Sonne scheint). Die Regeln für die Wetterprognose bekommen damit die Form bedingter Wahrscheinlichkeiten,

$$p_{11} = P(X_{n+1} = 1 | X_n = 1), \quad p_{12} = P(X_{n+1} = 2 | X_n = 1), \dots,$$

wobei wegen der Normierung der Wahrscheinlichkeiten $p_{11} + p_{12} = p_{21} + p_{22} = 1$ gilt.

Definition 3.5 (Stochastische Matrix) Es sei S eine abzählbare Menge. Eine $|S| \times |S|$ -Matrix $P = (p_{ij})_{i,j \in S}$ heißt *stochastische Matrix*, wenn $p_{ij} \geq 0$ für alle $i, j \in S$ und $\sum_{j \in S} p_{ij} = 1$ für jedes $i \in S$.

Die Einträge einer stochastischen Matrix lassen sich als Übergangswahrscheinlichkeiten eines diskreten dynamischen Systems interpretieren:

Definition 3.6 (Markovkette) Es sei S eine abzählbare Menge, μ eine Zähldichte auf S und $P \in \mathbf{R}^{|S| \times |S|}$ eine stochastische Matrix. Eine Folge $(X_n)_{n \in \mathbf{N}_0}$ von Zufallsvariablen auf S heißt (*homogene*) *Markovkette mit Übergangsmatrix P und Anfangsverteilung μ* , falls

$$P(X_0 = x) = P_{X_0}(x) = \mu(x) \quad \forall x \in S$$

und für jedes $n \in \mathbf{N}_0$, $y \in S$ sowie alle Tupel $(x_0, x_1, \dots, x_n) \in S$ mit $P(X_0 = x_0, \dots, X_n = x_n) > 0$ die Markoveigenschaft gilt:

$$P(X_{n+1} = y | X_0 = x_0, \dots, X_n = x_n) = P(X_{n+1} = y | X_n = x_n) = p_{xy}. \quad (3.2)$$

Im Gegensatz zu den bei den im letzten Kapitel betrachteten Folgen von Zufallsvariablen, sind die Folgenglieder einer Markovkette nicht unabhängig, wobei sich die Abhängigkeit aus den Übergangswahrscheinlichkeiten p_{xy} ergibt. Die Markoveigenschaft besagt gerade, dass X_{n+1} nur von X_n abhängt, nicht aber von X_{n-1}, \dots, X_0 . Da die Übergangswahrscheinlichkeiten p_{xy} nicht von n abhängen, wird die Markovkette *homogen* genannt.

Den Laufindex der X_0, X_1, X_2, \dots werden wir im Folgenden als Zeit interpretieren, entsprechend $(X_n)_{n \in \mathbf{N}_0}$ als stochastischen Prozess, d.h., eine diskrete oder kontinuierliche Familie von Zufallsvariablen, die bestimmte Messbarkeitseigenschaften erfüllt. Konkret stattdessen wir den Zustandsraum S mit seiner natürlichen σ -Algebra $\mathcal{F} = 2^S$ aus. Entsprechend ist μ ein Wahrscheinlichkeitsmaß auf (S, \mathcal{F}) , das sich induktiv mittels

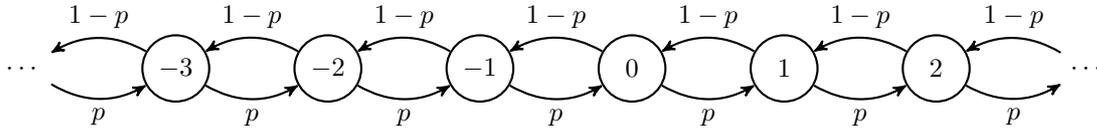
$$P_\mu(X_0 = x_0, X_1 = x_1, \dots, X_N = x_N) = \mu(x_0) \prod_{n=0}^{N-1} P(X_{n+1} = x_{n+1} | X_n = x_n)$$

zu einem Wahrscheinlichkeitsmaß auf dem Produktmessraum $(S^{N+1}, \mathcal{F}^{N+1})$ erweitern lässt, das die Verteilung der Trajektorie (X_0, X_1, \dots, X_N) bis zum Zeitpunkt N angibt. Da die Verteilungen der Trajektorien einer bestimmten Länge eine konsistente Familie von Wahrscheinlichkeitsmaßen bilden, garantiert der Satz von Ionescu-Tulcea [9, Satz 14.32], dass sogar ein eindeutiges Wahrscheinlichkeitsmaßes auf dem Produkt-Messraum $(S^\infty, \mathcal{F}^\infty)$ existiert, das die Verteilung der Kette $(X_n)_{n \in \mathbf{N}_0}$ zu allen Zeiten angibt. (Der Konsistenzsatz von Kolmogorov aus Abschnitt 2.5 gilt nur für Maße auf Borelmengen.)

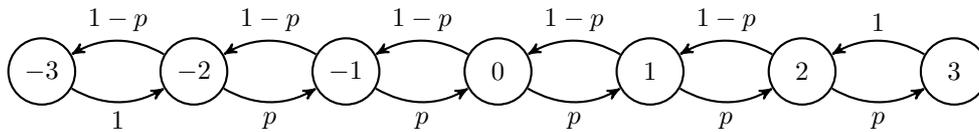
Satz 3.7 (Existenz einer Markovkette) Sei S abzählbar, P eine stochastische Matrix und μ ein Wahrscheinlichkeitsmaß auf S (bzw. eine Zähldichte). Dann existiert eine Folge $(X_n)_{n \in \mathbf{N}_0}$ von Zufallsvariablen auf S , die die Markoveigenschaft erfüllen, und für die $P_{X_0} = \mu$ ist.

Beweis. Für den Beweis verweisen wir auf [10, Anhang A.2 bzw. Satz A.11]. □

Beispiel 3.8 (Zufallsspaziergang auf \mathbf{Z}) Sei $S = \mathbf{Z}$ und P die (unendliche) Matrix mit den Einträgen $p_{i,i+1} = p$ und $p_{i,i-1} = 1 - p$ und 0 sonst. Wenn wir $\mu = \delta_0$ setzen, so ist $(X_n)_{n \in \mathbf{N}}$ eine Markovkette, die fast sicher in $X_0 = 0$ startet und die mit Wahrscheinlichkeit p bzw. $1 - p$ einen Schritt nach rechts bzw. links macht.



Beispiel 3.9 (Eingeschränkter Zufallsspaziergang) Wir betrachten die gleiche Situation wie eben, aber schränken X_n auf die Menge $S = \{-3, \dots, 3\}$ ein. Dazu definieren wir folgende Regel: Wenn die Kette auf den Rand bei $x = \pm 3$ trifft, so soll der Prozess im nächsten Schritt mit Wahrscheinlichkeit 1 zum benachbarten inneren Zustand zurückkehren. Dargestellt als Graph sieht das dann so aus:



Man spricht in diesem Fall von reflektierenden Randbedingungen. Analog lassen sich endliche Markovketten mit periodischen ($p_{\pm 3, \mp 3} \neq 0$) bzw. absorbierenden ($p_{\pm 3, \pm 3} = 1$) Randbedingungen konstruieren.

Simulation von Markovketten I

Wie simuliert man eine Markovkette? Um eine konkrete Realisierung X_1, X_2, \dots zu erzeugen, muss in jedem Schritt ausgehend von $X_n = x_n$ entsprechend der Übergangswahrscheinlichkeiten $P(\cdot | X_n = x_n)$ ein $X_{n+1} \in S$ ausgewählt werden. Im letzten Beispiels bietet es sich an, in jedem Schritt eine uniform verteilte Zufallszahl $u \in [0, 1]$ auszuwürfeln und dann für $u \in [0, p)$ einen Schritt nach rechts zu machen bzw. einen Schritt nach links, wenn $u \in (p, 1]$ liegt (bei $u = p$ darf getrost eine Münze geworfen werden).

Satz 3.10 (Kanonische Darstellung von Markovketten) *Es sei $(\xi_n)_{n \in \mathbf{N}}$ eine Folge von IID Zufallsvariablen mit Werten in $W \subset \mathbf{R}$, die unabhängig von $X_0: \Omega \rightarrow S$ ist. Für alle $f: S \times W \rightarrow S$ definiert die Iteration*

$$X_{n+1} = f(X_n, \xi_{n+1})$$

eine homogene Markovkette auf S mit Übergangswahrscheinlichkeiten $p_{xy} = P(f(x, \xi_1) = y)$.

Beweis. ÜA. □

Beispiel 3.11 (Zufallsspaziergang auf \mathbf{Z}) Sei ξ_n gleichverteilt auf $W = \{-1, +1\}$ und $f(x, y) = x + y$. Die zugehörige Folge $X_{n+1} = X_n + \xi_{n+1}$ ist der symmetrische Zufallsspaziergang aus Beispiel 3.8.

Anmerkung 3.12 Alle homogenen Markovketten mit Übergangsmatrix $P = (p_{xy})_{x,y \in S}$ besitzen eine kanonische Darstellung der Form $X_{n+1} = f(X_n, U_{n+1})$ mit

$$U_n \sim \mathcal{U}(0, 1) \quad \text{und} \quad f(x, u) = y, \quad \text{falls} \quad \sum_{z=1}^{y-1} p_{xz} \leq u < \sum_{z=1}^y p_{xz}.$$

Simulation von Markovketten II

Der eben beschriebene Weg ist nicht die einzige Möglichkeit, eine Markovkette $(X_n)_{n \in \mathbf{N}}$ zu simulieren. Anstatt Realisierungen x_0, x_1, x_2, \dots mit einem aus der Startverteilung μ gezogenen Anfangswert x_0 zu erzeugen, hätten wir ebenso die Verteilung der X_1, X_2, \dots direkt simulieren können (siehe Abb. 7).

Zu diesem Zweck bezeichnen wir mit μ_k die Verteilung der Kette zum Zeitpunkt $k \geq 0$, d.h., $\mu_k(A) = P_\mu(X_k \in A)$ für jedes $A \in \mathcal{F}$; die Bezeichnung $P_\mu(\cdot)$ gibt dabei an, dass die Kette mit Verteilung $\mu_0 = \mu$ gestartet wurde.²⁹

²⁹ Analog schreiben wir $P_x(\cdot)$, wenn $\mu = \delta_x$, d.h., wenn die Kette fast sicher bei $X_0 = x$ gestartet wurde.

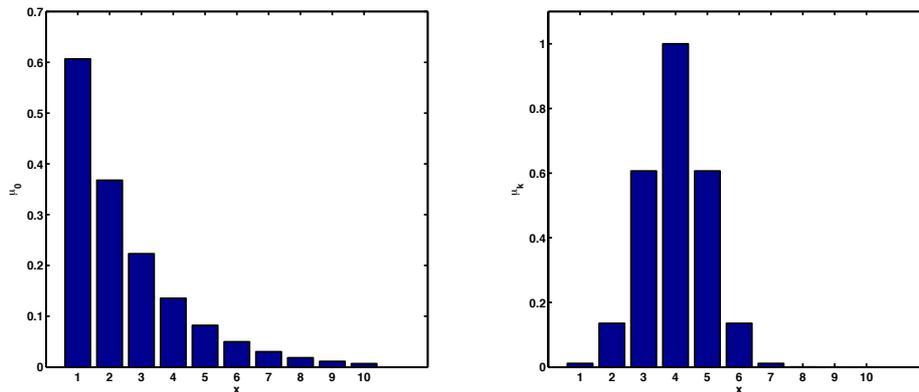


Abb. 7 Verteilung μ_n einer Markovkette mit $|S| = 10$ Zuständen zu verschiedenen Zeitpunkten $n \in \mathbf{N}_0$.

Lemma 3.13 (Propagation der Startverteilung) Sei $(X_n)_{n \in \mathbf{N}_0}$ eine Markovkette mit Übergangsmatrix $P = (p_{xy})_{x,y \in S}$ und Anfangsverteilung μ . Für die Verteilung μ_k der Kette zum Zeitpunkt $k \geq 0$ gilt

$$\mu_k(y) = \sum_{x \in S} \mu_{k-1}(x)p_{xy} \tag{3.3}$$

bzw.

$$\mu_k(y) = \sum_{x \in S} \mu(x)p_{xy}^{(k)}, \tag{3.4}$$

wobei $p_{xy}^{(k)}$ die Einträge der iterierten Übergangsmatrix, P^k , bezeichnet.

Beweis. Nach dem Satz von der totalen Wahrscheinlichkeit ist

$$\begin{aligned} P_\mu(X_k = y) &= \sum_{x \in S} P(X_k = y | X_{k-1} = x) P_\mu(X_{k-1} = x) \\ &= \sum_{x \in S} \mu_{k-1}(x)p_{xy}. \end{aligned}$$

Der zweite Teil der Aussage folgt durch Induktion. □

Anmerkung 3.14 Wir vereinbaren, dass $\mu \in \mathbf{R}^{|S|}$ ein Spaltenvektor ist. In Matrix-Vektor-Schreibweise lässt sich die Iteration für die Verteilung dann wie folgt schreiben:

$$\mu_k^T = \mu_{k-1}^T P \quad \text{bzw.} \quad \mu_k^T = \mu_0^T P^k \quad (\mu_0 = \mu).$$

Die Einträge von P^k geben die Wahrscheinlichkeit an, in k Schritten von x nach y zu gelangen, d.h.

$$p_{xy}^{(k)} = P(X_k = y | X_0 = x).$$

Damit ist insbesondere P^k für alle $k \in \mathbf{N}_0$ (mit der Vereinbarung $P^0 = \text{Id}$) wieder eine stochastische Matrix. Eine wichtige Eigenschaft der iterierten Übergangsmatrix ist ihre Halbgruppeneigenschaft, die vor allem unter dem Namen *Chapman-Kolmogorov-Gleichung* bekannt ist.³⁰

Satz 3.15 (Chapman-Kolmogorov-Gleichung) Sei $(X_n)_{n \in \mathbf{N}_0}$ eine homogene Markovkette mit Übergangsmatrix $P = (p_{xy})_{x,y \in S}$. Für alle $n, m \in \mathbf{N}_0$ gilt

$$p_{xy}^{(n+m)} = \sum_{z \in S} p_{xz}^{(n)} p_{zy}^{(m)} = \sum_{z \in S} p_{xz}^{(m)} p_{zy}^{(n)} \tag{3.5}$$

³⁰ Sidney Chapman, FRS (1888–1970), engl. Mathematiker und Geophysiker

Beweis. Es ist

$$\begin{aligned} P(X_{n+m} = y | X_0 = x) &= \sum_{z \in S} P(X_{n+m} = y, X_n = z | X_0 = x) \\ &= \sum_{z \in S} P(X_{n+m} = y | X_n = z, X_0 = x) P(X_n = z | X_0 = x) \\ &= \sum_{z \in S} P(X_{n+m} = y | X_n = z) P(X_n = z | X_0 = x) \\ &= \sum_{z \in S} p_{xz}^{(n)} p_{zy}^{(m)}, \end{aligned}$$

wobei wir in der zweiten Zeile die Definition der bedingten Wahrscheinlichkeit und in der dritten die Markoveigenschaft der Kette ausgenutzt haben; der letzte Schritt ergibt sich aus der Homogenität der Kette. Da das Argument gleichermaßen mit $X_m = z$ funktioniert, ist die Gleichung symmetrisch in n und m . \square

Anmerkung 3.16 Anschaulich bedeutet die Halbgruppeneigenschaft, dass die Wahrscheinlichkeit dafür, dass die Kette in $n+m$ Schritten von x nach y gelangt, die gleiche ist wie die Wahrscheinlichkeit, zunächst in n (bzw. m) Schritten zu irgendeinem Zustand z und anschließend in m (bzw. n) Schritten zu y zu laufen.

Wir erwähnen zwei wichtige Erweiterungen der Markoveigenschaft.

1. **Kontinuierlicher Zustandsraum:** Eine naheliegende Verallgemeinerung sind Markovketten $(X_n)_{n \in \mathbf{N}_0}$ auf einem überabzählbaren Zustandsraum $Z \subset \mathbf{R}$ mit Ein-Schritt-Übergangswahrscheinlichkeiten

$$p(x, A) = P(X_{n+1} \in A | X_n = x) = \int_A q(x, y) dy$$

und Startverteilung

$$\mu_0(A) = P(X_0 \in A) = P_{X_0}(A)$$

Dabei ist $A \subset Z$ eine Borelmenge und $q(x, y)$ die Lebesgue-dichte von $p(x, \cdot)$ (von der wir stillschweigend annehmen, dass sie existiert). Für die Mehrschritt-Übergangswahrscheinlichkeiten

$$p_k(x, A) = P(X_{n+k} \in A | X_n = x)$$

gilt analog zum diskreten Fall, Gleichung (3.5), die Chapman-Kolmogorov-Gleichung

$$\begin{aligned} p_{n+m}(x, A) &= \int_A \left(\int_Z q_n(x, z) q_m(z, y) dz \right) dy \\ &= \int_A \left(\int_Z q_m(x, z) q_n(z, y) dz \right) dy. \end{aligned}$$

Auf die gleiche Art und Weise lassen sich die Ausdrücke (3.3) bzw. (3.4) für die Vorwärtspropagation der Verteilung μ_k der Kette zum Zeitpunkt k schreiben:

$$\mu_{k+1}(A) = \int_Z p(z, A) d\mu_k(z)$$

bzw.

$$\mu_{k+1}(A) = \int_Z p_{k+1}(z, A) d\mu_0(z).$$

Beispiel 3.17 (Zufallsspaziergang auf \mathbf{R}) Wir betrachten eine Markovkette $(X_n)_{n \in \mathbf{N}_0}$, die durch

$$X_{n+1} = X_n + \sqrt{\Delta t} \xi_{n+1}, \quad X_0 = 0$$

mit IID Zufallsvariablen $\xi_n \sim \mathcal{N}(0, 1)$ gegeben ist. Wie wir aus Beispiel 2.45 wissen, konvergiert X_n für $n \rightarrow \infty$ und $\Delta t \rightarrow 0$ unter der Bedingung $n\Delta t_n \rightarrow t$ nach dem Zentralen Grenzwertsatz schwach gegen die Brownsche Bewegung mit Verteilung $\mathcal{N}(0, t)$. Die Dichte der Übergangswahrscheinlichkeiten ist

$$q_{\Delta t}(x, y) = \frac{1}{\sqrt{2\pi\Delta t}} \exp\left(-\frac{(y-x)^2}{2\Delta t}\right).$$

Unsere Markovkette stellt damit eine zeitlich diskrete Approximation der Brownschen Bewegung dar.

2. **Gestoppte Markovketten:** Die für unsere Zwecke wichtigste Verallgemeinerung betrifft die Erweiterung der Markoveigenschaft auf Stoppzeiten. Dazu bezeichne $\Omega = S^\infty$ die Menge aller Folgen $(X_n)_{n \in \mathbf{N}_0}$ von Zufallsvariablen auf S , die wir mit ihrer natürlichen σ -Algebra \mathcal{F} ausstatten (vgl. Abschnitt 2.5). Eine Funktion $\tau: \Omega \rightarrow \mathbf{N}_0 \cup \{+\infty\}$ heißt *Stoppzeit*, wenn das Ereignis $\{\omega: \tau(\omega) \leq n\}$ für alle $n \in \mathbf{N}_0$ messbar bezüglich \mathcal{F}_n ist, wobei $\mathcal{F}_n = \sigma(\{X_k: k \leq n\})$ die von den ersten $n+1$ Folgengliedern erzeugte σ -Algebra ist.³¹ Anschaulich bedeutet das, dass über das Eintreten des Ereignisses $\{\tau \leq n\}$ allein aufgrund der Kenntnis von X_0, \dots, X_n entschieden werden kann. Beispiele für Stoppzeiten sind deterministische Stoppzeiten

$$\tau = n$$

für festes $n \in \mathbf{N}_0$ oder die Eintrittszeit in eine Menge $A \subset S$,

$$\tau_A = \inf\{n \geq 0: X_n \in A\},$$

wobei wir $\inf \emptyset = \infty$ vereinbaren, d.h. $\tau_A = \infty$, falls A niemals erreicht wird. Keine Stoppzeit ist dagegen

$$T_A = \sup\{n \geq 0: X_n \in A\},$$

die Zeit des letzten Besuchs von $A \subset S$, denn offenbar hängt das Ereignis $\{T_A = n\}$ nicht nur von X_0, \dots, X_n ab. (Um zu entscheiden, ob $T_A = n$ ist, müssen auch X_{n+1}, X_{n+2}, \dots bekannt sein.) Die für uns im Folgenden wichtigste Stoppzeit ist die Rückkehr- oder Wiederkehrzeit eines Zustands $z \in S$,

$$\tau_z = \inf\{n \geq 1: X_n = z\}.$$

Man beachte, dass die Wiederkehrzeit anders als die Eintrittszeit ≥ 1 ist, d.h., aus $X_0 = z$ folgt nicht $\tau_z = 0$. Eine praktische Eigenschaft von Markovketten ist nun, dass die Markoveigenschaft auch für Stoppzeiten gilt. Sei also $(X_n)_{n \in \mathbf{N}_0}$ eine Markovkette auf S . Mit

$$\mathcal{F}_\tau = \{A: A \in \mathcal{F} \text{ und } A \cap \{\tau = n\} \in \mathcal{F}_n \text{ für jedes } n \geq 0\}$$

bezeichnen wir die σ -Algebra der τ -Vergangenheit der Kette. Der nachfolgende Satz erweitert die Markoveigenschaft von festen Zeiten $n \in \mathbf{N}_0$ auf Stoppzeiten $\tau(\omega)$, die Zufallsvariable sind.

Satz 3.18 (Starke Markoveigenschaft) *Für alle Stoppzeiten τ mit $P(\tau < \infty) = 1$ erfüllt $(X_n)_{n \in \mathbf{N}_0}$ die Markoveigenschaft, d.h., die bedingte Verteilung von $X_{\tau+1}, X_{\tau+2}, \dots$ für gegebenes \mathcal{F}_τ mit $X_\tau = x$ entspricht der Verteilung der Kette X_1, X_2, \dots mit Startwert $X_0 = x$. mit anderen Worten:*

$$P(X_{\tau+1} = x_1, X_{\tau+2} = x_2, \dots | X_\tau = x) = P_x(X_1 = x_1, X_2 = x_2, \dots) \quad \forall (x_1, x_2, \dots) \in S^\infty.$$

Beweis. Wir betrachten alle möglichen Ereignisse $\{\tau = n\}$, $n \geq 0$, die eintreten können, und wenden die gewöhnliche Markoveigenschaft (3.2) an. Es gilt:

$$\begin{aligned} P(X_{\tau+1} = x_1, \dots | X_\tau = x) &= \sum_{n=0}^{\infty} P(X_{\tau+1} = x_1, X_{\tau+2} = x_2, \dots | X_\tau = x) P(\tau = n | X_\tau = x) \\ &= \sum_{n=0}^{\infty} P(X_{n+1} = x_1, X_{n+2} = x_2, \dots | X_n = x) P(\tau = n | X_\tau = x) \\ &= P(X_1 = x_1, X_2 = x_2, \dots | X_0 = x) \sum_{n=0}^{\infty} P(\tau = n | X_\tau = x) \\ &= P_x(X_1 = x_1, X_2 = x_2, \dots), \end{aligned}$$

wobei wir im vorletzten Schritt die Homogenität der Kette und im letzten Schritt die Tatsache ausgenutzt haben, dass $\sum_n P(\tau = n | \cdot) = 1$. \square

Anmerkung 3.19 Die starke Markoveigenschaft besagt, dass, wenn $X_\tau = x$ ist, die Kette $(X_n)_{n \geq \tau}$ ununterscheidbar von $(X_n)_{n \geq 0}$ mit Startverteilung δ_x ist; insbesondere ist $(X_n)_{n \geq \tau}$ unabhängig von \mathcal{F}_τ .

³¹ Die von endlich vielen Folgengliedern erzeugte σ -Algebra \mathcal{F}_n lässt sich analog zu den Überlegungen in Abschnitt 2.5 aus dem durchschnittstabilen System der Zylindermengen $F_{E,n} = \{\text{alle Folgen } \omega: (X_0, \dots, X_n) \in E \subset S^{n+1}\}$ erzeugen.

3.3 Stationäre Verteilungen von Markovketten

Es sei nun wieder $(X_n)_{n \in \mathbf{N}_0}$ unsere homogene Markovkette mit Übergangsmatrix $P \in \mathbf{R}^{|S| \times |S|}$ und Startverteilung μ_0 . Wir haben uns bereits in Lemma 3.13 klargemacht, dass die Verteilung der Kette zu einem Zeitpunkt k (sprich: die Verteilung von X_k) iterativ durch Linksmultiplikation an die Übergangsmatrix,

$$\mu_{k+1}^T = \mu_k^T P,$$

aus der Startverteilung berechnet werden kann. Da liegt es nahe, zu fragen, ob die Iteration der μ_k Fixpunkte besitzt oder ob die Iteration für $k \rightarrow \infty$ gegen einen Fixpunkt konvergiert. Solche Fixpunkte werden (so sie denn existieren) aus offensichtlichen Gründen *stationäre Verteilungen* oder *invariante Maße* genannt.

Definition 3.20 (Stationäre Verteilung) Sei $(X_n)_{n \in \mathbf{N}}$ eine Markovkette mit Übergangsmatrix P . Ein Wahrscheinlichkeitsmaß π heißt *stationär*, falls $\pi^T = \pi^T P$, d.h., wenn $\pi(y) = P_\pi(X_1 = y)$.

Anmerkung 3.21 Im Allgemeinen kann eine Markovkette mehrere invariante Verteilungen haben oder auch gar keine. Ein Vertreter der ersten Kategorie ist eine Markovkette mit der (zugegebenermaßen recht uninteressanten) Übergangsmatrix $P = \text{Id}$, für die die Eigenwertgleichung $\pi^T = \pi^T P$ offensichtlich unendlich viele nichtnegative Lösungen hat, die sich alle auf $\|\pi\|_1 = 1$ normieren lassen. Hingegen hat die (unendliche) Übergangsmatrix des symmetrischen Zufallsspaziergangs auf \mathbf{Z} als Linkseigenvektor zum Eigenwert 1 die bis auf Skalierung eindeutige Lösung $\pi^T = (1, 1, 1, \dots)$, die jedoch nicht normierbar und mithin kein Wahrscheinlichkeitsmaß ist.

Neben der bereits erwähnten Interpretation der stationären Verteilung als (normierterer) Linkseigenvektor von P zum Eigenwert 1 lassen sich invariante Maße mit der Erhaltung von Wahrscheinlichkeitsflüssen zwischen Teilmengen des Zustandsraums in Verbindung bringen. Konkret sei $(X_n)_{n \in \mathbf{N}_0}$ eine Markovkette auf S mit stationärer Verteilung π ; für zwei Mengen $A, B \subset S$ heißt

$$J_\pi(A, B) = P_\pi(X_1 \in B, X_0 \in A)$$

der *Wahrscheinlichkeitsfluss von A nach B*, wobei

$$\begin{aligned} P_\pi(X_1 \in B, X_0 \in A) &= \sum_{x \in A} \pi(x) P(X_1 \in B | X_0 = x) \\ &= \sum_{x \in A} \sum_{y \in B} \pi(x) p_{xy}. \end{aligned}$$

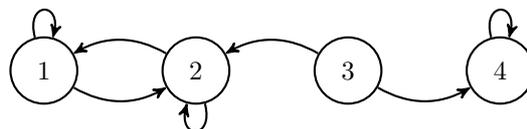
Satz 3.22 (Stationäres Flussgleichgewicht) Sei $(X_n)_{n \in \mathbf{N}_0}$ eine Markovkette mit stationärer Verteilung π . Dann gilt für beliebige $A \subset S$, dass

$$J_\pi(A, A^c) = J_\pi(A^c, A).$$

Beweis. ÜA. □

Kommunikationsklassen

Wir betrachten die folgende symbolische Markovkette mit $|S| = 4$ Zuständen; die Pfeile zwischen einem Zustand i und einem Zustand j deuten an, dass $p_{ij} > 0$ ist.



Offenbar kann man zwischen den Zuständen $\{1, 2\}$ beliebig hin und her wechseln, ohne jedoch die Zustände $\{3, 4\}$ erreichen zu können. Auch von Zustand $\{4\}$ gibt es keine Möglichkeit, in $\{1, 2, 3\}$ zu wechseln (Zustand $\{4\}$ ist absorbierend). Zustand $\{3\}$ hat fast sicher die Wiederkehrzeit $\tau_3 = \infty$ und wird deshalb, nachdem die Kette ihn einmal besucht hat, nie wieder erreicht. Unsere Beispielkette zerfällt demnach in zwei Teilmengen $C_1 = \{1, 2\}$ und $C_2 = \{4\}$, die nicht wieder verlassen werden können, und eine Menge $T = \{3\}$, von der aus man in eine der beiden ersten springen kann. Die Mengen $C_1, C_2 \subset S$ werden *abgeschlossen* genannt, T heißt *transient*.

Definition 3.23 (kommunizierend, irreduzibel) 2 Zustände $x, y \in S$ heißen *kommunizierend* (symbolisch: $x \leftrightarrow y$), falls $p_{xy}^{(n)}, p_{yx}^{(m)} > 0$ für zwei $n, m \in \mathbf{N}_0$. Sind alle Zustände kommunizierend, so heißt die Kette *irreduzibel*.

Lemma 3.24 *Kommunikation ist eine Äquivalenzrelation auf S .*

Beweis. ÜA. □

Die Äquivalenzklassen von \leftrightarrow werden *Kommunikationsklassen* genannt. Da insbesondere $p_{xx}^{(0)} = 1$ für alle $x \in S$ gilt, kommuniziert jeder Zustand mit sich selbst und bildet damit *per se* eine Kommunikationsklasse.

Rekurrenente und transiente Zustände

Wir werden nun die Zustandsklassifikation aus dem obigen Beispiel präzisieren.

Definition 3.25 (rekurrent, transient) Es sei $\tau_z = \inf\{n \geq 1: X_n = z\}$ die Rückkehrzeit nach $z \in S$. Der Zustand z heißt *rekurrent*, wenn $P_z(\tau_z < \infty) = 1$; wenn $P_z(\tau_z < \infty) < 1$, so wird z *transient* genannt.

Anmerkung 3.26 Ein Zustand $z \in S$ ist entweder transient oder rekurrent, denn

$$P_z(\tau_z < \infty) = \sum_{n=1}^{\infty} P_z(\tau_z = n) = 1 - P_z(\tau_z = \infty).$$

Insbesondere gilt, dass wenn zwei Zustände x, y derselben Kommunikationsklasse angehören, d.h. $x \leftrightarrow y$, so sind x und y entweder beide transient oder beide rekurrent. Rekurrenz und Transienz sind somit Klasseneigenschaften.

Anmerkung 3.27 Rekurrente Klassen sind abgeschlossen. Ist also $R \subset S$ eine rekurrente Klasse, dann gilt für alle $x \in R$ und $y \in R^c = S \setminus R$, dass

$$P(X_n = y | X_0 = x) = 0 \quad \forall n \in \mathbf{N}.$$

Anmerkung 3.28 Jede Markovkette lässt sich eindeutig in eine Menge transienter Zustände und disjunkte irreduzible, rekurrente Klassen zerlegen,

$$S = T \cup R_1 \cup \dots \cup R_m, \quad R_i \cap R_j = \emptyset,$$

wobei für die transienten Zustände gilt:

$$\lim_{n \rightarrow \infty} p_{xx}^{(n)} = 0, \quad x \in T.$$

Wir werden uns gleich der Frage zuwenden, unter welchen Bedingungen Markovketten eine eindeutige stationäre Verteilung $\pi > 0$ haben. Da sich jede Markovkette eindeutig in eine Menge transienter Zustände und disjunkte irreduzible, rekurrente Klassen zerlegen lässt, wobei für die transienten Zustände $y \in S$ gilt, dass $\pi(y) = 0$, beschränken wir uns von vornherein auf irreduzible, rekurrente Markovketten. Dass für transiente Zustände $\pi(y) = 0$ gilt, sieht man daran, dass für alle $n \in \mathbf{N}_0$ gilt, dass

$$\pi(y) = \sum_{x \in S} \pi(x) p_{xy}^{(n)}$$

und somit

$$\pi(y) = \lim_{n \rightarrow \infty} \sum_{x \in S} \pi(x) p_{xy}^{(n)} = \sum_{x \in S} \pi(x) \lim_{n \rightarrow \infty} p_{xy}^{(n)} = 0.$$

Die Vertauschbarkeit von Limes und Summation erlaubt hierbei der Satz der monotonen Konvergenz; die Konvergenz $p_{xy}^{(n)} \rightarrow 0$ folgt aus der Tatsache, dass y transient ist.

Existenz und Eindeutigkeit einer stationären Verteilung

Dass die Eigenwertgleichung $P^T \pi = \pi$ immer eine Lösung hat, folgt aus der Tatsache, dass wegen der Zeilensummeneigenschaft stochastischer Matrizen der Einsvektor $\mathbf{1} = (1, \dots, 1)^T$ immer ein Eigenvektor von P zum Eigenwert $\lambda = 1$ ist und dass die Matrix P und ihre Transponierte, P^T , dasselbe Spektrum haben, denn wegen

$$\det(P - \lambda I) = \det((P - \lambda I)^T) = \det(P^T - \lambda I)$$

haben P und P^T dasselbe charakteristische Polynom. (Mit I wird hier die Einheitsmatrix bezeichnet.)

Mit Hilfe eines ähnlichen Arguments zeigt man, dass der Eigenwert $\lambda = 1$ einfach ist, wenn die Matrix P irreduzibel ist (sprich: wenn P die Übergangsmatrix einer irreduziblen Markovkette ist). Dazu nehmen wir an, mit η würde ein weiterer Eigenvektor von P^T zum Eigenwert $\lambda = 1$ existieren, der kein Vielfaches von π ist. Wir setzen $\xi = \eta - c\pi$, wobei $c \in \mathbf{R}$ so gewählt wird, dass $\xi \geq 0$, wobei $\xi(z) = 0$ für ein $z \in S$. Wegen der Irreduzibilität existiert ein $k = k(y)$ mit $y \neq z$, so dass $p_{yz}^{(k)} > 0$; aus der nicht-Negativität von ξ und

$$0 = \xi(z) = \sum_{x \in S} \xi(x) p_{xz}^{(k)} \geq \pi(y) p_{yz}^{(k)} \geq 0,$$

folgt dann, dass $\pi(y) = 0$ ist. Wiederholtes Anwenden des Arguments zeigt, dass $\xi = 0$, d.h., η ist ein Vielfaches von π . Damit ist π eindeutig.³² Wir wollen nun untersuchen, wann die stationäre Verteilung sogar strikt positiv ist und notwendige Bedingungen für die Existenz und Eindeutigkeit einer stationären Verteilung $\pi > 0$ formulieren.

Definition 3.29 (positiv rekurrent, null-rekurrent) Ein rekurrenter Zustand $z \in S$ heißt:

1. *positiv rekurrent*, falls $\mathbf{E}_z[\tau_z] = \sum_{n \geq 1} n P_z(\tau_z = n) < \infty$,
2. *null-rekurrent*, falls $\mathbf{E}_z[\tau_z] = \infty$.

Lemma 3.30 *Positive Rekurrenz ist eine Klasseneigenschaft, d.h., ist ein Zustand einer irreduziblen Markovkette positiv rekurrent, so sind alle Zustände positiv rekurrent.*

Beweis. Seien $x, y \in S$ zwei beliebige Zustände, von denen x positiv rekurrent ist. Wir wollen zeigen, dass dann auch y positiv rekurrent ist und erinnern zunächst daran, dass Rekurrenz eine Klasseneigenschaft ist. Wegen der Irreduzibilität können wir annehmen, dass y mit positiver Wahrscheinlichkeit besucht wird, wenn die Kette in x gestartet wird, d.h., $P_x(\tau_y < \infty) > 0$. Wegen der starken Markoveigenschaft der Kette gilt dann, dass auch die Wahrscheinlichkeit, dass y besucht wird, bevor die Kette zu x zurückkehrt positiv ist, also $P_x(\tau_y < \tau_x) > 0$; wäre dies nicht der Fall, würde y niemals besucht, denn die Kette ist nach der ersten Rückkehr zu x von der Kette, die bei $n = 0$ gestartet wurde, ununterscheidbar. Sei also $p = P_x(\tau_y < \tau_x) > 0$. Aufgrund der Markov- bzw. Tschebyscheff-Ungleichung für den Betrag gilt, dass für alle Realisierungen unserer Kette mit Start in x

$$P_x(\tau_x > \tau_y) \leq \frac{\mathbf{E}_x[\tau_x]}{\tau_y}$$

ist und damit

$$\mathbf{E}_x[\tau_y] \leq \frac{\mathbf{E}_x[\tau_x]}{p} < \infty.$$

Andererseits lässt sich mit dem Satz von Fubini zeigen, dass

$$\mathbf{E}_y[\tau_x] \leq \frac{\mathbf{E}_x[\tau_x]}{p} < \infty,$$

woraus folgt, dass

$$\mathbf{E}_y[\tau_y] \leq \mathbf{E}_x[\tau_y] + \mathbf{E}_y[\tau_x] \leq \frac{2\mathbf{E}_x[\tau_x]}{p} < \infty.$$

Also ist y positiv rekurrent, womit sich die Klasseneigenschaft der positiven Rekurrenz ergibt. \square

Korollar 3.31 *Ist $(X_n)_{n \in \mathbf{N}_0}$ eine irreduzible Markovkette auf S , so sind entweder alle Zustände transient, null-rekurrent oder positiv rekurrent.*

Anmerkung 3.32 Für Markovketten mit endlichem Zustandsraum sind Rekurrenz und positive Rekurrenz gleichbedeutend.

Beispiel 3.33 (*Committor-Funktion*) Sei $(X_n)_{n \in \mathbf{N}_0}$ eine Markovkette auf S und $A, B \subset S$ zwei Teilmengen des Zustandsraums. Wir definieren die sogenannte *Committor-Funktion* als $q(x) = P_x(\tau_B < \tau_A)$, wobei τ_A, τ_B die Zeiten des ersten Eintritts der Kette in A oder B bezeichnen ($q(x)$ entspricht im wesentlichen der Funktion $P_x(\tau_x > \tau_y)$ aus dem Beweis von Lemma 3.30). Konkret betrachten wir die folgende Situation (vgl. [3]): Eine Maus,

³² Auf den genauen Zusammenhang zwischen Eigenwerten und Irreduzibilität wird später eingegangen werden.

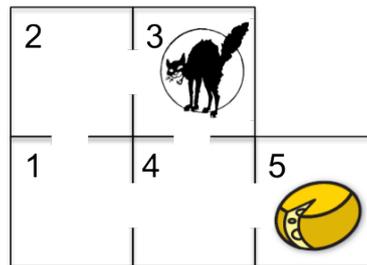


Abb. 8 Labyrinth mit Katze, Maus und einem Stück Käse. Die Maus will an den Käse, ohne der Katze zu begegnen, wobei sie in jedem Schritt zufällig mit Wahrscheinlichkeit $1/k$ einen der benachbarten k Räume auswählt.

die zusammen mit einer Katze in einem Labyrinth von Räumen gefangen ist, möchte an ein Stück Käse gelangen, das sich in einem der Räume befindet. Die Katze – von Natur aus faul – bewegt sich nicht, wohingegen die Maus zufällig die Räume durchstreift; dabei wählt sie in jedem Schritt zufällig gleichverteilt einen der benachbarten Räume (siehe Abbildung 8). Die Frage ist, mit welcher Wahrscheinlichkeit sie zum Käse gelangt, ohne zuvor der Katze zu begegnen, d.h., wie groß ist $q(x) = P_x(\tau_B < \tau_A)$, wenn x den Anfangszustand der Maus angibt, A das Ereignis $\{\omega: \text{Maus trifft auf Katze}\}$ und B das Ereignis $\{\omega: \text{Maus findet Käse}\}$?

Wir setzen $A = \{3\}$, $B = \{5\}$ und nehmen an, dass sich die Maus anfänglich in Raum (d.h., Zustand) $\{1\}$ befindet. Wir werden versuchen, $q(x)$ iterativ zu berechnen. Randbedingungen

$$q(3) = 0 \quad \text{und} \quad q(5) = 1$$

sind klar. Von Raum Nr. 1 kann die Maus entweder in Raum 2 oder 4 laufen. Da sie dies jeweils mit Wahrscheinlichkeit $1/2$ tut, ist

$$q(1) = \frac{1}{2}q(2) + \frac{1}{2}q(4)$$

mit

$$q(2) = \frac{1}{2}q(1) + \frac{1}{2}q(3) = \frac{1}{2}q(1)$$

und

$$q(4) = \frac{1}{3}q(1) + \frac{1}{3}q(3) + \frac{1}{3}q(5) = \frac{1}{3}q(1) + \frac{1}{3}$$

(Aus den Räumen $\{2\}$ bzw. $\{4\}$ geht es weiter in die Räume $\{1, 3\}$ bzw. $\{1, 3, 5\}$.) Wir erhalten schließlich

$$q(1) = \frac{2}{7}.$$

Anmerkung 3.34 Für eine Markovkette auf S mit Übergangsmatrix P ist die *Committer*-Funktion $P_x(\tau_B < \tau_A)$, $A, B \subset S$ allgemein als Lösung des linearen Gleichungssystems³³

$$\begin{aligned} (I - P)q &= 0 \\ q(A) &= 0 \\ q(B) &= 1 \end{aligned} \tag{3.6}$$

gegeben.

Stationäre Verteilung positiv rekurrenter Ketten

Wir kommen nun auf das Problem der stationären Verteilung zurück. Unsere Strategie wird dabei sein, die stationäre Wahrscheinlichkeit $\pi(x)$ eines Zustandes $x \in S$ mit der relativen Häufigkeit, mit der der Zustand besucht wird, zu identifizieren. Der folgende Satz gibt hinreichende und notwendige Bedingungen für die Existenz einer (elementweise) positiven stationären Verteilung an (siehe auch [10, Kap. 9.4]).

³³ ÜA.

Satz 3.35 Sei $(X_n)_{n \in \mathbb{N}_0}$ irreduzibel. Eine hinreichende und notwendige Bedingung für die Existenz einer eindeutigen stationären Verteilung $\pi > 0$ ist, dass die Kette positiv rekurrent ist. In diesem Fall ist

$$\pi(x) = (\mathbf{E}_x[\tau_x])^{-1}, \quad x \in S. \quad (3.7)$$

Beweis. Es bezeichne $P = (p_{xy})_{x,y \in S}$ die Übergangsmatrix der Kette und π eine stationäre Verteilung mit $\pi^T = \pi^T P$. Wir zeigen zunächst, dass π strikt positiv ist. Dazu zerlegen wir S in 2 disjunkte Mengen U und V , so dass U alle Zustände u mit $\pi(u) > 0$ enthält und für $v \in V$ entsprechend $\pi(v) = 0$ gilt. Da π eine Wahrscheinlichkeitsverteilung ist, muss für mindestens einen Zustand u gelten, dass $\pi(u) > 0$ ist, d.h., $U \neq \emptyset$. Umgekehrt ist für alle $v \in V$

$$\pi(v) = \sum_{x \in S} \pi(x) p_{xv} = 0$$

und damit $p_{uv} = 0$ für alle $u \in U$ und $v \in V$. Das steht aber im Widerspruch zur Irreduzibilität der Kette, denn die besagt, dass man mit positiver Wahrscheinlichkeit in einer bestimmten Anzahl von Schritten von jedem beliebigen Zustand in U zu jedem beliebigen Zustand in V wechseln kann; da aber $p_{uv} = 0$ für alle $u \in U$ und $v \in V$ gilt, kommt man nicht von U nach V oder von V nach U ; insbesondere existieren damit keine natürlichen Zahlen $n, m \geq 0$, so dass $p_{uv}^{(n)}, p_{vu}^{(m)} > 0$. Folglich ist $V = \emptyset$ bzw. $S = U$, woraus $\pi(x) > 0$ für alle $x \in S$ folgt. Aus der Irreduzibilität der Markovkette folgt zudem, dass π eindeutig ist. .

“ \Rightarrow ” Wir zeigen als erstes, dass positive Rekurrenz ein notwendiges Kriterium für die Existenz einer normierbaren stationären Verteilung ist. Nach dem oben gesagten ist die Kette rekurrent, denn wäre sie transient, so wäre $\pi \equiv 0$. Wegen der Stationarität von π gilt nun, dass

$$\begin{aligned} P_\pi(\tau_x < \infty) &= \lim_{n \rightarrow \infty} P_\pi(\tau_x \leq n) \\ &= \lim_{n \rightarrow \infty} \left(\sum_{l=1}^{n-1} P_\pi(X_{n-l} = x, X_{n-l+1} \neq x, \dots, X_n \neq x) + P_\pi(X_n = x) \right) \\ &= \lim_{n \rightarrow \infty} \left(P_\pi(X_0 = x) + \sum_{l=1}^{n-1} P_\pi(X_0 = x, X_1 \neq x, \dots, X_l \neq x) \right) \\ &= \sum_{l=0}^{\infty} P_\pi(X_0 = x, \tau_x > l) \\ &= P_\pi(X_0 = x) \sum_{l=0}^{\infty} P_x(\tau_x > l) \\ &= \pi(x) \sum_{l=0}^{\infty} \sum_{k=l+1}^{\infty} P_x(\tau_x = k) \\ &= \pi(x) \sum_{k=1}^{\infty} k P_x(\tau_x = k) \\ &= \pi(x) \mathbf{E}_x[\tau_x]. \end{aligned}$$

Im zweiten Schritt haben wir ausgenutzt, dass

$$\begin{aligned} \{\tau_x \leq n\} &= \{X_1 = x, X_2 \neq x, \dots, X_n \neq x\} \cup \dots \cup \{X_{n-1} = x, X_n \neq x\} \cup \{X_n = x\} \\ &= \bigcup_{l=1}^{n-1} \{X_{n-l} = x, X_{n-l+1} \neq x, \dots, X_n \neq x\} \cup \{X_n = x\} \end{aligned}$$

eine disjunkte Zerlegung des Ereignisses $\{\tau_x \leq n\}$ ist, und im dritten Schritt wurde benutzt, dass wegen der Homogenität der Übergangswahrscheinlichkeiten gilt, dass

$$P_\pi(X_{n-l} = x, X_{n-l+1} \neq x, \dots, X_n \neq x) = P_\pi(X_0 = x, X_1 \neq x, \dots, X_l \neq x), \quad 0 \leq l \leq n-1.$$

Aufgrund der Rekurrenz ist insbesondere $P_\pi(\tau_x < \infty) = 1$. Daraus folgt aber sofort

$$\mathbf{E}_x[\tau_x] = (\pi(x))^{-1} < \infty$$

für jedes $x \in S$ und damit die positive Rekurrenz der Kette.

“ \Leftarrow ” Wir definieren zunächst das Zählmaß

$$\mu_x(y) = \mathbf{E}_x \left[\sum_{n=1}^{\tau_x} \chi_{\{X_n=y\}} \right] \quad (3.8)$$

als die erwartete Anzahl der Besuche von y bei Start der Kette in x . Offenbar ist $\mu_x(x) = 1$ und es lässt sich zeigen, dass $\mu_x^T = \mu_x^T P$ eine stationäre Verteilung ist, die bis auf Skalierung $\mu_x \mapsto \alpha \mu_x$, $\alpha > 0$ eindeutig ist, wobei α von x abhängt. Die Stationarität von μ_x sieht man wie folgt: Zunächst macht man sich klar, dass nach dem Satz von der monotonen Konvergenz gilt:

$$\begin{aligned} \mu_x(y) &= \mathbf{E}_x \left[\sum_{n=1}^{\tau_x} \chi_{\{X_n=y\}} \right] \\ &= \mathbf{E}_x \left[\sum_{n=1}^{\infty} \chi_{\{X_n=y\}} \chi_{\{\tau_x \geq n\}} \right] \\ &= \sum_{n=1}^{\infty} \mathbf{E}_x [\chi_{\{X_n=y\}} \chi_{\{\tau_x \geq n\}}] \\ &= \sum_{n=1}^{\infty} P_x(X_n = y, \tau_x \geq n) \end{aligned}$$

Somit ist

$$\begin{aligned} \sum_{y \in S} \mu_x(y) p_{yz} &= \mu_x(x) p_{xz} + \sum_{y \neq x} \mu_x(y) p_{yz} \\ &= p_{xz} + \sum_{y \neq x} \sum_{n=1}^{\infty} P_x(X_n = y, \tau_x \geq n) p_{yz} \\ &= p_{xz} + \sum_{y \neq x} \sum_{n=1}^{\infty} P_x(X_{n+1} = z, X_n = y, \tau_x \geq n) \\ &= P_x(X_1 = z) + \sum_{n=1}^{\infty} P_x(X_{n+1} = z, \tau_x \geq n+1) \\ &= \sum_{n=1}^{\infty} P_x(X_n = z, \tau_x \geq n) \\ &= \mu(z), \end{aligned}$$

wobei wir in der vierten Gleichung ausgenutzt haben, dass

$$\{X_{n+1} = z, \tau_x \geq n+1\} = \bigcup_{y \neq x} \{X_{n+1} = z, X_n = y, \tau_x \geq n\}$$

eine disjunkte Vereinigung ist. Wir nehmen also an, die Kette sei positiv rekurrent. Dann ist

$$\begin{aligned} \sum_{y \in S} \mu_x(y) &= \sum_{y \in S} \mathbf{E}_x \left[\sum_{n=1}^{\tau_x} \chi_{\{X_n=y\}} \right] \\ &= \mathbf{E}_x \left[\sum_{n=1}^{\infty} \left(\sum_{y \in S} \chi_{\{X_n=y\}} \right) \chi_{\{\tau_x \geq n\}} \right] \\ &= \mathbf{E}_x \left[\sum_{n=0}^{\infty} \chi_{\{\tau_x > n\}} \right] \\ &= \sum_{n=0}^{\infty} P_x(\tau_x > n) \\ &= \mathbf{E}_x[\tau_x], \end{aligned}$$

und somit $\|\mu_x\|_1 < \infty$. Um aus μ_x ein Wahrscheinlichkeitsmaß zu machen, müssen wir es normieren:

$$\pi = (\mathbf{E}_x[\tau_x])^{-1} \mu_x.$$

Da $x \in S$ beliebig war und μ_x bis auf Skalierung eindeutig ist, erhalten wir

$$\pi(x) = (\mathbf{E}_x[\tau_x])^{-1} \mu_x(x) = (\mathbf{E}_x[\tau_x])^{-1}, \quad x \in S.$$

Damit ist die Behauptung bewiesen. □

Wir illustrieren die letzten Schritte anhand von 2 Beispielen zu Rekurrenz und Transienz.

Beispiel 3.36 (Reduzible Markovkette) Sei $(X_n)_{n \in \mathbf{N}_0}$ eine Markovkette auf $S = \{1, 2\}$ mit Übergangsmatrix

$$P = \begin{pmatrix} 1 - \alpha & \alpha \\ 0 & 1 \end{pmatrix}, \quad \alpha \in (0, 1).$$

Da der Zustand $x = 1$ transient ist, kann die stationäre Verteilung π nicht strikt positiv sein. In der Tat ist

$$\pi = (0, 1)^T,$$

wobei $\pi(1) = \lim_{n \rightarrow \infty} p_{11}^{(n)} = 0$ das Gewicht des transienten Zustands $x = 1$ und $\pi(2) = 1$ das positive Gewicht des rekurrenten Zustands $x = 2$ ist.

Beispiel 3.37 (Zufallsspaziergang auf \mathbf{Z}^d) Auf unendlich abzählbaren Zustandsräumen müssen Null-Rekurrenz und positive Rekurrenz unterschieden werden. Wir betrachten zunächst den Zufallsspaziergang auf $S = \mathbf{Z}$ (also $d = 1$) mit den Übergangswahrscheinlichkeiten

$$p_{i,i+1} = p \quad \text{und} \quad p_{i,i-1} = 1 - p = q, \quad i \in \mathbf{Z}$$

Die zugehörige Markovkette $(X_n)_{n \in \mathbf{N}_0}$ ist irreduzibel und wegen der Translationssymmetrie der Übergangswahrscheinlichkeiten können wir uns einen beliebigen Zustand, z.B. $x = 0$ herauspicken. Sicherlich kann die Kette bei Start in $X_0 = 0$ nur in einer geraden Anzahl von Schritten zu $x = 0$ zurückkehren. Folglich ist

$$p_{00}^{(2n+1)} = 0 \quad \forall n \in \mathbf{N}_0.$$

Für die Rückkehr zur 0 in einer geraden Anzahl von Schritten liefert die Stirling-Formel

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n \sim \frac{(4pq)^n}{\sqrt{\pi n}}, \quad n \gg 1.$$

Wir erinnern an die Definition von Rekurrenz und Transienz, Definition 3.25; es lässt sich beweisen, dass ein Zustand x genau dann rekurrent ist, wenn $P_x(\{X_n = x : \text{unendlich oft}\}) = 1$ ist bzw. transient, falls das Ereignis $\{X_n = x\}$ fast sicher nur endlich oft vorkommt. Nun ist für $p \neq q$

$$\sum_{n=1}^{\infty} p_{00}^{(2n)} < \sum_{n=1}^{\infty} (4p(1-p))^n < \infty,$$

und das impliziert nach dem Borel-Cantelli-Lemma, Lemma 2.39, dass die Kette nur endlich oft zu $x = 0$ zurückkehrt. Der asymmetrische Zufallsspaziergang ist damit transient. Für den symmetrischen Zufallsspaziergang $p = q = 1/2$ erhalten wir dagegen

$$\sum_{n=1}^{\infty} p_{00}^{(2n)} \sim \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty,$$

woraus nach dem Borel-Cantelli-Lemma die Rekurrenz der Kette folgt. Die Kette ist aber nicht positiv rekurrent, denn das stationäre Zählmaß $\mu = (1, 1, 1, \dots)^T$ ist auf \mathbf{Z} nicht normierbar.³⁴

In Dimension $d = 2$ ergibt sich im wesentlichen das gleiche Bild, wenn die Markovkette in einem Schritt jeweils ihre nächsten Nachbarn erreichen kann. Dagegen ist in höheren Dimensionen, $d \geq 3$, der Zufallsspaziergang immer transient, d.h., X_n verschwindet mit positiver Wahrscheinlichkeit nach unendlich.

³⁴ Auf abzählbar unendlichen Mengen gibt es keine Gleichverteilung.

3.4 Asymptotisches Verhalten und Kopplung von Markovketten

Wir wollen nun untersuchen, unter welchen Bedingungen die Verteilung einer Markovkette $(X_n)_{n \in \mathbf{N}_0}$ asymptotisch gegen ihre (eindeutige) stationäre Verteilung konvergiert. Mit anderen Worten: Wann gilt $\mu_k \rightarrow \pi$ für $k \rightarrow \infty$ und welche Startverteilungen μ_0 sind dabei erlaubt?

Periodizität und Aperiodizität

Damit überhaupt eine eindeutige stationäre Wahrscheinlichkeitsverteilung $\pi > 0$ existiert, muss unsere Markovkette irreduzibel und positiv rekurrent sein. Dass dies allein nicht reicht, illustriert das folgende Beispiel.

Beispiel 3.38 (Periodische Markovkette) Wir betrachten eine Markovkette auf $S = \{1, 2\}$ mit der Übergangsmatrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Da die Kette in jedem Schritt mit Wahrscheinlichkeit in den jeweils anderen Zustand wechselt, ist die Kette irreduzibel und rekurrent, folglich auch positiv rekurrent. In der Tat ist die eindeutige stationäre Verteilung – der Eigenwert $\lambda = 1$ ist einfach – durch $\pi = (1/2, 1/2)^T$ gegeben. Man sieht aber sofort, dass die Verteilung $\mu_k = P_{X_k}$ für kein $\mu_0 \neq \pi$ konvergiert, denn für alle $k \in \mathbf{N}_0$ ist

$$P^{2k+1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{und} \quad P^{2k} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

und das bedeutet, dass die Kette periodisch zwischen den beiden Zuständen $x = 1$ und $x = 2$ hin und her springt, ohne jemals die Gleichverteilung, sprich: ihre stationäre Verteilung zu erreichen.

Definition 3.39 (periodisch, aperiodisch) Als Periode $d(x)$ eines Zustands $x \in S$ definieren wir

$$d(x) = \text{ggT}\{n \geq 1 : p_{xx}^{(n)} > 0\},$$

wobei wir im Falle $p_{xx}^{(n)} = 0$ für alle $n \geq 1$ vereinbaren, dass $d(x) = \infty$. Ist $d(x) = 1$, so heißt der Zustand x *aperiodisch*, andernfalls wird x *periodisch* genannt.

Lemma 3.40 *Periodizität ist eine Klasseneigenschaft; insbesondere haben alle Zustände $x \in C$ einer Kommunikationsklasse $C \subset S$ dieselbe Periode $d(x)$.*

Beweis. Ist $(X_n)_{n \in \mathbf{N}_0} \subset S$ reduzibel, so gibt es natürliche Zahlen $k, m \in \mathbf{N}$, so dass $p_{xy}^{(k)}, p_{yx}^{(m)} > 0$, woraus folgt, dass $d(x) < \infty$ und $d(y) < \infty$ sind. Wir nehmen nun an, $x \in S$ periodisch. Mit Hilfe der Chapman-Kolmogorov-Gleichung (3.5) sehen wir sofort, dass dann

$$p_{xx}^{(k+j+m)} \geq p_{xy}^{(k)} p_{yy}^{(j)} p_{yx}^{(m)}$$

ist, außerdem dass $1 < d(x) < \infty$ Teiler von $k + m$ ist, denn nach Voraussetzung ist $p_{xx}^{(k+m)} \geq p_{xy}^{(k)} p_{yx}^{(m)} > 0$. Auch für diejenigen j für die $p_{yy}^{(j)} > 0$ gilt, ist $d(x)$ ein Teiler von $m + j + k$, folglich muss $d(x)$ auch Teiler von j , mithin von $d(y)$ sein. Da Kommunikation eine symmetrische Relation ist, können wir die Rollen von x und y vertauschen, woraus folgt, dass $d(y)$ Teiler von $d(x)$ ist. Also ist

$$d(x) = d(y),$$

womit die Behauptung bewiesen ist. □

Wir behaupten nun, dass Aperiodizität der fehlende Baustein ist – mit anderen Worten, dass für alle irreduziblen und aperiodischen Markovketten mit stationärer Verteilung $\pi > 0$ gilt:

$$\mu_k(A) = P(X_k \in A) \rightarrow \pi(A), \quad A \subset S,$$

für $k \rightarrow \infty$ und unabhängig von der Anfangsverteilung der Kette.

Kopplung von Zufallsvariablen

Unsere Behauptung werden wir mit Hilfe von Kopplungspaaren (*Copulae*) beweisen. Für unsere Zwecke ist die folgende Definition ausreichend.

Definition 3.41 (Kopplungspaar) Ein Kopplungspaar von 2 Zufallsvariablen $X, Y: \Omega \rightarrow S$ ist eine Zufallsvariable $Z = f(X, Y)$, $f: S \times S \rightarrow S \times S$ mit der Eigenschaft

$$P_X = P_Z \circ \pi_X^{-1} \quad \text{und} \quad P_Y = P_Z \circ \pi_Y^{-1}.$$

wobei $\pi_{X,Y}$ die Projektionen auf die X - bzw. Y -Komponenten von f bezeichnet.

Beispiel 3.42 (Periodische Markovkette, Fortsetzung) Wir betrachten zwei unabhängige Kopien $(X_n)_{n \in \mathbb{N}_0}$ und $(Y_n)_{n \in \mathbb{N}_0}$ der Markovkette aus Beispiel 3.38. Eine mögliches Kopplungspaar von $(X_n)_n$ und $(Y_n)_n$ ist die Produkt-Markovkette (Unabhängigkeitscopula)

$$(Z_n)_{n \in \mathbb{N}_0} = (X_n, Y_n)_{n \in \mathbb{N}_0} \subset S \times S$$

mit den Übergangswahrscheinlichkeiten

$$q_{ij} = P(Z_{n+1} = j | Z_n = i) = p_{i_x j_x} p_{i_y j_y}, \quad i = (i_x, i_y), j = (j_x, j_y) \in S \times S,$$

wobei p_{kl} die Einträge der Übergangsmatrix P bezeichnet; in unserem Fall ist

$$q_{ij} = (1 - \delta_{i_x j_x})(1 - \delta_{i_y j_y}).$$

Lemma 3.43 (Kopplungsungleichung I) Für unabhängige Zufallsvariable $X, Y: \Omega \rightarrow S$ mit Verteilungen $\mu = P_X$ und $\nu = P_Y$ gilt

$$\|\mu - \nu\|_{\text{TV}} \leq P(X \neq Y).$$

Beweis. Wir erinnern an die Definition des totalen Variationsabstands:³⁵

$$\|\nu - \mu\|_{\text{TV}} = \sup_{A \subset S} |\nu(A) - \mu(A)|.$$

Für alle $A \subset S$ ist nun

$$\begin{aligned} |\mu(A) - \nu(A)| &= |P(X \in A) - P(Y \in A)| \\ &= |P(X \in A, X \neq Y) + P(X \in A, X = Y) - P(Y \in A, X \neq Y) - P(Y \in A, X = Y)| \\ &= |P(X \in A, X \neq Y) - P(Y \in A, X \neq Y)| \\ &\leq P(X \neq Y). \end{aligned}$$

Da die Ungleichung für alle Teilmengen A gilt, haben wir die Behauptung bewiesen. \square

Anmerkung 3.44 Zu jedem Paar von Zufallsvariablen X, Y lassen sich stets unendlich viele Kopplungspaare Z finden – die Unabhängigkeitscopula ist nur eine Möglichkeit. Wegen

$$P(X \neq Y) = \sum_{x \neq y} P(Z = (x, y)) = 1 - \sum_x P(Z = (x, x))$$

zeichnet sich ein gutes Kopplungspaar gerade dadurch aus, dass $P(Z = (x, y))$ klein für $x \neq y$ bzw. groß für $x = y$ ist. Für unsere Zwecke bedeutet die Kopplungsungleichung, dass wir eine unabhängige Kopie $(Y_n)_{n \in \mathbb{N}_0}$ unserer Markovkette $(X_n)_{n \in \mathbb{N}_0}$ mit stationärer Anfangsverteilung $\nu_0 = \pi$ starten können und den Abstand der beiden Randverteilungen $\mu_k = P_{X_k}$ und $\pi = P_{Y_k}$ durch die Wahrscheinlichkeit des Ereignisses $\{X_k = Y_k\}$ abschätzen können. Dabei ist die folgende Definition hilfreich.

Definition 3.45 (Kopplungszeit) Seien $(X_n)_{n \in \mathbb{N}_0}, (Y_n)_{n \in \mathbb{N}_0}$ zwei unabhängige Markovketten auf S mit Übergangsmatrix P , stationärer Verteilung $\pi > 0$ und Anfangsverteilungen $\mu_0, \nu_0 = \pi$. Die Stoppzeit

$$\tau_c = \inf\{n \geq 1: X_n = Y_n\}$$

wird *Kopplungszeit* genannt.

³⁵ Manche Autoren definieren den TV-Abstand $\|\mu - \nu\|_{\text{TV}}$ mit dem Vorfaktor 2, weil die dadurch definierte Norm gerade der totalen Variation von $\mu - \nu$ entspricht; siehe z.B. [15, Kap. 1.5.3].

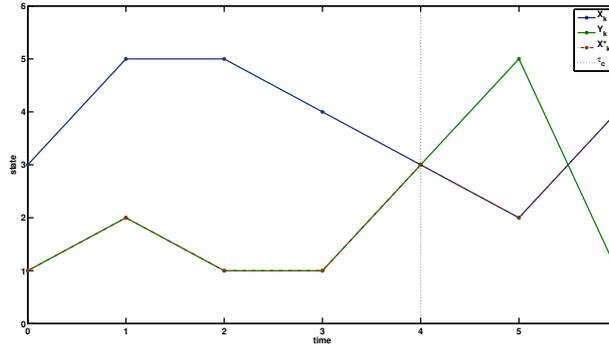


Abb. 9 Typische Realisierung des gestoppten Prozesses $(X_n^*)_{n \in \mathbf{N}_0}$.

Lemma 3.46 (Kopplungsungleichung II) *Es gilt*

$$\|\mu_k - \pi\|_{\text{TV}} \leq P(\tau_c > k).$$

Beweis. Wir definieren eine neue Markovkette (siehe Abb. 9)

$$X_k^* : \Omega \rightarrow S, \quad X_k^* = \begin{cases} Y_k, & k < \tau_c \\ X_k, & k \geq \tau_c \end{cases}.$$

Wegen der starken Markoveigenschaft ist $X^* = (X_n^*)_{n \in \mathbf{N}_0}$ wieder ein Markovprozess mit Übergangsmatrix P und der stationären Verteilung π . Da die zweite Markovkette zudem mit $Y_0 \sim \pi$ gestartet wird, ist

$$P_{X_k^*} = \pi \quad \forall n \in \mathbf{N}_0.$$

Nach Lemma 3.43 gilt dann für alle $A \subset S$, dass

$$|\mu_k(A) - \pi(A)| = |P(X_k \in A) - P(X_k^* \in A)| \leq P(X_k \neq X_k^*) \leq P(\tau_c > k),$$

wobei die letzte Ungleichung aus der Monotonie von P und $\{X_k \neq X_k^*\} \subset \{\tau_c > k\}$ folgt. □

Satz 3.47 (Ergodensatz für Markovketten) *Für irreduzible, aperiodische Markovketten mit Übergangsmatrix P und stationärer Verteilung $\pi > 0$ gilt*

$$\|\mu_k - \pi\|_{\text{TV}} \rightarrow 0$$

für $k \rightarrow \infty$ und unabhängig von der Startverteilung μ_0 .

Beweis. Nach Lemma 3.46 reicht es, zu zeigen, dass $P(\tau_c > k)$ für $k \rightarrow \infty$ gegen 0 geht, was gleichbedeutend mit der Aussage $P(\tau_c < \infty) = 1$ ist. Dazu setzen wir $Z_k = (X_k, Y_k)$ und betrachten die Unabhängigkeitscopula $(Z_n)_{n \in \mathbf{N}_0}$, die nach Konstruktion $\pi_Z(x, y) = \pi(x)\pi(y) > 0$ als stationäre Verteilung hat, folglich positiv rekurrent ist. Wenn wir zeigen können, dass die Kette zudem reduzibel ist, wissen wir, dass sie mit Wahrscheinlichkeit 1 irgendwann die Diagonale $\{(x, x) : x \in S\}$ trifft, und wir sind fertig. Da die Komponenten der Produktkette unabhängig sind, sind die Mehrschritt-Übergangswahrscheinlichkeiten gerade durch

$$q_{ij}^{(k)} = P(Z_{n+k} = j | Z_n = i) = p_{i_x j_x}^{(k)} p_{i_y j_y}^{(k)}$$

gegeben, wobei $i = (i_x, i_y)$, $j = (j_x, j_y)$ die Zustände der Produktkette bezeichnet. Wegen der Aperiodizität der Übergangsmatrix P gibt es ein $k^* \in \mathbf{N}$, so dass

$$p_{xx}^{(k)} > 0 \quad \forall k \geq k^* \quad \forall x \in S.$$

Aus der Irreduzibilität von P und der Chapman-Kolmogorov-Gleichung (3.5) folgt damit, dass wir ein $n^* \in \mathbf{N}$ finden können, so dass für alle $i, j \in S \times S$, $n \geq n^*$ und $k \geq k^*$ gilt:

$$q_{ij}^{(n+k)} = p_{i_x j_x}^{(n+k)} p_{i_y j_y}^{(n+k)} \geq p_{i_x j_x}^{(n)} p_{j_x j_x}^k p_{i_y j_y}^{(n)} p_{j_y j_y}^{(k)} > 0.$$

Die Produktkette ist somit irreduzibel und die Kopplungszeit τ_c fast sicher endlich, was zu beweisen war □

Anmerkung 3.48 Ein unmittelbare Folgerung aus dem Ergodensatz ist, dass die Zeilen der iterierten Übergangsmatrix P^k gegen π konvergieren. (Um dies zu sehen, setze man z.B. $\mu_0 = \delta_x$ und rechne $\mu_0 P^k$ aus.)

Beispiel 3.49 (Fortsetzung von Beispiel 3.42) Wir illustrieren den Ergodensatz mit einem (Gegen-)Beispiel und betrachten die Produktkette $(Z_n)_{n \in \mathbf{N}_0} = (X_n, Y_n)_{n \in \mathbf{N}_0}$ mit den Übergangswahrscheinlichkeiten

$$q_{ij} = P(Z_{n+1} = j | Z_n = i) = (1 - \delta_{i_x j_x})(1 - \delta_{i_y j_y}).$$

Wir starten die Produktkette mit Anfangswerten $X_0 = 1$ (f.s.) und $Y_0 \sim \pi$, d.h., $P(Y_0 = 1) = P(Y_0 = 2) = 1/2$. Mit Wahrscheinlichkeit $P(Y_0 = 2) = 1/2$ startet die Produktkette vom Zustand $(1, 2)$, von dem aus mit Wahrscheinlichkeit 1 der Zustand $(2, 1)$ angesteuert wird, bevor die Kette wegen der Periodizität wieder in den Anfangszustand zurückkehrt. Das bedeutet aber, dass die Produktkette nicht irreduzibel ist und, da die Diagonale $z = (1, 1)$ oder $z = (2, 2)$ nicht für alle Startwerte Z_0 erreicht wird, dass $P(\tau_c < \infty) < 1$.

Beispiel 3.50 (Konsensbildung) Eine interessante Anwendung von Markovketten ist die Modellierung von Verhandlungsstrategien zwischen Individuen. Das folgende Beispiel stammt aus [10, S. 262] und beschreibt das Problem, in einer Gruppe von N Personen einen Konsens herbeizuführen. Konkret geht es um die Aufgabe, gemeinschaftlich eine Schätzung für eine Größe G abzugeben. Dazu $S = \{1, \dots, N\}$ die Menge der Personen und $\mu_k(i) \geq 0$ das relative Gewicht, mit dem die Schätzung G_i der i -ten Person nach k Runden der Konsensfindung berücksichtigt wird. Die gemeinschaftliche Schätzung unter der Annahme $\mu_k \rightarrow \pi$ ist entsprechend

$$G_\pi = \sum_{i=1}^N G_i \pi_i.$$

Für eine basisdemokratisch organisierte Personengruppe, bei der die Schätzung einfach als arithmetisches Mittel aus den individuellen Schätzungen G_i gebildet wird, hat die Übergangsmatrix die schlichte Form

$$P = \begin{pmatrix} 1/N & \cdots & 1/N \\ \vdots & \ddots & \vdots \\ 1/N & \cdots & 1/N \end{pmatrix}.$$

Es lässt sich leicht nachprüfen, dass P irreduzibel, aperiodisch und positiv rekurrent ist (vgl. die untenstehende Bemerkung). Da die Spalten von P alle linear abhängig (bzw. sogar identisch) sind hat P genau einen Eigenwert 1 und $N - 1$ Eigenwerte 0. Die Übergangsmatrix $P = P^T$ ist also eine Projektionsmatrix, die jede beliebige Anfangsverteilung in einem Schritt auf die eindeutige stationäre Verteilung

$$\pi = (1/N, \dots, 1/N)^T$$

abbildet. In diesem Fall wird ein Konsens in einem Schritt erzielt. Alternativ könnte man ein Modell entwickeln, bei dem die i -te Personen allen anderen Personen (einschließlich sich selbst) ein Gewicht $p_{ij} > 0$, $j = 1, \dots, N$ zuordnet, dass dem Mitspracherecht bei der gemeinschaftlichen Schätzung entspricht. Sind alle Gewichte positiv und gemäß $\sum_j p_{ij} = 1$ normiert, ist die zugehörige stochastische Matrix $P = (p_{ij})_{i,j \in S}$ aperiodisch mit eindeutiger stationärer Verteilung $\pi > 0$ (die Eindeutigkeit folgt aus der Irreduzibilität von P). Da vermutlich nicht allen Personen von vornherein die gleichen Mitspracherechte zugebilligt werden, werden die Gewichte zeilenweise unterschiedlich sein, so dass ein Konsens ausgehandelt werden muss – ansonsten wären wie im ersten Fall die Zeilen der Übergangsmatrix alle gleich und ein Konsens nach einem Schritt erreicht. Das Aushandeln der Mitspracherechte (d.h., wer billigt wem wie viel Kompetenz beim Schätzen zu?) geschieht durch Iteration $\mu_{k+1}^T = \mu_k^T P$, wobei nach Konstruktion der Markovkette μ_k asymptotisch gegen π konvergiert.

3.5 Satz von Perron-Frobenius

Die Aperiodizität der Übergangsmatrix P in Beispiel 3.50 ergibt sich aus einem Satz über Gerschgorin-Kreise, der im wesentlichen besagt, dass P nur einen Eigenwert λ mit der Eigenschaft $|\lambda| = 1$ haben kann, wenn die Diagonalelemente p_{ii} alle positiv sind [8]; da die Zeilensummennorm einer stochastischen Matrix gleich eins und letztere eine obere Schranke der Spektralnorm ist, müssen alle übrigen Eigenwerte von P vom Betrage kleiner als 1 sein (vgl. Definition 3.51 weiter unten). Aperiodizität impliziert folglich, dass die Projektionen auf die

Eigenräume von P^k zu Eigenwerten $|\lambda| < 1$ asymptotisch verschwinden, so dass für $k \rightarrow \infty$ nur der Eigenraum zum einzigen Eigenwert 1, der von der stationären Verteilung π aufgespannt wird, übrigbleibt.

Wir wollen den Zusammenhang zwischen den Eigenwerten einer stochastischen Matrix und ihrem asymptotischen Verhalten nun etwas genauer untersuchen, wobei wir uns der Einfachheit halber auf Markovketten auf endlichem Zustandsraum beschränken werden. Als Beispiel betrachten wir die Familie von Übergangsmatrizen

$$P^\epsilon = \begin{pmatrix} \epsilon & 1 - \epsilon \\ 1 - \epsilon & \epsilon \end{pmatrix}, \quad \epsilon \in (0, 1),$$

von irreduziblen Markovketten auf $S = \{1, 2\}$. Dazu betrachten wir das Eigenwertproblem

$$P_\epsilon^T v^\epsilon = \lambda^\epsilon v^\epsilon.$$

mit der Lösung

$$(\lambda_1^\epsilon, v_1^\epsilon) = (1, \pi), \quad (\lambda_2^\epsilon, v_2^\epsilon) = (2\epsilon - 1, \eta)$$

wobei die beiden Eigenvektoren $\pi = (1/2, 1/2)^T$ und $\eta = (-1, 1)$ unabhängig von ϵ sind. Da π und η linear unabhängig sind, lässt sich jede Startverteilung μ_0 wie folgt in der Eigenbasis darstellen:

$$\mu_0 = \pi + c\eta,$$

wobei der Koeffizient c eindeutig durch μ_0 und die Bedingung $\mu_0(1) + \mu_0(2) = 1$ festgelegt ist. Nach Lemma 3.13 ist die Verteilung der Kette nach n Schritten

$$\mu_n^T = (\pi + c\eta)^T P_\epsilon^n = \pi^T + c(\lambda_2^\epsilon)^n \eta^T$$

gegeben. Man beachte, dass $\eta(1) + \eta(2) = 0$ ist, so dass μ_n für alle $n \geq 0$ eine Wahrscheinlichkeitsverteilung ist. Um die Verteilung der Kette für $n \rightarrow \infty$ zu untersuchen, betrachten wir zunächst die zwei Fälle:

1. Aperiodische Markovkette ($0 < \epsilon < 1$): In diesem Fall ist $|\lambda_2^\epsilon| < 1$ und somit gilt

$$\lim_{n \rightarrow \infty} \mu_n^T = \pi^T + c \lim_{n \rightarrow \infty} (\lambda_2^\epsilon)^n \eta^T = \pi^T,$$

unabhängig von der Startverteilung.

2. Periodische Markovkette ($\epsilon = 0$): Hier ist $\lambda_2^\epsilon = -1$ und die Kette springt periodisch zwischen den Verteilungen

$$\mu_n = \begin{cases} \pi + c\eta, & n \text{ gerade} \\ \pi - c\eta, & n \text{ ungerade} \end{cases}$$

hin und her.

Dieses einfache Beispiel lässt vermuten, dass Aperiodizität einer Übergangsmatrix P (bzw. einer Markovkette) gerade bedeutet, dass der Eigenwert $\lambda_1 = 1$ einfach ist und dass für alle übrigen Eigenwerte gilt, dass $|\lambda_i| < 1$, so dass die entsprechenden Terme in der Spektralzerlegung von P asymptotisch verschwinden, d.h.

$$P^n \rightarrow (\pi^T | \dots | \pi^T), \quad n \rightarrow \infty.$$

Definition 3.51 (Spektralnorm) Ist $A \in \mathbf{R}^{n \times n}$ eine reelle Matrix mit Eigenwerten $\lambda_1, \dots, \lambda_n \in \mathbf{C}$, so heißt

$$\rho(A) = \max_{i=1, \dots, n} |\lambda_i|$$

Spektralradius von A .

Lemma 3.52 Stochastische Matrizen $P \in \mathbf{R}^{n \times n}$ haben Spektralradius $\rho(P) = 1$.

Beweis. Es sei v_i der Eigenvektor von P zum Eigenwert λ_i . Für jede induzierte Matrixnorm $\|\cdot\|$ gilt, dass die Spektralnorm eine untere Schranke darstellt, denn

$$\|P\| = \sup_{u \neq 0} \frac{|Pu|}{|u|} \geq \frac{|Pv_i|}{|v_i|} = |\lambda_i|.$$

Das heißt, $\|P\| \geq \rho(P)$. Umgekehrt gilt für die von der Maximumsnorm induzierte Zeilensummennorm:

$$\|P\|_\infty = \max_{i=1, \dots, n} \sum_{j=1}^n p_{ij} = 1 \geq \rho(P)$$

Da wegen der Zeilensummeneigenschaft der Eigenwert $\lambda = 1$ immer existiert, muss $\rho(P) = 1$ gelten. \square

Lemma 3.52 besagt, das für eine stochastische Matrix $P \in \mathbf{R}^{n \times n}$ der Eigenwert $\lambda_1 = 1$ stets der betragsmäßig größte Eigenwert ist, aber nicht notwendigerweise der einzige. Der berühmte Satz von Perron und Frobenius gibt genauer darüber Auskunft, unter welchen Bedingungen $\lambda_1 = 1$ der einzige Eigenwert mit Betrag 1 ist.³⁶

Satz 3.53 (Satz von Perron-Frobenius) *Es sei $P \in \mathbf{R}^{n \times n}$ eine irreduzible und aperiodische stochastische Matrix. Dann gelten folgende Aussagen:*

1. $\lambda_1 = 1$ ist einfacher Eigenwert mit strikt positiven Links- und Rechtseigenvektoren

$$\pi^T = \pi^T P, \quad \mathbf{1} = P \mathbf{1}.$$

2. Für alle Eigenwerte $\lambda_i \neq 1$ gilt $|\lambda_i| < 1$.
3. Sind die Eigenwerte $\lambda_1, \dots, \lambda_r$, $r \leq n$ so geordnet, dass

$$1 = \lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_r|$$

ist, so gilt

$$P^n = (\pi^T | \dots | \pi^T) + \mathcal{O}(n^{m-1} |\lambda_2|^n), \quad (3.9)$$

wobei m die algebraische Vielfachheit von λ_2 angibt.

Beweis. Den ersten Teil der Aussage, der allein aus der Irreduzibilität von P folgt, haben wir bereits auf S. 59 bewiesen. Wir beweisen den zweiten Teil, ohne jedoch den Ergodensatz für Markovketten, Satz 3.47, zu benutzen. Dazu nehmen wir an, dass $\mu \neq 1$ ein Eigenwert von P mit der Eigenschaft $|\mu| = 1$ sei und führen die Annahme zum Widerspruch. (Wegen $\rho(P) = 1$ kann es keine betragsmäßig größeren Eigenwerte geben.)

- Wegen der Irreduzibilität und Aperiodizität der Kette existiert ein $k^* \in \mathbf{N}$, so dass die Matrix P^{k^*} für alle $k \geq k^*$ strikt positive Einträge hat: Aperiodizität bedeutet, dass es ein $n^* \geq 1$ gibt, so dass $p_{xx}^{(n)} > 0$ für alle $x \in S$ und für alle $n \geq n^*$. Zusammen mit der Irreduzibilität folgt aus Chapman-Kolmogorov-Gleichung dann, dass es zu jedem $y \in S$ ein $l = l(x, y)$ gibt, so dass

$$p_{xy}^{n+l} \geq p_{xx}^{(n)} p_{xy}^{(l)} > 0, \quad n \geq n^*.$$

Folglich gilt $P^k > 0$ für alle für $k \geq k^* \geq n^* + l^*$ mit $l^* = \max_{x, y \in S} l(x, y)$.

- Sei nun $k^* \geq n^* + l^*$ so gewählt, dass μ^{k^*} strikt negativ ist; ein solches k^* existiert, da $|\mu| = 1$ und $\mu \neq 1$. Damit gibt es ein $\epsilon > 0$, so dass die Matrix

$$T = P^{k^*} - \epsilon I$$

positiv ist. Nach Konstruktion hat T den Eigenwert $\mu^{k^*} - \epsilon$, wobei $|\mu^{k^*} - \epsilon| > 1$.

- Nun gilt eintragsweise $0 \leq T \leq P^{k^*}$, woraus mit Hilfe der Gelfandschen Formel,

$$\rho(A) = \lim_{k \rightarrow \infty} \sqrt[k]{\|A^k\|},$$

folgt, dass

$$\rho(T) \leq \rho(P^{k^*}) \leq (\rho(P))^{k^*} = 1,$$

im Widerspruch zur Annahme $|\mu| = 1$.

Der Beweis der Konvergenzaussage findet sich im Buch von Meyn und Tweedie [13]. \square

³⁶ Oskar Perron (1880–1975), dt. Mathematiker; Ferdinand Georg Frobenius (1849–1917), dt. Mathematiker

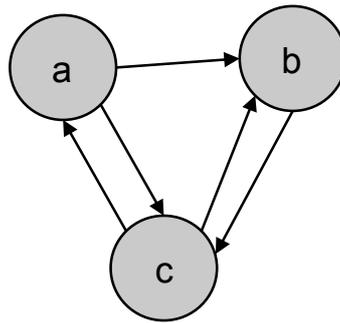


Abb. 10 Das Internet, bestehend aus drei miteinander verlinkten Webseiten.

Eine Anwendung: Googles PageRank-Algorithmus

Eine Anwendung, bei der der zweite Eigenwert einer stochastischen Matrix P eine zentrale Rolle spielt, ist der Google PageRank-Algorithmus [4], mit dem Google die Relevanz von Webseiten bewertet. Die Idee des Verfahrens ist es, jeder Webseite i einen PageRank $\pi(i)$ zuzuordnen, der sich aus dem PageRank der Seiten ergibt, die auf sie verweisen. Der PageRank entspricht demnach der stationären Verteilung eines Zufallsspaziergangs auf dem Zustandsraum aller Webseiten, wobei die Übergangswahrscheinlichkeiten von Webseite i zu Webseite j als

$$p_{ij} = (\#\text{Links, die von } i \text{ ausgehen})^{-1}$$

definiert sind. Als Beispiel betrachten wir ein System aus drei Webseiten, das in Abbildung 10 dargestellt ist. Die zugehörige Übergangsmatrix $P^{3 \times 3}$ ist durch

$$P = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \\ 1/2 & 1/2 & 0 \end{pmatrix}$$

gegeben, und der PageRank π ergibt sich in diesem Fall aus der Lösung der Eigenwertgleichung

$$\pi = P^T \pi \quad \Rightarrow \quad \pi = (2/9, 1/3, 4/9)^T. \quad (3.10)$$

Die Definition von P entspricht der Tatsache, dass von $\{a\}$ zwei Links abgehen, von $\{b\}$ einer und von $\{c\}$ wieder zwei Links; Links, die auf sich selbst verweisen gibt es in diesem Modell nicht, d.h., ein Zufallssurfer bewegt sich in jedem Schritt auf eine andere Website, indem er zufällig einem Link auf der jeweiligen folgt.

Für das Internet mit seinen ca. 700 Millionen registrierten Webseiten ist die Eigenwertgleichung (3.10) ein Gleichungssystem mit 700 Millionen Unbekannten, das sich numerisch nicht direkt lösen lässt.³⁷ Algorithmisch lässt sich jedoch die Tatsache ausnutzen, dass $\mu_0^T P^n \rightarrow \pi^T$, sofern die Übergangsmatrix irreduzibel und aperiodisch ist (Potenzmethode). Im Allgemeinen muss eine wie oben definierte Übergangsmatrix jedoch weder irreduzibel noch aperiodisch sein, daher verwendet der Algorithmus eine regularisierte Übergangsmatrix der Form

$$P_\epsilon = (1 - \epsilon)P + \epsilon I,$$

die nach Konstruktion für $\epsilon \in (0, 1)$ aperiodisch ist gilt, dass

$$(P_\epsilon^T)^n \mu_0 \rightarrow \pi_\epsilon, \quad n \rightarrow \infty.$$

Ist die Matrix P_ϵ irreduzibel, so ist der Grenzwert unabhängig von μ_0 , wobei die Geschwindigkeit des Verfahrens stets davon abhängt, wie nahe der zweite Eigenwert λ_2 von P_ϵ am Einheitskreis liegt. (In der Regel braucht eine Neuberechnung des PageRanks allein aufgrund der Größe des Zustandsraums mehrere Wochen.)

³⁷ Der Aufwand bei anwächstender voller Besetzung steigt asymptotisch mit der Anzahl der Webseiten hoch 3.

3.6 Markovketten-Monte-Carlo

Wir haben im letzten Abschnitt bewiesen, dass die Punkte jeder unendlich langen Realisierung einer ergodischen Markovkette nach der stationären Verteilung π verteilt sind. Da liegt die Frage nahe, ob sich auf die gleiche Weise Erwartungswerte bezüglich π berechnen lassen.

Gesetz der großen Zahlen für Markovketten

Es sei $(X_n)_{n \in \mathbf{N}_0}$ eine Markovkette mit stationärer Verteilung. Wir betrachten Mittelwerte

$$S_n(f) = \frac{1}{n+1} \sum_{i=0}^n f(X_i)$$

von integrierbaren Funktionen $f: S \rightarrow \mathbf{R}$, d.h., Funktionen mit der Eigenschaft

$$\sum_{x \in S} |f(x)| \pi(x) < \infty.$$

Wären die Zufallsvariablen X_i alle unabhängig und gemäß π verteilt, so würde nach dem Gesetz der großen Zahlen $S_n(f)$ fast sicher gegen den Erwartungswert $\mathbf{E}[f] = \sum_x f(x) \pi(x)$ konvergieren. Im Falle einer ergodischen Markovkette $(X_k)_{k \in \mathbf{N}_0}$ mit stationärer Verteilung $\pi > 0$, sind die Folgenglieder X_i zwar asymptotisch π -verteilt, aber als Realisierung einer Markovkette nicht unabhängig. Satz 2.41 ist damit nicht direkt anwendbar, aber die starke Markoveigenschaft liefert auch hier den fehlenden Baustein (vgl. die untenstehende Bemerkung).

Satz 3.54 (Starkes Gesetz der großen Zahlen für Markovketten) *Sei $(X_n)_{n \in \mathbf{N}_0}$ eine irreduzible Markovkette auf S mit stationärer Verteilung $\pi > 0$. Dann gilt für alle Anfangswerte $X_0 = x \in S$, dass*

$$P_x \left(\lim_{n \rightarrow \infty} S_n(f) = \mathbf{E}[f] \right) = 1.$$

Beweis. Wir definieren zunächst das Zählmaß

$$\mu_x(y) = \mathbf{E}_x \left[\sum_{n=0}^{\tau_x} \chi_{X_n=y} \right] < \infty.$$

Nach Satz 3.35 gilt $\mu_x(y) = C_x \pi(y)$ mit $C_x = \mathbf{E}_x[\tau_x] < \infty$.

- Nun sei $U_0 = \sum_{k=0}^{\tau_x} f(X_k)$ eine neue Zufallsvariable mit dem Erwartungswert

$$\begin{aligned} \mathbf{E}[U_0] &= \mathbf{E}_x \left[\sum_{k=0}^{\tau_x} f(X_k) \right] \\ &= \mathbf{E}_x \left[\sum_{k=0}^{\tau_x} \left(\sum_{y \in S} \chi_{X_k=y} \right) f(X_k) \right] \\ &= \sum_{y \in S} f(y) \left(\mathbf{E}_x \left[\sum_{k=0}^{\tau_x} \chi_{X_k=y} \right] \right) \\ &= \sum_{y \in S} f(y) \mu_x(y). \end{aligned}$$

Dazu definieren wir Zufallsvariable $U_r = \sum_{k=T_r+1}^{T_{r+1}} f(X_k)$, wobei $T_r = \tau_x^{(r)}$ die Zeit der r -ten Wiederkehr zu $x \in S$ bezeichnet. Wegen der starken Markoveigenschaft sind U_0, U_1, \dots unabhängig und identisch verteilt mit endlichem Erwartungswert $\mathbf{E}[U_0] < \infty$, und es folgt aus dem starken Gesetz der großen Zahlen, dass

$$\frac{1}{n+1} \sum_{i=0}^n U_i \xrightarrow{\text{f.s.}} \mathbf{E}[U_0].$$

Das heißt aber nichts anderes als

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{i=0}^{T_{n+1}} f(X_i) = \sum_{y \in S} f(y) \mu_x(y) \quad (\text{f.s.}).$$

- Wir nehmen an, dass $f \geq 0$ sei und definieren $N_x(n) = \sum_{i=0}^n \chi_{X_i=x}$ als die Zahl der Besuche von x in den ersten n Schritten der Markovkette. Nach Definition gilt

$$T_{N_x(n)} \leq n < T_{N_x(n)+1}$$

und somit

$$\frac{1}{N_x(n)} \sum_{i=0}^{T_{N_x(n)}} f(X_i) \leq \frac{1}{N_x(n)} \sum_{i=0}^n f(X_i) \leq \frac{1}{N_x(n)} \sum_{i=0}^{T_{N_x(n)+1}} f(X_i)$$

Wegen der Rekurrenz von $(X_n)_{n \in \mathbf{N}_0}$ geht $N_x(n) \rightarrow \infty$; da zudem τ_x fast sicher endlich ist, konvergieren die beiden äußeren Summen der letzten Ungleichung für $n \rightarrow \infty$ fast sicher gegen $\sum_y f(y)\mu_x(y)$, d.h.,

$$\frac{1}{N_x(n)} \sum_{i=0}^n f(X_i) = \sum_{y \in S} f(y)\mu_x(y) \quad (\text{f.s.}).$$

- Aus der letzten Gleichung können wir die fehlende Normierungskonstante bestimmen; für $g \equiv 1$ erhalten wir

$$\frac{1}{N_x(n)} \sum_{i=0}^n g(X_i) = \frac{n+1}{N_x(n)} \xrightarrow{\text{f.s.}} \sum_{y \in S} \mu_x(y) = C_x,$$

woraus folgt, dass (f.s.)

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{i=0}^n f(X_i) &= \lim_{n \rightarrow \infty} \frac{1}{N_x(n)} \frac{N_x(n)}{n+1} \sum_{i=0}^n f(X_i) \\ &= C_x^{-1} \sum_{y \in S} f(y)\mu_x(y) \\ &= \sum_{y \in S} f(y)\pi(y). \end{aligned}$$

- Zu guter Letzt lassen wir die Beschränkung $f \geq 0$ fallen. Mit der Zerlegung von f in positiven und negativen Anteil, d.h., $f = f^+ - f^-$ mit $f^\pm = \max\{0, \pm f\}$ folgt die Aussage für alle f mit endlichem Erwartungswert. □

Die Aussage des letzten Satzes lässt sich durch die Formel

$$\text{Zeitmittel} = \text{Scharmittel}$$

zusammenfassen. Grob gesprochen bedeutet das, dass das Langzeitverhalten einer Markovkette (Zeitmittel) dem Mittel über alle Zustände mit der stationären Verteilung entspricht (Scharmittel). Anders als bei Satz 3.47 haben wir dabei keine Aperiodizität vorausgesetzt, und das folgende Beispiel zeigt, dass das starke Gesetz der großen Zahlen für Markovketten tatsächlich eine abgeschwächte Form des Ergodensatzes impliziert.

Beispiel 3.55 Angenommen, die Übergangsmatrix P wäre entweder nicht bekannt oder zu groß, um das Eigenwertproblem $\pi^T = \pi^T P$ numerisch oder gar analytisch lösen zu können. Was tun? Wir setzen $f(x) = \chi_A(x)$ für irgendeine Teilmenge $A \subset S$ und zählen, wie oft $(X_n)_{n \in \mathbf{N}_0}$ die Menge A besucht, d.h.,

$$\frac{1}{n+1} \sum_{i=0}^n \chi_A(X_i) \xrightarrow{\text{f.s.}} \pi(A), \quad A \subset S.$$

Da wir das insbesondere für jeden Zustand $A = \{x: x \in S\}$ tun können, liefern die Anzahl der Besuche von x geteilt durch die Länge der Realisierung eine Approximation von $\pi(x)$.

Anmerkung 3.56 Da die Aussage von Satz 3.54 für alle π -integrierbaren Funktionen f gilt, entspricht die Konvergenz

$$\frac{1}{n+1} \sum_{i=0}^n f(X_i) \xrightarrow{\text{f.s.}} \sum_{x \in S} f(x)\pi(x)$$

im wesentlichen der schwachen Konvergenz der empirischen Verteilung

$$\mu_n = \frac{1}{n+1} \sum_{i=0}^n \delta_{X_i}$$

gegen die stationäre Verteilung π , sprich: es gilt fast sicher (vgl. Definition 2.19)

$$\lim_{n \rightarrow \infty} \int f d\mu_n = \int f d\pi, \quad \forall f \in \ell_\pi^1(S).$$

Dabei interpretieren wir π in dem Integral auf der rechten Seite als Zählmaß auf S und bezeichnen mit $L^1(S, \pi)$ den Vektorraum aller bezüglich π integrierbaren Funktionen, auf dem der Erwartungswert ein lineares Funktional der Form $\ell_\pi(f) = \mathbf{E}_\pi[f]$ darstellt. Wegen $\pi > 0$ gilt dann für alle $n \in \mathbf{N}_0$, dass $L^1(S, \pi) \subset L^1(S, \mu_n)$, wodurch sichergestellt ist, dass auch $\mathbf{E}_{\mu_n}[f]$ für alle $n \in \mathbf{N}_0$ existiert.

Der wesentliche Punkt beim Beweis von Satz 3.54 besteht darin, durch Summieren der Folgenglieder zwischen den Erneuerungszeiten (Wiederkehrzeiten) der Markovkette neue Zufallsvariable zu bilden, die wegen der starken Markoveigenschaft der Kette unabhängig sind, so dass die üblichen Grenzwertsätze angewendet werden können. Die folgende Aussage lässt sich so beweisen.

Korollar 3.57 Sei $g: S \times S \rightarrow \mathbf{R}$ messbar mit der Eigenschaft

$$\sum_{x,y \in S} |g(x,y)| \pi(x) p_{xy} < \infty.$$

Dann gilt für alle $X_0 = x \in S$, dass

$$P_x \left(\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n g(X_k, X_{k+1}) = \sum_{x,y \in S} g(x,y) \pi(x) p_{xy} \right) = 1.$$

Beweis. ÜA. (Hinweis: Man zeige, dass $\pi(x)p_{xy}$ die stationäre Verteilung von $Y_n = (X_n, X_{n+1})$ ist.) \square

Beispiel 3.58 Neben der stationären Verteilung lassen sich auch die Übergangswahrscheinlichkeiten aus einer Realisierung der Markovkette schätzen – Korollar 3.57 zeigt, wie es geht. Wir setzen

$$g(x,y) = \chi_{(a,b)}(x,y) = \begin{cases} 1, & x = a, y = b \\ 0, & \text{sonst} \end{cases}.$$

Das Schätzen des Übergangsmatrixelements aus Simulationsdaten entspricht dann dem Zählen der Übergänge zwischen den Zuständen a und b in einem Schritt:

$$\frac{1}{n+1} \sum_{i=0}^n \chi_{(a,b)}(X_i, X_{i+1}) \xrightarrow{\text{f.s.}} \pi(a) p_{ab}, \quad a, b \in S.$$

Ist $\pi > 0$ bekannt (z.B. durch eine Schätzung auf Basis einer langen Realisierung der Kette), so ist

$$p_{ab} = (\pi(a))^{-1} \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{i=0}^n \chi_{(a,b)}(X_i, X_{i+1}).$$

Reversible Markovketten

Eine wichtige Klasse von Markovketten sind reversible Ketten. Eine Markovkette $(X_n)_{n \in \mathbf{N}_0}$ wird reversibel genannt, wenn der rückwärts laufende Prozess $Y_k = X_{T-k}$ für beliebige $T \in \mathbf{N}$ dieselbe Verteilung wie der vorwärts laufende Prozess, d.h., wenn für alle $x_0, \dots, x_T \in S$ und $T \in \mathbf{N}$ gilt, dass

$$P_\mu(Y_0 = x_T, \dots, Y_T = x_0) = P_\mu(X_0 = x_0, \dots, X_T = x_T).$$

Dabei ist $\mu > 0$ irgendeine Startverteilung, mit der die Kette gestartet wurde. Anschaulich bedeutet Reversibilität, dass die vorwärts und die rückwärts laufende Realisierung einem Betrachter gleich plausibel vorkommen würden (z.B. im dem Sinne, dass sie dieselbe Likelihood haben). Bezogen auf einen Schritt heißt das insbesondere, dass

$$P_\mu(Y_0 = y, Y_1 = x) = P_\mu(X_0 = x, X_1 = y) \quad \forall x, y \in S,$$

Mit der Abkürzung $q_{yx} = P(Y_1 = x | Y_0 = y)$ erhalten wir für die Übergangsmatrix $Q = (q_{xy})_{x,y \in S}$ der umgekehrten Kette $(Y_n)_{n \in \mathbb{N}_0}$ die Bedingung

$$\mu(y)q_{yx} = \mu(x)p_{xy} \quad \forall x, y \in S,$$

die, da Q eine stochastische Matrix ist, nur dann erfüllt sein kann, wenn μ eine stationäre Verteilung von P ist:

$$\mu(y) \sum_{x \in S} q_{yx} = \sum_{x \in S} \mu(x)p_{xy} \quad \forall y \in S \quad \iff \quad \mu^T = \mu^T P.$$

Das motiviert die folgende Definition.

Definition 3.59 Eine Markovkette heißt reversibel, wenn für ihre Übergangswahrscheinlichkeiten gilt:

$$\pi(y)p_{yx} = \pi(x)p_{xy} \quad \forall x, y \in S. \quad (3.11)$$

Die Bedingung (3.11) wird *detailliertes Gleichgewicht* (engl.: *detailed balance*) genannt.

Reversibilität bedeutet im Allgemeinen nicht, dass die Übergänge von x nach y bzw. y nach x gleich wahrscheinlich sind, sondern nur dass die Wahrscheinlichkeitsflüsse gleich sind, d.h.,

$$J_\pi(\{x\}, \{y\}) = J_\pi(\{y\}, \{x\}) \quad \forall x, y \in S.$$

Das bedeutet, dass der Übergang $x \rightarrow y$ im Mittel genauso oft vorkommt wie der Übergang $y \rightarrow x$, wobei sich der Zusatz *im Mittel* auf die Gewichtung der Zustände x und y mit der stationären Verteilung π bezieht. Im Vergleich dazu besagt der Satz über das stationäre Flussgleichgewicht, Satz 3.22, das für beliebige Markovketten mit stationärer Verteilung die Flüsse in einen Zustand und aus dem Zustand heraus gleich sind:

$$J_\pi(\{x\}, \{x\}^c) = J_\pi(\{x\}^c, \{x\}) \quad \forall x \in S.$$

Das detaillierte Gleichgewicht wird daher auch als *mikroskopisches Flussgleichgewicht* bezeichnet.

Anmerkung 3.60 Ist $\pi > 0$, so ist die Übergangsmatrix P zu einer symmetrischen Matrix ähnlich, d.h., es gibt eine Ähnlichkeitstransformation $P \mapsto V P V^{-1}$, die P symmetrisch macht (z.B. ist $V = \text{diag}\{\sqrt{\pi_1}, \sqrt{\pi_2}, \dots\}$ eine solche Transformation). Übergangsmatrizen von reversiblen Markovketten haben folglich reelle Eigenwerte.

Der Metropolis-Hastings-Algorithmus

Wir drehen nun den Spieß um und suchen zu einer gegebenen Verteilung $\pi > 0$ auf einem abzählbaren Zustandsraum, im folgenden *Zielverteilung* genannt, eine Markovkette, die π als eindeutige stationäre Verteilung hat. Die einfachste Möglichkeit ist, eine Übergangsmatrix zu raten, indem wir Einträge $p_{x,y} \in (0, 1)$, $y > x$ der oberen Dreiecksmatrix vorgeben und durch $p_{yx} = \pi(x)p_{yx}/\pi(y)$ und die Nebenbedingung $\sum_y p_{xy} = 1$ zu einer stochastischen Matrix ergänzen. Die so berechnete Matrix P ist nach Konstruktion reversibel, positiv rekurrent und hat π als eindeutige stationäre Verteilung

Eine andere, meist praktikablere Möglichkeit ist der Metropolis-Algorithmus, dessen Idee darin besteht, die Markovkette in einen Vorschlags- und einen Akzeptanzschritt zu zerlegen (engl.: *acceptance rejection algorithm*). Sei also $(X_n)_{n \in \mathbb{N}_0}$ eine Markovkette mit Übergangsmatrix P , die sich wie folgt zusammensetzt:

1. **Vorschlagsschritt:** Von $X_n = x$ wird nach der Vorschlagsmatrix $Q = (q_{xy})$ der Vorschlag $X_n^* = y$ generiert, wobei wir voraussetzen, dass Q stochastisch und irreduzibel sei und dass $q_{xy} \neq 0$ genau dann, wenn $q_{yx} \neq 0$.
2. **Akzeptanzschritt:** Der Vorschlag $X_n^* = y$ wird mit der Wahrscheinlichkeit

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)q_{yx}}{\pi(x)q_{xy}} \right\}$$

akzeptiert und $X_{n+1} = y$ gesetzt; wird der Vorschlag verworfen – das geschieht gerade mit Wahrscheinlichkeit $r(x, y) = 1 - a(x, y)$ – so wird $X_{n+1} = X_n$ gesetzt und ein neuer Vorschlag generiert.

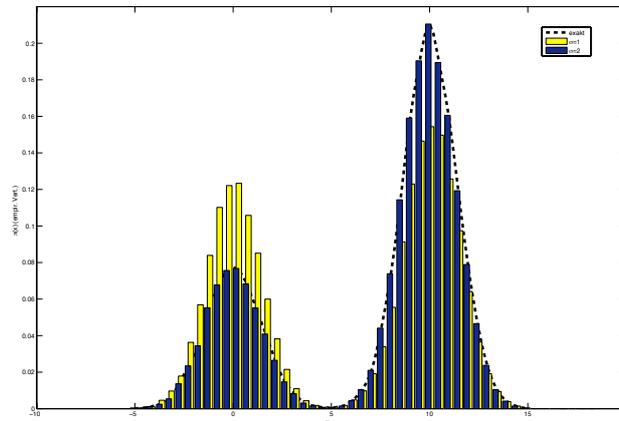


Abb. 11 Aus $N = 10^5$ Monte-Carlo-Punkten und verschiedenen Vorschlagswahrscheinlichkeiten $q(\cdot, \cdot; \sigma)$ erzeugte empirische Schätzer (Histogramme) für π (gelb: $\sigma = 1$, blau: $\sigma = 2$).

Die beiden letzten Schritte werden iteriert, bis eine ausreichend lange Realisierung von $(X_n)_{n \in \mathbb{N}_0}$ erzeugt wurde. Die Übergangsmatrix $P = (p_{xy})_{x, y \in S}$, die sich aus Vorschlags- und Akzeptanzwahrscheinlichkeiten zusammensetzt, ist reversibel mit stationärer Verteilung $\pi > 0$ und hat die Einträge

$$p_{xy} = \begin{cases} q_{xy}a_{xy}, & x \neq y \\ 1 - \sum_{z \in S, z \neq y} q_{xz}a_{xz}, & x = y \end{cases}.$$

Ist der Vorschlagsschritt symmetrisch, ist also $q_{xy} = q_{yx}$, so vereinfacht sich die Akzeptanzwahrscheinlichkeit zu

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\},$$

d.h., ein Vorschlag y wird immer dann sicher akzeptiert, wenn er die Likelihood – gemessen durch die Zielverteilung π – gegenüber dem alten Zustand x erhöht; ansonsten wird er entsprechend des Likelihood-Quotienten von x und y verworfen. Ein Vorteil des Metropolis-Verfahrens ist, dass es reicht, die Zielverteilung bis auf die Normierungskonstante anzugeben, da für den Akzeptanzschritt allein das Verhältnis $\pi(y)/\pi(x)$ relevant ist.

Beispiel 3.61 (Summe von Normalverteilungen) Wir betrachten ein Beispiel mit kontinuierlichem Zustandsraum. Unsere Zielverteilung sei die konvexe Linearkombination zweier eindimensionaler Normalverteilungen

$$\pi(x) = w\mathcal{N}(\mu_1, \sigma_1^2) + (1 - w)\mathcal{N}(\mu_2, \sigma_2^2), \quad w \in (0, 1).$$

Zur Berechnung von π generieren wir symmetrische Vorschläge gemäß der Vorschlagsdichte

$$q(x, y) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - y)^2}{2\sigma^2}\right),$$

die (wegen der Symmetrie von q) mit Wahrscheinlichkeit

$$a(x, y) = \min \left\{ 1, \frac{\pi(y)}{\pi(x)} \right\}$$

akzeptiert werden. Abbildung 11, die die empirischen Verteilungen der entsprechenden Markovkette nach $N = 10^5$ Schritten für verschiedene Vorschlagswahrscheinlichkeiten $q(x, y) = q(x, y; \sigma)$ zeigt, demonstriert, dass die Konvergenz des Metropolis-Hastings-Algorithmus stark von der Wahl der Vorschlagswahrscheinlichkeit abhängt.

4 Zeitdiskrete Martingale

Martingale sind die mathematische Variante dessen, was man gemeinhin unter einem fairen Spiel versteht. Unter einem *Spiel* werden wir im folgenden einen zeitdiskreten stochastischen Prozess $S = (S_n)_{n \geq 0}$ verstehen. Pro Einsatz sei $S_n - S_{n-1}$ der Gewinn in Runde $n \geq 1$ (der auch negativ ausfallen kann). Wir bezeichnen das Spiel als *fair*, wenn der erwartete Gewinn pro Runde jeweils 0 ist, also

$$\mathbf{E}[S_n - S_{n-1} | \text{Vorgeschichte}] = 0,$$

bzw. als *unfair*, wenn der erwartete Gewinn ungleich 0 ist, sprich:

$$\mathbf{E}[S_n - S_{n-1} | \text{Vorgeschichte}] \leq 0 \quad (\text{zu unseren Ungunsten}).$$

Definition 4.1 (Endliche Martingalfolge) Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum. Eine Martingalfolge der Länge n ist eine Folge von Zufallsvariablen X_1, \dots, X_n und Sub- σ -Algebren $\mathcal{F}_1, \dots, \mathcal{F}_n$, für die gilt:

1. Jedes X_i ist eine integrierbare Zufallsvariable, die bezüglich \mathcal{F}_i messbar ist.
2. Die $\mathcal{F}_1, \dots, \mathcal{F}_n$ bilden eine aufsteigende Kette (Filtration), d.h. $\mathcal{F}_i \subset \mathcal{F}_{i+1}$.
3. Für jedes $i = 1, \dots, n-1$ gilt fast sicher, dass $X_i = \mathbf{E}[X_{i+1} | \mathcal{F}_i]$.

Anmerkung 4.2 Jeder stochastische Prozess ist bezüglich seiner eigenen Filtration $\{\mathcal{F}_k : k \geq 0\}$, $\mathcal{F}_k = \sigma(\{X_0, \dots, X_k\})$ messbar. Bezüglich einer Filtration messbare Prozesse heißen *adaptiert*.

4.1 Bedingte Wahrscheinlichkeiten und Erwartungswerte, Forts.

Der Martingal-Begriff basiert wesentlich auf dem Konzept der bedingten Erwartung. Dazu erinnern wir uns zunächst an die Definition der bedingten Erwartung auf Seite 52: Es sei $X : \Omega \rightarrow \mathbf{R}$ eine integrierbare Zufallsvariable auf (Ω, \mathcal{F}, P) und $\mathcal{G} \subset \mathcal{F}$ eine Sub- σ -Algebra. Die bedingte Erwartung $Z = \mathbf{E}[X | \mathcal{G}]$ ist \mathcal{G} -messbar und es gilt

$$\int_G X dP = \int_G Z dP \quad \forall G \in \mathcal{G}.$$

Satz von Radon-Nikodym

Um zu zeigen, dass die Definition der bedingten Erwartung sinnvoll ist, d.h., dass eine solche Zufallsvariable Z existiert und dass sie fast sicher eindeutig ist, benötigen wir die folgenden Werkzeuge.

Definition 4.3 (Absolute Stetigkeit) Für zwei Maße μ, λ auf (Ω, \mathcal{F}) heißt μ *absolut stetig* bezüglich λ (symbolisch: $\mu \ll \lambda$), wenn für jedes $F \in \mathcal{F}$ mit $\lambda(F) = 0$ auch $\mu(F) = 0$ gilt.

Satz 4.4 (Satz von Radon-Nikodym³⁸) *Es seien μ, λ Maße auf (Ω, \mathcal{F}) und λ σ -endlich. Dann sind die folgenden Aussagen äquivalent:*

1. $\mu \ll \lambda$
2. *Es gibt eine Dichtefunktion $f \geq 0$, so dass für alle $F \in \mathcal{F}$ gilt*

$$\mu(F) = \int_F f(\omega) d\lambda(\omega)$$

Beweis. Die Implikation 1. \Rightarrow 2. ist leicht; an Stelle der Umkehrung beweisen wir eine "entschärfte" Version des Satzes (vgl. [10, Kapitel 2.5] für den vollen Beweis des Satzes):

Lemma 4.5 *Sind μ, λ endlich mit $\mu \leq \lambda$, d.h. $\mu(E) \leq \lambda(E) \forall E \in \mathcal{F}$, so gibt es eine Dichtefunktion $f : \Omega \rightarrow [0, 1]$ mit $f = d\mu/d\lambda$.*

³⁸ Johann Radon (1887–1956), österr. Mathematiker; Otton Marcin Nikodym (1887–1974), poln. Mathematiker

Wir definieren zunächst die gewichteten L^p -Räume

$$L^p(\Omega, \nu) := \left\{ u: \Omega \rightarrow \mathbf{R} : \left(\int_{\Omega} |u(\omega)|^p d\nu(\omega) \right)^{1/p} < \infty \right\}.$$

mit der zugehörigen Norm $\|\cdot\|_{p,\nu}$. Wegen $\mu \leq \lambda$ und $\lambda(\Omega) < \infty$ gilt dann $L^2(\Omega, \lambda) \subset L^2(\Omega, \mu) \subset L^1(\Omega, \mu)$, woraus folgt, dass die Linearform

$$l: L^2(\Omega, \lambda) \rightarrow \mathbf{R}, g \mapsto \int g d\mu$$

endlich (d.h., $|l(g)| \leq C\|g\|_{2,\lambda}$), mithin stetig ist. Nach dem Darstellungssatz von Riesz (siehe z.B. [17, Satz V.3.6]) gibt es dann ein eindeutiges $f \in L^2(\Omega, \lambda)$ mit der Eigenschaft

$$\int g d\mu = \langle g, f \rangle_{\lambda} := \int gf d\lambda \quad \forall g \in L^2(\Omega, \lambda).$$

Dieses f ist die gesuchte Dichte. Insbesondere gilt dann für $g = \chi_A$, $A \in \mathcal{F}$, dass

$$\mu(A) = \int \chi_A d\mu = \int \chi_A f d\lambda,$$

d.h., $f = d\mu/d\lambda$. Es bleibt zu zeigen, dass fast sicher $0 \leq f \leq 1$. Dazu nehmen wir an, dass $\lambda(\{f < 0\}) > 0$ und erhalten einen Widerspruch; denn dann ist

$$\mu(\{f < 0\}) = \int_{\{f < 0\}} f d\lambda < 0,$$

im Widerspruch zur Maßeigenschaft von μ . Analog folgt aus $\lambda(\{f > 1\}) > 0$, dass

$$\mu(\{f > 1\}) = \int_{\{f > 1\}} f d\lambda > \lambda(\{f > 1\}),$$

im Widerspruch zu $\mu \leq \lambda$. Damit ist fast sicher $0 \leq f \leq 1$. □

Für endliche Maße ist die absolute Stetigkeit von Maßen eng mit der Stetigkeit von Funktionen verwandt. Das ist nicht weiter überraschend, wenn man sich klar macht, dass Maße nichts anderes als stetige Funktionen sind, deren Argumente messbare Mengen sind. Die folgende Aussage erklärt den Zusammenhang.

Lemma 4.6 (ϵ - δ -Kriterium) *Ein endliches Maß μ ist genau dann bezüglich einem σ -endlichem Maß λ absolut stetig, wenn es zu jedem $\epsilon > 0$ ein $\delta > 0$ gibt, so dass für alle $F \in \mathcal{F}$ gilt:*

$$\lambda(F) < \delta \quad \Rightarrow \quad \mu(F) < \epsilon. \tag{4.1}$$

Beweis. ÜA. □

Anmerkung 4.7 Der Ausdruck $f = d\mu/d\lambda$ heißt *Radon-Nikodym-Ableitung* von μ und λ . Ist $\mu \ll \lambda$ und $f > 0$, so ist zudem $\lambda \ll \mu$ und es gilt $f^{-1} = d\lambda/d\mu$.

Existenz und fast sichere Eindeutigkeit

Kommen wir zurück zur Frage von Existenz und Eindeutigkeit der bedingten Erwartung $\mathbf{E}[X|\mathcal{G}]$.

Existenz: O.B.d.A. sei $X \geq 0$. Dann definiert

$$\mu(F) = \int_F X dP, \quad F \in \mathcal{F}$$

ein Maß auf $\mathcal{G} \subset \mathcal{F}$ mit $\mu \ll P$, dass wegen $\mathbf{E}[|X|] < \infty$ endlich ist. Aus dem Satz von Radon-Nikodym folgt dann, dass es eine Funktion $f \geq 0$ gibt, die bezüglich \mathcal{G} messbar ist und für die gilt

$$\int_G X dP = \mu(G) = \int_G f dP \quad \forall G \in \mathcal{G}.$$

Folglich haben wir mit

$$f(\omega) = \frac{d\mu}{dP} \Big|_{\mathcal{G}}$$

einen Kandidaten für die bedingte Erwartung $Z(\omega) = \mathbf{E}[X|\mathcal{G}]$ gefunden. Für integrierbare Zufallsvariable X , die positive und negative Werte annehmen, setzen wir einfach wieder

$$\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X^+|\mathcal{G}] - \mathbf{E}[X^-|\mathcal{G}],$$

wobei $X^\pm = \max\{0, \pm X\}$ den Positiv- und Negativanteil von X bezeichnet.

Fast sichere Eindeutigkeit: Es seien Z und \tilde{Z} zwei Kandidaten für $\mathbf{E}[X|\mathcal{G}]$. Wegen der Linearität gilt

$$\int_G (Z - \tilde{Z}) dP = 0 \quad \forall G \in \mathcal{G}.$$

Insbesondere folgt damit für $E = \{\omega: Z(\omega) > \tilde{Z}(\omega)\} \in \mathcal{G}$ die Aussage $P(E) = 0$ bzw. analog $P(F) = 0$ für $F = \{\omega: Z(\omega) < \tilde{Z}(\omega)\} \in \mathcal{G}$. Also ist fast sicher $Z = \tilde{Z}$.

Eigenschaften der bedingten Erwartung

Die bedingte Erwartung $Z = \mathbf{E}[X|\mathcal{G}]$ erbt die folgenden elementaren Eigenschaften des Lebesgueintegrals bzw. von $\mathbf{E}[\cdot]$, allerdings gelten die Aussagen, da Sie sich nun auf Zufallsvariable beziehen, nur noch fast sicher:

- (Positivität) Aus $X \geq 0$ folgt $Z \geq 0$.
- (Linearität) Für alle $\alpha_1, \alpha_2 \in \mathbf{R}$ ist $\mathbf{E}[\alpha_1 X_1 + \alpha_2 X_2|\mathcal{G}] = \alpha_1 \mathbf{E}[X_1|\mathcal{G}] + \alpha_2 \mathbf{E}[X_2|\mathcal{G}]$.
- (Monotonie) Ist $X_1 \leq X_2$, so ist auch $\mathbf{E}[X_1|\mathcal{G}] \leq \mathbf{E}[X_2|\mathcal{G}]$.

Darüber hinaus gelten die Konvergenzsätze aus Abschnitt 1.4 analog für die bedingte Erwartung, was man leicht einsieht, indem man jeweils $\mathbf{E}[\cdot]$ durch $\mathbf{E}[\cdot|\mathcal{G}]$ ersetzt (vgl. [10, Kap. 8.2]).

Lemma 4.8 (Turmeigenschaft) Für σ -Algebren $\mathcal{H} \subset \mathcal{G} \subset \mathcal{F}$ gilt

$$\mathbf{E}[\mathbf{E}[X|\mathcal{G}]|\mathcal{H}] = \mathbf{E}[X|\mathcal{H}] = \mathbf{E}[\mathbf{E}[X|\mathcal{H}]|\mathcal{G}] \quad (f.s.)$$

Damit ist insbesondere fast sicher $\mathbf{E}[Z] = \mathbf{E}[X]$.

Beweis. Das rechte Gleichheitszeichen gilt nach Definition der bedingten Erwartung, weil wegen $\mathcal{H} \subset \mathcal{G}$ die Zufallsvariable $\mathbf{E}[X|\mathcal{H}]$ auch \mathcal{G} -messbar ist. Um die Gleichheit links zu beweisen, betrachten wir ein beliebiges $H \in \mathcal{H}$. Wegen $H \in \mathcal{G}$ gilt dann

$$\mathbf{E}[\mathbf{E}[\mathbf{E}[X|\mathcal{G}]|\mathcal{H}]\chi_H] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]\chi_H] = \mathbf{E}[X\chi_H] = \mathbf{E}[\mathbf{E}[X|\mathcal{H}]\chi_H].$$

Die Ausdrücke links und rechts unter dem äußeren Erwartungswert sind \mathcal{H} -messbar, woraus die behauptete Gleichheit folgt. Mit $\mathcal{H} = \{\emptyset, \Omega\}$ folgt sofort der Rest der Aussage. \square

Lemma 4.9 (Produkteigenschaft) Es sei Y eine beschränkte und \mathcal{G} -messbare Zufallsvariable. Dann ist

$$\mathbf{E}[XY|\mathcal{G}] = Y\mathbf{E}[X|\mathcal{G}] \quad (f.s.)$$

Beweis. Sei $E \in \mathcal{G}$. Dann gilt für alle $G \in \mathcal{G}$, dass

$$\mathbf{E}[\chi_E \mathbf{E}[X|\mathcal{G}]\chi_G] = \mathbf{E}[\mathbf{E}[X|\mathcal{G}]\chi_{E \cap G}] = \mathbf{E}[X\chi_{E \cap G}] = \mathbf{E}[(X\chi_E)\chi_G]$$

Damit gilt die Aussage für Funktionen $Y = \chi_E$ mit $E \in \mathcal{G}$. Wegen der Linearität der bedingten Erwartung gilt sie auch für Treppenfunktionen. Den Rest der Behauptung folgt aus dem Satz von der beschränkten Konvergenz und der gleichmäßigen Approximierbarkeit beschränkter messbarer Funktionen durch Treppenfunktionen. \square

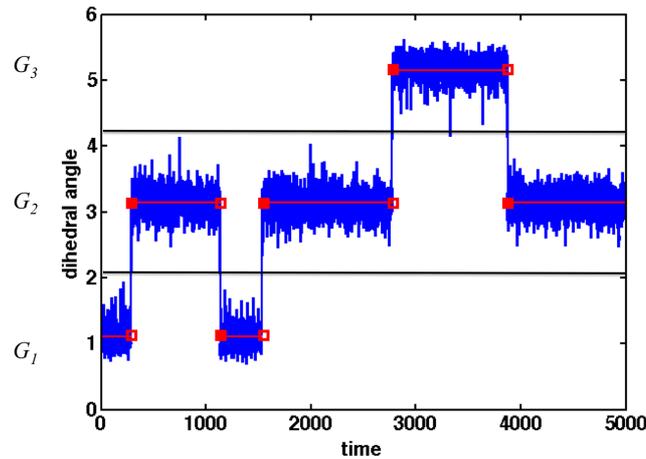


Abb. 12 Vergrößerte Zufallsvariable $Z \in \{\mathbf{E}[X|G_i], i = 1, 2, 3\}$.

Anmerkung 4.10 Lemma 4.9 gilt auch ohne die Annahme, dass Y beschränkt ist, sofern $\mathbf{E}[|XY|] < \infty$. Der Beweis folgt demselben Muster.

Lemma 4.11 (Bedingte Jensens-Ungleichung) Für konvexe, integrierbare Funktionen $\varphi: \mathbf{R} \rightarrow \mathbf{R}$ gilt fast sicher die Jensen-Ungleichung

$$\varphi(Z) \leq \mathbf{E}[\varphi(X)|\mathcal{G}].$$

Insbesondere ist also $\mathbf{E}[\varphi(Z)] \leq \mathbf{E}[\varphi(X)]$.

Beweis. Wir rufen uns in Erinnerung (vgl. Abschnitt 2.5), dass sich jede konvexe Funktion als Legendre-Transformierte gemäß $\varphi(x) = \sup\{sx - \psi(s) : s \in \mathbf{Q}\}$ mit einer geeigneten Funktion $\psi: \mathbf{Q} \rightarrow \mathbf{R}$ darstellen lässt. (Wegen der Dichtheit von \mathbf{Q} in \mathbf{R} reicht es, das Supremum über \mathbf{Q} zu bilden.) Da \mathbf{Q} abzählbar ist, folgt mit der Monotonie der bedingten Erwartung, dass

$$\begin{aligned} \mathbf{E}[\varphi(X)|\mathcal{G}] &= \mathbf{E}[\sup_s \{sX - \psi(s)\}|\mathcal{G}] \\ &\geq \sup_s \{s\mathbf{E}[X|\mathcal{G}] - \psi(s)\} \\ &= \varphi(\mathbf{E}[X|\mathcal{G}]). \end{aligned}$$

Damit ist der erste Teil der Behauptung bewiesen; der Rest folgt aus der Turmeigenschaft, Lemma 4.8, zusammen mit der Monotonie des Erwartungswertes. \square

Bedingte Erwartung als Zufallsvariable

Wir wollen anhand einiger Beispiele verdeutlichen, in welchem Sinne die bedingte Erwartung als Zufallsvariable interpretiert werden kann. Dazu sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum und $G = \{G_i\}_{i=1, \dots, M}$ eine Partition von Ω , d.h., die $G_i \in \mathcal{F}$ sind paarweise disjunkte Teilmengen von Ω mit der Eigenschaft $\cup_i G_i = \Omega$. Wir betrachten eine Zufallsvariable X auf (Ω, \mathcal{F}, P) und definieren die "vergrößerte" Zufallsvariable $Z = \mathbf{E}[X|G]$ durch

$$Z: \Omega \rightarrow \mathbf{R}, \quad \omega \mapsto \sum_i \chi_{G_i}(\omega) \mathbf{E}[X|G_i].$$

Die so definierte Zufallsvariable $Z = \mathbf{E}[X|G]$ lässt sich problemlos zu einer Zufallsvariable $\mathbf{E}[X|\mathcal{G}]$ mit $\mathcal{G} = \sigma(G) \subset \mathcal{F}$ als der von G erzeugten σ -Algebra erweitern. Die bedingte Erwartung $\mathbf{E}[X|\mathcal{G}]$ wird damit eine Zufallsvariable, die jedem ω jeweils den Erwartungswert von X über die $\omega \in G_i \in \mathcal{G}$ zuweist (siehe Abb. 12).

Beispiel 4.12 Wir illustrieren den letzten Punkt mit einigen Beispielen.

1. *Würfeln:* Es sei $X(\omega) \in \{1, \dots, 6\}$ die Augenzahl beim Würfeln. Nun definiert beispielsweise

$$G = \{\{1, 3, 5\}, \{2, 4, 6\}\}$$

eine Partition der Zahlen 1 bis 6 in ungerade und gerade Augenzahlen. Es lässt sich leicht nachrechnen, dass $\mathbf{E}[X|\text{ungerade AZ}] = 3$ bzw. $\mathbf{E}[X|\text{gerade AZ}] = 4$. Die vergrößerte Zufallsvariable $Z = \mathbf{E}[X|G]$ ist dann die Summe der Erwartungswerte von ungeraden und geraden Zahlen, d.h.,

$$\mathbf{E}[X|G] = 3\chi_{\{1,3,5\}}(X(\omega)) + 4\chi_{\{2,4,6\}}(X(\omega))$$

nimmt die Werte 3 bzw. 4 an.

2. *Zufallsvariable I:* Sind $X: \Omega \rightarrow \mathbf{R}$ und $\xi: \Omega \rightarrow \mathbf{R}$ Zufallsvariable auf (Ω, \mathcal{F}, P) , wobei ξ nur endlich viele Werte a_1, \dots, a_M annimmt, so wird durch

$$G = \bigcup_{i=1}^M G_i, \quad G_i = \{\omega: X(\omega) = a_i\}$$

eine Partition von Ω definiert. Dann folgt sofort, dass $\mathbf{E}[X|G]$ bezüglich $\mathcal{G} = \sigma(G)$ messbar ist und dass gilt

$$\mathbf{E}[X|\xi = a_i] = \mathbf{E}[X|G_i].$$

($\mathbf{E}[X|\xi] := \mathbf{E}[X|G]$ ist eine Funktion von ξ .)

3. *Zufallsvariable II:* Nun sei ξ stetig verteilt, und wir nehmen an, dass X und ξ eine gemeinsame Dichte $f(x, a)$ haben. Für eine integrable Funktion g lässt sich $\mathbf{E}[g(X, \xi)|\xi]$ berechnen, indem wir ξ zunächst durch eine diskrete Zufallsvariable approximieren und dann wie oben verfahren: Sei also

$$\tilde{\xi} = \sum_{k \in \mathbf{N}} h \left(k + \frac{1}{2} \right) \chi_{[hk, h(k+1))}(\xi)$$

eine diskrete Approximation für ein hinreichend kleines $h > 0$. Offensichtlich gilt dann

$$P\left(\tilde{\xi} = h \left(k + \frac{1}{2} \right)\right) = \int_{hk}^{h(k+1)} \left(\int_{-\infty}^{\infty} f(x, a) dx \right) da$$

und folglich für h mit $a_k = h(k + 1/2) \rightarrow a$:

$$\begin{aligned} \mathbf{E}[g|\xi = a] &\approx \mathbf{E}[g|\tilde{\xi} = a_k] \\ &\approx \frac{\int_{hk}^{h(k+1)} \left(\int_{-\infty}^{\infty} g(x, a) f(x, a) dx \right) da}{h \int_{-\infty}^{\infty} f(x, a_k) dx} \\ &\rightarrow \frac{\int_{-\infty}^{\infty} g(x, a) f(x, a) dx}{\int_{-\infty}^{\infty} f(x, a) dx}, \end{aligned}$$

wobei wir im Nenner der zweiten Zeile die Rechtecksumme zur Approximation des Integrals verwendet haben. Der Grenzübergang $h \rightarrow 0$ (d.h. $h(k + 1/2) \rightarrow a$ im Limes $k \rightarrow \infty$) in der letzten Zeile entspricht dann in wesentlichen dem Berechnen der Radon-Nikodym-Ableitung.

Bestapproximationseigenschaft bedingter Erwartungen

Wir erinnern daran, dass für eine Zufallsvariable X , deren erste beide Momente existieren gilt:

$$\mathbf{E}[X] = \operatorname{argmin}_{c=\text{const.}} \mathbf{E}[(X - c)^2].$$

Mit anderen Worten: $\mathbf{E}[X]$ ist die beste Approximation einer Zufallsvariable $X(\omega)$ durch eine Konstante c im Sinne der kleinsten quadratischen Abweichungen (=Varianz). Nun ist die bedingte Erwartung $\mathbf{E}[X|\mathcal{G}]$ gerade der Erwartungswert von X , wenn für jedes $G \in \mathcal{G}$ nur die Information $\omega \in G$ bzw. $\omega \notin G$ bekannt ist, aber nicht mehr. Der folgende Satz besagt gerade, dass der bedingte Erwartungswert die Bestapproximation von X als Funktion der Information, die durch \mathcal{G} gegeben ist, liefert.

Satz 4.13 (Bestapproximation) Sei X eine Zufallsvariable auf (Ω, \mathcal{F}, P) mit endlicher Varianz und $\mathcal{G} \subset \mathcal{F}$ eine Sub- σ -algebra von \mathcal{F} . Dann nimmt die Funktion

$$h_X: Y \mapsto \mathbf{E}[(X - Y)^2],$$

unter der Bedingung, dass Y bezüglich \mathcal{G} messbar und dass $\mathbf{V}(Y \chi_G) < \infty$ für alle $G \in \mathcal{G}$ ist, an der Stelle

$$Y_* = \mathbf{E}[X|\mathcal{G}]$$

ihr Minimum an.

Beweis. Unter der Annahme, dass Y \mathcal{G} -messbar sei, gilt nach Lemma 4.9, dass

$$\mathbf{E}[XY|\mathcal{G}] = Y\mathbf{E}[X|\mathcal{G}] = YY_*$$

und damit

$$\mathbf{E}[XY] = \mathbf{E}[YY_*].$$

Also ist

$$\begin{aligned} \mathbf{E}[(X - Y)^2] - \mathbf{E}[(X - Y_*)^2] &= \mathbf{E}[Y^2] - \mathbf{E}[Y_*^2] + 2\mathbf{E}[XY_*] - 2\mathbf{E}[XY] \\ &= \mathbf{E}[Y^2] - 2\mathbf{E}[YY_*] + \mathbf{E}[Y_*^2] \\ &= \mathbf{E}[(Y - Y_*)^2] \end{aligned}$$

wobei wir in der letzten Zeile ausgenutzt haben, dass $\mathbf{E}[XY_*] = \mathbf{E}[X(Y = Y_*)] = \mathbf{E}[Y_*^2]$. Somit ist

$$h_X(Y) = \mathbf{E}[(X - Y_*)^2] + \mathbf{E}[(Y - Y_*)^2],$$

woraus folgt, dass die Funktion h_X gerade an der Stelle $Y_* = \mathbf{E}[X|\mathcal{G}]$ minimal wird. □

Beispiel 4.14 Wir schließen unseren kleinen Exkurs zu bedingten Erwartungen mit einigen Beispielen.

1. *Keine Information (minimale σ -Algebra)*: Ist $\mathcal{G} = \{\emptyset, \Omega\}$ die minimale Sub- σ -Algebra von \mathcal{F} , so ist keine Information über X vorhanden, d.h., wir haben fast sicher

$$\mathbf{E}[X|\{\emptyset, \Omega\}](\omega) = \mathbf{E}[X] \quad \forall \omega \in \Omega.$$

2. *Vollständige Information (maximale σ -Algebra)*: Ist dagegen $\mathcal{G} = \sigma(X)$, so kennen wir alle möglichen Möglichkeiten $\mathcal{G} = \mathcal{F}$ für $X(\cdot)$, und die optimale Approximation von X ist X selbst, d.h.,

$$\mathbf{E}[X|\sigma(X)](\omega) = X(\omega) \quad \forall \omega \in \Omega.$$

3. *Martingalfolge*: Hier ist nach Definition $X_i = \mathbf{E}[X_{i+1}|\sigma(X_1, \dots, X_i)]$ für alle $i = 1, \dots, n-1$, d.h., X_i ist die optimale Vorhersage für X_{i+1} , wenn X_1, X_2, \dots, X_i bekannt sind.

4.2 Spielsysteme und stochastische Integrale

Wir wollen uns nun wieder den Martingalen zuwenden und zunächst einige Begriffe klären.

Definition 4.15 (Sub- und Supermartingal) Es sei $(X_n)_{n \geq 0}$ ein stochastischer Prozess, der bezüglich einer Filtration $\{\mathcal{F}_n: n \geq 0\}$ adaptiert ist. Ist X_n für jedes $n \geq 0$ integrierbar, so heißt $(X_n)_{n \geq 0}$

- *Martingal*, wenn fast sicher $X_n = \mathbf{E}[X_{n+1}|\mathcal{F}_n]$ ist,
- *Submartingal*, wenn fast sicher $X_n \leq \mathbf{E}[X_{n+1}|\mathcal{F}_n]$ ist,

- *Supermartingal*, wenn fast sicher $X_n \geq \mathbf{E}[X_{n+1}|\mathcal{F}_n]$ ist.

Anmerkung 4.16 Erfüllt $(X_n)_{n \geq 0}$ die Martingaleigenschaft, so bedeutet das, dass die mittleren Zuwächse von $X_{i+1} - X_i$ null sind. Mit anderen Worten: $(X_n)_{n \geq 0}$ ist genau dann ein Martingal, wenn

$$\mathbf{E}[X_{i+1} - X_i|\mathcal{F}_i] = 0 \quad i \geq 0.$$

Insbesondere folgt daraus, dass $\mathbf{E}[X_i] = \mathbf{E}[X_0]$ für alle $i \geq 0$, denn

$$\mathbf{E}[X_i] = \mathbf{E}[\mathbf{E}[X_{i+1}|\mathcal{F}_i]] = \mathbf{E}[X_{i+1}].$$

Mit anderen Worten: Martingale haben konstante Erwartung. Analog zeigt man, dass der Erwartungswert von Submartingalen im Mittel steigt bzw. dass der Erwartungswert von Supermartingalen im Mittel sinkt.

Beispiel 4.17 Wir betrachten zwei Folgen von Zufallsvariablen, die die Martingaleigenschaft erfüllen.

1. Ist $(X_i)_{i \geq 1}$ eine Folge von unabhängigen Zufallsvariablen mit jeweils endlichen Erwartungswerten, so ist

$$M_n = \sum_{i=1}^n (X_i - \mathbf{E}[X_i])$$

ein Martingal bezüglich der durch $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ erzeugten Filtration, denn

$$\mathbf{E}[M_{n+1} - M_n|\mathcal{F}_n] = \mathbf{E}[X_{n+1} - \mathbf{E}[X_{n+1}]|\mathcal{F}_n] = 0.$$

2. Sei X integrierbar und $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \dots$ eine Filtration. Dann definiert

$$X_i = \mathbf{E}[X|\mathcal{F}_i], \quad i = 1 \geq 0$$

ein Martingal. Mit anderen Worten: X_i ist die sukzessive Prognose von X für gegebene, mit jedem Zeitschritt besser werdende Information \mathcal{F}_i . Dass X_i ein Martingal ist, sieht man mit Hilfe der Turmeigenschaft:

$$\begin{aligned} \mathbf{E}[X_{i+1}|\mathcal{F}_i] &= \mathbf{E}[\mathbf{E}[X|\mathcal{F}_{i+1}]|\mathcal{F}_i] \\ &= \mathbf{E}[X|\mathcal{F}_i] \\ &= X_i. \end{aligned}$$

Lemma 4.18 Ist $\{(X_i, \mathcal{F}_i) : i \geq 1\}$ ein Martingal und φ eine konvexe (konkave) Funktion, so dass $\varphi(X_i)$ für alle $i \in \mathbf{N}$ integrierbar ist, dann ist $\{(\varphi(X_i), \mathcal{F}_i) : i \geq 1\}$ eine Submartingal (Supermartingal).

Beweis. Die Behauptung folgt aus der bedingten Jensenschen Ungleichung, Lemma 4.11). □

Faire und unfaire Spiele

Wir werden nun der Frage nachgehen, ob sich ein faires Spiel durch die geschickte Wahl der Einsätze in ein unfaires verwandeln lässt. (Die intuitive Antwort "nein" ist, wie sich gleich herausstellen wird, richtig.)

Definition 4.19 (Prävisibler Prozess) Es sei $(H_n)_{n \geq 1}$ ein stochastischer Prozess und $(\mathcal{F}_n)_{n \geq 0}$ eine Filtration. Falls H_n für alle $n \in \mathbf{N}$ bezüglich \mathcal{F}_{n-1} messbar ist, so heißt $(H_n)_{n \geq 1}$ *prävisibel* bzw. *vorhersagbar*.

Definition 4.20 (Spielsystem, stochastisches Integral, Martingaltransformation) Es sei $(H_n)_{n \geq 1}$ prävisibel und $(X_n)_{n \geq 0}$ adaptiert, so dass $H_n(X_n - X_{n-1})$ für alle $n \in \mathbf{N}$ integrierbar ist.

- $H = (H_n)_{n \geq 1}$ heißt *Spielsystem*.
- Als *stochastisches Integral* bezeichnen wir

$$(H \circ X)_n := X_0 + \sum_{k=1}^n H_k(X_k - X_{k-1}) \quad (n \geq 1).$$

- Ist $(X_n)_{n \geq 0}$ ein Martingal, so wird $(H \circ X)_n$ *Martingaltransformation* genannt.

Beispiel 4.21 Wir betrachten irgendein Glücksspiel. Dazu nehmen wir an, dass H_n für alle n beschränkt und X_n für alle n integrel sei. Ferner bezeichne

- X_0 das Startkapital,
- $X_n - X_{n-1}$ den Gewinn in Runde $n \geq 1$ pro Euro Spieleinsatz,
- H_n den Einsatz in Runde $n \geq 1$.

H_n wird nach $n - 1$ gespielten Runden gesetzt und erfüllt damit unsere Messbarkeitseigenschaft bezüglich \mathcal{F}_{n-1} (konkret: H_n hängt nur von X_0, \dots, X_{n-1} ab). Die Bilanz nach N Spielrunden ist dann

$$(H \circ X)_N = X_0 + \sum_{n=1}^N H_n(X_n - X_{n-1}).$$

Wir werden nun die mathematische Antwort auf die eingangs gestellte Frage nach der Möglichkeit, ein faires Spiel durch den Wetteinsatz zu unseren Gunsten zu beeinflussen, geben.

Satz 4.22 Sei $H = (H_n)_{n \geq 1}$ ein Spielsystem, das an eine Filtration $\mathcal{F} = (\mathcal{F}_n)_{n \geq 0}$ adaptiert sei, und $X = (X_n)_{n \geq 0}$ sei integrel. Dann gelten die folgenden Aussagen:

- Ist X ein Martingal (bezüglich \mathcal{F}), so ist auch $H \circ X$ ein Martingal.
- Ist $H \geq 0$ und X ein Submartingal (Supermartingal), so ist auch $H \circ X$ ein Submartingal (Supermartingal).

Beweis. Nach Voraussetzung ist $H \circ X$ integrel und adaptiert. Wegen der \mathcal{F}_{n-1} -Messbarkeit von H_n gilt

$$\begin{aligned} \mathbf{E}[(H \circ X)_n - (H \circ X)_{n-1} | \mathcal{F}_{n-1}] &= \mathbf{E}[H_n(X_n - X_{n-1}) | \mathcal{F}_{n-1}] \\ &= H_n \mathbf{E}[(X_n - X_{n-1}) | \mathcal{F}_{n-1}], \end{aligned}$$

woraus folgt, dass

$$\mathbf{E}[(H \circ X)_n | \mathcal{F}_{n-1}] - (H \circ X)_{n-1} = \begin{cases} 0, & X \text{ ist Martingal} \\ \geq 0, & H \geq 0 \text{ und } X \text{ ist Submartingal} \\ \leq 0, & H \geq 0 \text{ und } X \text{ ist Supermartingal.} \end{cases}$$

□

Beispiel 4.23 (Diskrete Brownsche Bewegung) Gegen sei eine beschränkte und Lipschitz-stetige Funktion $\sigma: \mathbf{R} \rightarrow \mathbf{R}$. Wir betrachten eine diskrete Brownsche Bewegung $(X_n)_{n \geq 0}$ zum Zeitschritt Δt mit

$$X_{n+1} = X_n + \sigma(X_n)\xi_{n+1}, \quad X_0 = 0,$$

wobei die $\xi_k \in \sim \mathcal{N}(0, \Delta t)$ IID Zufallsvariable sind (vgl. Beispiel 2.45). Aus der Darstellung

$$\begin{aligned} X_N &= \sum_{n=1}^N (X_n - X_{n-1}) + X_0 \\ &= \sum_{n=1}^N \sigma(X_{n-1})\xi_n. \end{aligned}$$

und Satz 4.22 folgt sofort, dass die diskrete Brownsche Bewegung X_n ein Martingal ist, denn der Integrand $\sigma(X_{n-1})$ ist nach Konstruktion beschränkt und bezüglich \mathcal{F}_{n-1} messbar, wenn X_n adaptiert ist und $\mathbf{E}[\xi_n] = 0$.

Anmerkung 4.24 Die obige Konstruktion des stochastischen Integrals und seiner Martingalerhaltungseigenschaft spielt in der Theorie der stochastischen Differentialgleichungen (siehe [14]) eine wichtige Rolle. Dabei betrachtet man Integranden der Form $H_n = H_n(X_{n-1})$ und trifft Aussagen über Ausdrücke der Form

$$\int_0^T H(X_t) dX_t := \lim_{N \rightarrow \infty} \sum_{n=1}^N H_n(X_n^N)(X_n^N - X_{n-1}^N).$$

4.3 Gleichgradige Integrierbarkeit

Ziel dieses Abschnittes ist es, den Begriff der Integrierbarkeit für Folgen von Zufallsvariablen zu präzisieren. Das zentrale Resultat wird eine Verallgemeinerung des Satzes von der majorisierten Konvergenz, Satz 1.40, sein, der beim Beweis der Stoppsätze im nächsten Abschnitt eine wichtige Rolle spielen wird.³⁹

Definition 4.25 (Gleichgradige Integrierbarkeit) Es sei $I \neq \emptyset$ eine nichtleere Indexmenge. Eine Familie $(X_i)_{i \in I}$ von Zufallsvariablen heißt *gleichgradig integrierbar*, wenn für $c \rightarrow \infty$ gilt:

$$\sup_{i \in I} \mathbf{E}[|X_i| \chi_{\{|X_i| \geq c\}}] \rightarrow 0. \quad (4.2)$$

Anmerkung 4.26 Ist I endlich, so besteht kein Unterschied zwischen Integrierbarkeit und gleichgradiger Integrierbarkeit, d.h., jede integrierbare Zufallsvariable erfüllt (4.2). Das sieht man, indem man durch

$$\mu(F) = \int_F |X| dP, \quad F \in \mathcal{F}$$

ein Maß definiert, dass wegen $\mathbf{E}[|X|] < \infty$ endlich ist und das bezüglich P absolut stetig ist. Daraus folgt aber sofort, dass $F = \{|X| = \infty\}$ eine μ -Nullmenge ist, denn $P(F) = 0$ und $\mu \ll P$. Wegen der Stetigkeit von Maßen ist insbesondere μ stetig und folglich lässt sich zu jedem $\epsilon > 0$ ein $c^* > 0$ finden, so dass

$$\mu(|X| \geq c) = \mathbf{E}[|X| \chi_{\{|X| \geq c\}}] < \epsilon \quad \forall c \geq c^*.$$

Da $\epsilon > 0$ beliebig ist, folgt die sofort die Aussage. Analog beweist man, dass jede endliche Familie von integrierbaren Zufallsvariablen gleichgradig integrierbar ist.

Lemma 4.27 Es sei Y eine integrierbare Zufallsvariable und $\{\mathcal{F}_i : i \in I\}$ eine Familie von Sub- σ -Algebren von \mathcal{F} . Dann ist $X_i = \mathbf{E}[Y | \mathcal{F}_i]$, $i \in I$, gleichgradig integrierbar.

Beweis. Wir definieren das endliche Maß

$$\mu(F) = \int_F |Y| dP, \quad F \in \mathcal{F} \quad (4.3)$$

und unterteilen den Beweis in drei kleine Schritte:

- (a) Sei $\epsilon > 0$ fest. Gemäß unserer Konstruktion ist $\mu \ll P$, und nach dem ϵ - δ -Kriterium für absolute Stetigkeit, Lemma 4.6, gibt es ein $\delta > 0$, so dass aus $P(F) < \delta$ folgt, dass $\mu(F) < \epsilon$.
- (b) Wegen der Monotonie der bedingten Erwartung gilt für alle $i \in I$ die Abschätzung $\mathbf{E}[Y | \mathcal{F}_i] \leq \mathbf{E}[|Y| | \mathcal{F}_i]$. Mit Hilfe der Tschebyscheff-Ungleichung für das erste Moment von $|Y|$ (Markov-Ungleichung) folgt nun, dass

$$\begin{aligned} \sup_{i \in I} P(|X_i| \geq c) &\leq \sup_{i \in I} \frac{\mathbf{E}[|X_i|]}{c} \\ &= \sup_{i \in I} \frac{\mathbf{E}[|\mathbf{E}[Y | \mathcal{F}_i]|]}{c} \\ &\leq \sup_{i \in I} \frac{\mathbf{E}[\mathbf{E}[|Y| | \mathcal{F}_i]]}{c} \\ &= \frac{\mathbf{E}[|Y|]}{c}, \end{aligned}$$

wobei der letzte Ausdruck für $c \rightarrow \infty$ wegen $\mathbf{E}[|Y|] < \infty$ gegen 0 strebt.

- (c) Sei nun $c > 0$ so, dass $\mathbf{E}[|Y|] < c\delta$. Dann gilt

$$\begin{aligned} \sup_{i \in I} \mathbf{E}[|X_i| \chi_{\{|X_i| \geq c\}}] &\leq \sup_{i \in I} \mathbf{E}[\mathbf{E}[|Y| | \mathcal{F}_i] \chi_{\{|X_i| \geq c\}}] \\ &= \sup_{i \in I} \mathbf{E}[\mathbf{E}[|Y| \chi_{\{|X_i| \geq c\}} | \mathcal{F}_i]] \\ &= \sup_{i \in I} \mathbf{E}[|Y| \chi_{\{|X_i| \geq c\}}], \end{aligned}$$

³⁹ Der Abschnitt folgt im Wesentlichen [10, Kap. 11.2].

wobei wir im zweiten Schritt die \mathcal{F}_i -Messbarkeit von $\chi_{\{|X_i| \geq c\}}$ ausgenutzt haben. Nach (4.3) ist der letzte Ausdruck gleich $\sup_{i \in I} \mu(|X_i| \geq c)$ und die Stetigkeit aus (a) impliziert, dass

$$\sup_{i \in I} \mathbf{E}[|X_i| \chi_{\{|X_i| \geq c\}}] \leq \sup_{i \in I} \mu(|X_i| \geq c) < \epsilon.$$

Da $\epsilon > 0$ beliebig ist, folgt $\sup_{i \in I} \mathbf{E}[|X_i| \chi_{\{|X_i| \geq c\}}] \rightarrow 0$, womit die Aussage bewiesen ist. \square

Beispiel 4.28 (Sukzessive Prognose, Forts.) Sei X integrierbar und $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}_3 \dots$ eine Filtration. Dann definiert

$$X_i = \mathbf{E}[X | \mathcal{F}_i], \quad i = 1 \geq 0$$

ein gleichgradig integrierbares Martingal (vgl. Beispiel 4.17).

Hinreichende Bedingungen

Um gleichgradige Integrierbarkeit einer Folge von Zufallsvariablen nachzuweisen, ist der folgende Satz nützlich.

Satz 4.29 (Hinreichende Bedingungen für gleichgradige Integrierbarkeit) *Eine Familie $(X_i)_{i \in I}$ von integrierbaren Zufallsvariablen ist gleichgradig integrierbar, wenn eine der beiden folgenden Bedingungen erfüllt ist:*

(a) *Es gibt eine integrierbare Majorante für $(X_i)_{i \in I}$, d.h. eine Zufallsvariable $Y \geq 0$ mit $\mathbf{E}[Y] < \infty$ und*

$$|X_i| \leq Y, \quad i \in I.$$

(b) *Für eine messbare Funktion $\varphi: \mathbf{R} \rightarrow \mathbf{R}_+$ mit der Eigenschaft $\lim_{x \rightarrow \infty} (x^{-1} \varphi(x)) = \infty$ gilt, dass*

$$\sup_{i \in I} \mathbf{E}[\varphi(|X_i|)] \leq C < \infty.$$

Beweis. Sei $\epsilon > 0$ beliebig fest.

(a) Nach Voraussetzung gilt, dass $|X_i(\omega)| \leq Y(\omega)$ punktweise und damit $\{|X_i| \geq c^*\} \subset \{Y \geq c^*\}$ für alle $c^* \geq 0$. Da Y integrierbar ist, kann c^* so groß gewählt werden, dass

$$\sup_{i \in I} \mathbf{E}[|X_i| \chi_{\{|X_i| \geq c\}}] \geq \sup_{i \in I} \mathbf{E}[Y \chi_{\{|X_i| \geq c\}}] < \epsilon \quad \forall c \geq c^*.$$

(b) Nach Voraussetzung gibt es ein $x^* > 0$, so dass

$$\frac{\varphi(x)}{x} > \frac{C}{\epsilon} \quad \forall x \geq x^*.$$

Damit erhalten wir für alle $c \geq x^*$ die Abschätzung

$$\begin{aligned} \sup_{i \in I} \mathbf{E}[|X_i| \chi_{\{|X_i| \geq c\}}] &\leq \epsilon \sup_{i \in I} \frac{\mathbf{E}[\varphi(|X_i|) \chi_{\{|X_i| \geq c\}}]}{C} \\ &\leq \epsilon \sup_{i \in I} \frac{\mathbf{E}[\varphi(|X_i|)]}{C} \\ &< \epsilon. \end{aligned}$$

\square

Der letzte Satz macht deutlich, dass es einen Unterschied zwischen *gleichmäßiger* und *gleichgradiger* Integrierbarkeit gibt: Nach Satz 4.29b reicht es für die gleichgradige Integrierbarkeit von $(X_i)_{i \in I}$, dass die X_i gleichmäßig in $L^{1+\delta}(\Omega, P)$ beschränkt sind – mit anderen Worten, dass es ein $\delta > 0$ gibt, so dass gilt:

$$\mathbf{E}[|X_i|^{1+\delta}] \leq C < \infty \quad \forall i \in I.$$

Ist I endlich, fallen gleichmäßige und gleichgradige Integrierbarkeit zusammen, denn $Y = |X_1| + \dots + |X_{|I|}$ ist eine integrierbare Majorante (vgl. Anmerkung 4.26). Für eine unendliche Familie von Zufallsvariablen reicht die gleichmäßige Integrierbarkeit jedoch nicht aus, wie das folgende Beispiel zeigt.

Beispiel 4.30 Sei $(X_n)_{n \in \mathbf{N}}$ eine Folge von Zufallsvariablen

$$X_n: [0, 1] \rightarrow \mathbf{R}, \quad X_n = n\chi_{[0, 1/n]}.$$

auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P) = ([0, 1], \mathcal{B}([0, 1]), \lambda|_{[0, 1]})$. Wegen

$$\mathbf{E}[|X_n|] = \int_{[0, 1]} |X_n| d\lambda = n\lambda([0, 1/n]) = 1,$$

unabhängig von $n \in \mathbf{N}$, ist die Folge $(X_n)_{n \in \mathbf{N}}$ gleichmäßig in L^1 beschränkt. Sie ist jedoch nicht gleichgradig integrierbar, denn für ein beliebiges $c > 0$ gilt

$$\mathbf{E}[|X_n| \chi_{\{|X_n| \geq c\}}] = \mathbf{E}[|X_n|] = 1 \quad \forall n > c.$$

Wir kommen nun zur zentralen Aussage dieses Abschnitts, der Verallgemeinerung des Satzes von der majorisierten Konvergenz, bei dem die Majorisierbarkeit einer Folge durch das schwächere Kriterium der gleichgradigen Integrierbarkeit ersetzt wird (vgl. Satz 1.40).

Satz 4.31 Für eine Zufallsvariable $X: \Omega \rightarrow \mathbf{R}$ und eine Folge $(X_n)_{n \geq 0}$ integrierbarer Zufallsvariablen auf (Ω, \mathcal{F}, P) sind folgende Aussagen äquivalent:

- (i) $(X_n)_{n \geq 0}$ ist gleichgradig integrierbar und $X_n \xrightarrow{P} X$.
- (ii) $X_n \rightarrow X$ in $L^1(\Omega, P)$, d.h., $\mathbf{E}[|X_n - X|] \rightarrow 0$ und $\mathbf{E}[|X|] < \infty$.

Beweis. Wir beweisen hier nur die Implikation “(i) \Rightarrow (ii)”; der Beweis der Umkehrung sei den Lesern als einfache Anwendung der Markov-Ungleichung und Satz 4.29 empfohlen.

“ \Rightarrow ”: Nach dem Lemma von Fatou, Satz, 1.34, und der gleichgradigen Integrierbarkeit von X_n gilt

$$\mathbf{E}[|X|] \liminf_{n \rightarrow \infty} \mathbf{E}[|X_n|] < \infty,$$

woraus sofort die Integrabilität von X folgt. Um $\mathbf{E}[|X_n - X|] \rightarrow 0$ zu zeigen, betrachten wir die gleichgradig integrierbare Folge $\{X_n: n \geq 0\} \cup \{X\}$ und wenden das ϵ - δ -Kriterium für absolute Stetigkeit von Maßen an (vgl. Lemma 4.27). Sei also $\epsilon > 0$. Dann existiert ein $\delta > 0$, so dass

$$P\left(|X_n - X| > \frac{\epsilon}{2}\right) < \delta \quad \Rightarrow \quad \mathbf{E}\left[(|X_n| + |X|)\chi_{\{|X_n - X| > \epsilon/2\}}\right] < \frac{\epsilon}{2},$$

wobei das Antecedens der Implikation wegen $X_n \xrightarrow{P} X$ immer erfüllt werden kann. Mit der Dreiecksungleichung erhalten wir somit, dass für alle für alle n , die größer als ein geeignetes $n^* = n^*(\epsilon, \delta)$ sind, gilt:

$$\begin{aligned} \mathbf{E}[|X_n - X|] &\leq \frac{\epsilon}{2} + \mathbf{E}[|X_n - X| \chi_{\{|X_n - X| > \epsilon/2\}}] \\ &\leq \frac{\epsilon}{2} + \mathbf{E}\left[(|X_n| + |X|)\chi_{\{|X_n - X| > \epsilon/2\}}\right] \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2}, \end{aligned}$$

wobei $\epsilon > 0$ wie immer beliebig klein gewählt werden kann.

“ \Leftarrow ”: Sei $\epsilon > 0$. Wegen

$$P(|X_n - X| > \epsilon) \leq \frac{\mathbf{E}[|X_n - X|]}{\epsilon}$$

impliziert L^1 -Konvergenz die Konvergenz in Wahrscheinlichkeit. Um die gleichgradige Integrierbarkeit reicht es zu zeigen, dass es ein $\delta > 0$ gibt, so dass gilt:

$$P(F) < \delta \quad \Rightarrow \quad \sup_{n \in \mathbf{N}_0} \mathbf{E}[|X_n| \chi_F] < \epsilon, \quad F \in \mathcal{F}.$$

Nun ist X als integrierbare Zufallsvariable auch gleichgradig integrierbar, und angewandt auf X können wir δ aus der letzten Gleichung so klein wählen, dass insbesondere

$$P(F) < \delta \quad \Rightarrow \quad \mathbf{E}[|X|\chi_F] < \frac{\epsilon}{2}, \quad F \in \mathcal{F}.$$

Sei nun $n^* \in \mathbf{N}$ so, dass $\mathbf{E}[|X_n - X|] < \epsilon/2$ für alle $n \geq n^*$. Somit existiert stets ein $\delta > 0$, so dass für alle $F \in \mathcal{F}$ mit der Eigenschaft $P(F) < \delta$ die Abschätzung

$$\begin{aligned} \mathbf{E}[|X_n|\chi_F] &\leq \mathbf{E}[|X|\chi_F] + \mathbf{E}[|X_n - X|\chi_F] \\ &\leq \frac{\epsilon}{2} + \mathbf{E}[|X_n - X|] \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \end{aligned}$$

gilt. Damit ist der Satz bewiesen. \square

Anmerkung 4.32 Neben einer Verallgemeinerung des Satzes von der majorisierten Konvergenz liefert die Aussage von Satz 4.31 eine hinreichende und notwendige Bedingung für Konvergenz in L^1 .

4.4 Stoppzeiten und Stoppsätze

Für Markovketten haben wir Stoppzeiten als an den Prozess adaptierte (im Sinne von “nicht vorausschauende”) Zufallsvariable kennengelernt. Für Spielsysteme – um bei der Analogie mit den Glücksspielen zu bleiben – definiert eine Stoppzeit einen in irgendeinem Sinne richtigen Zeitpunkt aus einem Spiel auszusteigen, also z.B. wenn ein bestimmter Gewinn erreicht ist. Die Messbarkeit verlangt wie schon bei den Markovketten, dass das über den Zeitpunkt des Aussteigens nur auf Grundlage der bis zur aktuellen Spielrunde vorliegenden Information entschieden werden kann. Wir erinnern an die Definition einer Stoppzeit:

Definition 4.33 (Stoppzeit) Eine Funktion $\tau: \Omega \rightarrow \mathbf{N}_0 \cup \{+\infty\}$ heißt *Stoppzeit*, wenn $\{\tau \leq n\}$ für alle $n \in \mathbf{N}_0$ bezüglich der \mathcal{F}_n messbar ist, wobei $\{\tau \leq n\}$ das Ereignis bezeichnet, dass bis zum Zeitpunkt n gestoppt wird.

Anmerkung 4.34 Äquivalent zu $\{\tau \leq n\} \in \mathcal{F}_n$ ist die Aussage, dass $\{\tau = n\}$ bezüglich \mathcal{F}_n messbar ist, denn

$$\{\tau = n\} = \{\tau \leq n\} \setminus \{\tau \leq n-1\}$$

bzw.

$$\{\tau \leq n\} = \bigcup_{k=0}^n \{\tau = k\}.$$

Als Beispiel betrachten wir ein Spielsystem und $(Y_n)_{n \geq 0} \subset \mathbf{N}_0$ mit $Y_n = (H \circ X)_n$ als dem ganzzahligen Gewinn nach n Spielrunden (z.B. in Euro). Ferner sei $A = \{0, \dots, 10\} \subset \mathbf{N}_0$. Die Zufallsvariable

$$\tau_A = \inf\{n \geq 0: Y_n \notin A\}$$

wird *Erstaustrittszeit* genannt. Mit anderen Worten: Wir steigen aus dem Spiel aus, wenn wir entweder pleite sind, $Y_n = 0$, oder wir genügend Geld beisammen haben, $Y_n \geq 10$. Dass τ_A tatsächlich eine Stoppzeit ist, sieht man indem man das Ereignis $\{\tau_A \leq n\}$ durch

$$\{\tau_A \leq n\} = \bigcup_{i=1}^n \{\omega \in \Omega: Y_i(\omega) \in A\}$$

darstellt; da $(Y_n)_{n \geq 0}$ adaptiert ist, gilt $\{\tau_A \leq n\} \in \mathcal{F}_n$, folglich ist τ_A eine Stoppzeit.

Definition 4.35 (Gestoppter Prozess) Es sei $(X_n)_{n \geq 0}$ ein adaptierter Prozess und τ eine Stoppzeit und $\tau \wedge n := \min\{\tau, n\}$. Als gestoppten Prozess bezeichnen wir (siehe Abb. 13)

$$X_n^\tau := X_{\tau \wedge n} = \begin{cases} X_n, & n < \tau \\ X_\tau, & n \geq \tau. \end{cases}$$

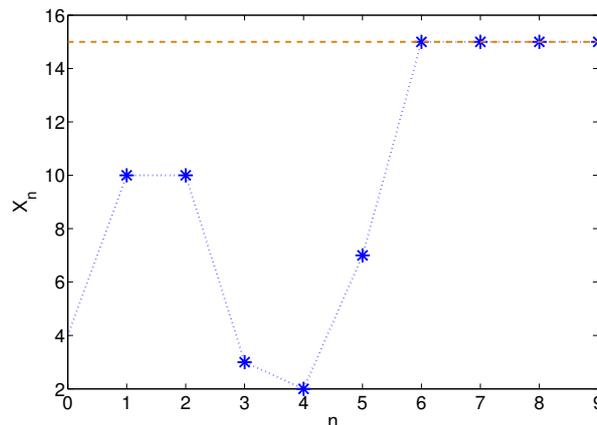


Abb. 13 Gestoppter Prozess $X_n^\tau = X_{\tau \wedge n}$ mit Stoppzeit $\tau = \inf\{n \geq 0: X_n \notin (0, 10)\}$

Anmerkung 4.36 Wie üblich setzen wir $\inf \emptyset = \infty$ und $X_n^\infty = X_n$.

Der folgende, berühmte Satz von Doob besagt, dass es bei einem fairen Spiel (Martingal) keine Abbruchstrategie gibt, mit der man seinen Gesamtgewinn fast sicher verbessern kann.⁴⁰

Satz 4.37 (*Optional Stopping Theorem*, auch: *Optional Sampling Theorem*) *Es sei $(X_n)_{n \geq 0}$ ein adaptierter Prozess und τ eine Stoppzeit. Dann gilt:*

1. *Ist $(X_n)_{n \geq 0}$ ein (Sub-, Super-)Martingal, so ist auch der gestoppte Prozess ein (Sub-, Super-)Martingal.*
2. *Ist $(X_n)_{n \geq 0}$ ein Martingal, so gilt*

$$\mathbf{E}[X_{\tau \wedge n}] = \mathbf{E}[X_0] \quad \forall n \in \mathbf{N}_0.$$

Ist $\tau < \infty$ (f.s.) oder $(X_n^\tau)_{n \geq 0}$ gleichgradig integrierbar, so gilt insbesondere

$$\mathbf{E}[X_\tau] = \mathbf{E}[X_0].$$

3. *Die Aussagen aus 2. gelten entsprechen für Sub- und Supermartingale mit “ \geq ” bzw. “ \leq ”.*

Beweis. Wir beweisen die ersten beiden Aussagen, die Aussage über Sub- und Supermartingale folgt analog.

1. Wir definieren $H = (H_n)_{n \geq 1}$ durch

$$H_n := \chi_{\{\tau \geq n\}} = 1 - \chi_{\{\tau < n\}}.$$

Da τ eine Stoppzeit ist, ist H prävisibel und, da H_n für alle n beschränkt ist, ein Spielsystem. Da die H_n nicht-negativ sind, ist nach Satz 4.22 auch $(H \circ X)$ ein (Sub-, Super-)Martingal und es gilt:

$$\begin{aligned} (H \circ X)_n &= X_0 + \sum_{k=1}^n H_k (X_k - X_{k-1}) \\ &= X_0 + \sum_{k=1}^n \chi_{\{\tau \geq k\}} (X_k - X_{k-1}) \\ &= X_{\tau \wedge n}. \end{aligned}$$

Damit ist auch X_n^τ ein (Sub-, Super-)Martingal.

⁴⁰ Doob (1910–2004), US-amerik. Mathematiker

2. Ist τ fast sicher beschränkt, so folgt aus

$$\tau_n = \tau \wedge n \xrightarrow{\text{f.s.}} \tau, \quad n \rightarrow \infty,$$

die fast sichere Konvergenz

$$\lim_{n \rightarrow \infty} X_{\tau \wedge n} = X_\tau.$$

Sei nun $\tau \leq N$ für ein $N \in \mathbf{N}$. Dann hat die Familie $(X_{\tau \wedge n})_{n \geq 0}$ eine gleichgradig integrierbare Majorante $\max_{0 \leq k \leq N} |X_k|$ und aus dem verallgemeinerten Satz von der majorisierten Konvergenz, Satz 4.31, folgt, dass

$$\lim_{n \rightarrow \infty} \mathbf{E}[|X_{\tau \wedge n} - X_\tau|] = 0.$$

Da $X_n^\tau = X_{\tau \wedge n}$ ein Martingal ist, gilt damit insbesondere, dass

$$\mathbf{E}[X_0] = \mathbf{E}[X_{\tau \wedge n}] \rightarrow \mathbf{E}[X_\tau], \quad n \rightarrow \infty.$$

3. ÜA.

□

Anmerkung 4.38 Im Beweis des ersten Teils des *Optional Stopping Theorem* haben wir ausgenutzt, dass sich jedes Martingal durch ein stochastisches Integral mit einem geeigneten Integranden darstellen lässt. Diese Aussage ist unter dem Namen *Martingal-Darstellungssatz* bekannt.

Beispiel 4.39 (Roulette) Wir betrachten einen asymmetrischen Zufallsspaziergang

$$S_n = \sum_{k=1}^n Y_k, \quad S_0 = 0,$$

wobei die $Y_k = \pm 1$ IID Zufallsvariable mit $p = P(X_k = 1)$ und $1 - p = P(X_k = -1)$ für $p \neq 1/2$ sind.

Wir betrachten das Spiel eines Spielers gegen die Bank, das durch S_n beschrieben werde. Dazu sei $a \in \mathbf{N}$ das Startkapital des Spielers und $b \in \mathbf{N}$ das der Bank; ist $S_n \leq -a$, so ist das Kapital des Spielers aufgebraucht und er geht pleite, bei $S_n \geq b$ geht die Bank bankrott. Wir definieren die Stoppzeit

$$\tau = \inf\{n \geq 0: S_n \notin \{-a, \dots, b\}, a, b \in \mathbf{N}\}$$

und fragen nach der Ruinwahrscheinlichkeit des Spielers, $P(S_\tau = -a)$. Dazu definieren wir eine Zufallsvariable

$$M_n = \left(\frac{1-p}{p}\right)^{S_n}, \quad M_0 = 1 \text{ (f.s.)}$$

die für alle $0 < p < 1$ ein gleichgradig integrierbares Martingal definiert (vgl. [10, Beispiel 11.8]). Nun ist

$$\mathbf{E}[M_\tau] = \left(\frac{1-p}{p}\right)^{-a} P(S_\tau = -a) + \left(\frac{1-p}{p}\right)^b P(S_\tau = b).$$

und nach dem Doobschen Stoppsatz gilt $\mathbf{E}[M_\tau] = \mathbf{E}[M_0] = 1$. Durch Auflösen der Gleichung nach der Ruinwahrscheinlichkeit $P(S_\tau = -a) = 1 - P(S_\tau = b)$ erhalten wir schließlich

$$P(S_\tau = -a) = \frac{1 - r^b}{r^{-a} - r^b},$$

wobei wir $r = (1-p)/p$ gesetzt haben. Im Falles des (unfairen) Roulettespiels ist $p = 18/37$, woraus folgt, dass für gegebenes $b < \infty$ die Ruinwahrscheinlichkeit des Spielers gegen einen festen Wert konvergiert:

$$\lim_{a \rightarrow \infty} P(S_\tau = -a) = 1 - r^{-b}.$$

Für $b = 100$ ist bereits $1 - r^{-b} \geq 0.99$. Abbildung 14 zeigt die Ruinwahrscheinlichkeit eines Roulettespielers für $b = 100$ – was für Casino gelinde gesagt untertrieben sein dürfte – als Funktion des Startkapitals des Spielers. Selbst für $a = b$ tritt der Ruin des Spielers noch mit einer Wahrscheinlichkeit von etwa 0.9955 ein, was im wesentlichen schon dem Grenzwert für $a \rightarrow \infty$ entspricht.

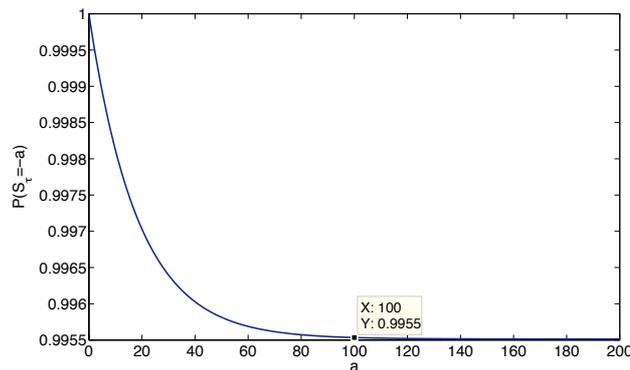


Abb. 14 Ruinwahrscheinlichkeit eines Roulettespielers als Funktion seines Startkapitals bei festem Kapital der Bank.

4.5 Optimales Stoppen

Als eine wichtige Anwendung des Doob'schen Stoppsätzen wollen wir einen kurzen Einblick in die Theorie des optimalen Stoppens geben. Während der Stoppsatz eine Aussage darüber ist, dass man auch mit einer cleveren Stoppstrategie aus einem fairen Spiel (Martingal) kein unfaires machen kann, werden wir nun der Frage nachgehen, wie man eine optimale Stoppstrategie für stochastische Prozesse findet, mit der bestimmte Zielfunktionale maximiert werden.⁴¹ Ein bekanntes Beispiel, bei dem das optimale Stoppen eine Rolle spielt, ist die Bewertung von Amerikanischen Aktienoptionen, mit denen man das Recht erwirbt, eine Aktie zu einem späteren Zeitpunkt zu einem fest Preis kaufen oder Verkaufen zu können. Für den Anbieter der Aktienoption stellt sich die daher Frage, wie viel Geld dem Kunden für ein solches Recht abzuknöpfen ist. Der Wert einer Option, bei der man zum Beispiel das Recht zum Aktienkauf erwirbt ("Call-Option"), ist durch

$$W(x) = \max_{\tau \leq T} \mathbf{E}_x[\max\{S_\tau - K, 0\}]$$

gegeben, wobei $T > 0$ die Laufzeit der Option ist, S_τ den Wert der Aktie zum Zeitpunkt des Ausübens der Option bezeichnet, x ihren Wert zur Zeit $t = 0$ und $K > 0$ den vorher vereinbarten Ausübungspreis. Das Maximum wird dabei über alle Stoppzeiten $\tau \leq T$ von $(S_t)_{t \leq T}$ gebildet.

Wir wollen nun den allgemeinen Rahmen für Stoppzeitenproblem formulieren und beschränken uns der Einfachheit halber auf Markovketten mit abzählbarem Zustandsraum, auch wenn die meisten der Aussagen unten auch für allgemeinere stochastische Prozesse gelten. Gegeben seien also

1. ein Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) ,
2. eine adaptierter stochastischer Prozess $\{X_n, \mathcal{F}_n : 0 \leq n \leq N\}$ auf S mit Übergangsmatrix $P = (P_{xy})_{x,y \in S}$,
3. eine Gewinnfunktion $f: \{0, \dots, N\} \times S \rightarrow \mathbf{R}$ mit der Eigenschaft $\mathbf{E}[f(n, X_n)] < \infty$ für alle $n \in \{0, \dots, N\}$.

Gesucht ist eine Stoppzeit $\tau^* \in \{0, 1, \dots, N\}$ bezüglich \mathcal{F}_n mit der Eigenschaft

$$\tau^* = \operatorname{argmax} \mathbf{E}[f(\tau, X_\tau)]$$

Dynamische Programmierung

Die Berechnung der optimalen Stoppstrategie basiert auf der iterativen Berechnung der Wertefunktion:

Definition 4.40 Es sei $\mathcal{U}_k = \{\tau : k \leq \tau \leq N \text{ und } \{\tau = n\} \in \mathcal{F}_n\}$ eine Menge von Stoppzeiten. Die Funktion

$$V(k, x) = \max_{\tau \in \mathcal{U}_k} \mathbf{E}[f(\tau, X_\tau) | X_k = x] \tag{4.4}$$

heißt *Wertefunktion* (engl: *optimal cost-to-go*).

⁴¹ In einem gewissen Sinne macht die optimale Stoppstrategie ein unfaires Spiel fair, indem sie aus einem " \leq " ein " $=$ " macht.

Das Prinzip der Werteiteration ist unter dem Namen *dynamische Programmierung* bekannt und geht auf Richard Bellman zurück.⁴² Zunächst macht man sich klar, dass nach Definition der Wertefunktion gilt:

$$V(N, x) = f(N, x) \quad (4.5)$$

Bei Start der Markovkette in $X_N = x$ ist die einzig erlaubte – und damit automatisch optimale – Strategie: Stoppen. Bei Start in $X_{N-1} = x$ ergeben sich entsprechend zwei Möglichkeiten: Stoppe oder mache einen Schritt; im ersten Fall ist wieder $V(N-1, x) = f(N-1, x)$, im zweiten Fall ist $V(N-1, x) = \mathbf{E}[V(N, X_N) | X_{N-1} = x]$. Wiederholtes Anwenden des letzten Arguments ergibt, dass V durch die Iteration

$$V(k-1, x) = \max\{f(k-1, x), \mathbf{E}[V(k, X_k) | X_{k-1} = x]\} \quad (4.6)$$

gegeben ist und dass der optimale Stoppzeitpunkt gerade durch

$$\tau^* = \inf\{k \geq 0: V(k, x) = f(k, x)\} \quad (4.7)$$

festgelegt ist. Gleichung (4.6) heißt *Bellman-Gleichung*. Der folgende Satz zeigt, dass die durch (4.5)–(4.7) definierten Stoppstrategien tatsächlich optimal sind.

Satz 4.41 Für alle Stoppzeiten $\tau \in \{0, 1, \dots, N\}$ gilt

$$V(0, x) \geq \mathbf{E}[f(\tau, X_\tau) | X_0 = x]$$

mit Gleichheit,

$$V(0, x) = \mathbf{E}[f(\tau^*, X_{\tau^*}) | X_0 = x],$$

für

$$\tau^* = \inf\{k \geq 0: V(k, x) = f(k, x)\}.$$

Beweis. Sei $Z_k = V(k, X_k)$, $0 \leq k \leq N$. Wir zeigen, dass Z_k ein Supermartingal ist, dass unter der optimalen Stoppstrategie (4.7) zu einem Martingal wird. Nach der Bellman-Gleichung (4.6) gilt

$$\begin{aligned} V(k, x) &\geq \sum_{y \in S} V(k+1, y) p_{xy} \\ &= \mathbf{E}[V(k+1, X_{k+1}) | X_k = x], \end{aligned}$$

Folglich ist

$$V(k, X_k) \geq \mathbf{E}[V(k+1, X_{k+1}) | \mathcal{F}_k].$$

bzw.

$$Z_k \geq \mathbf{E}[Z_{k+1} | \mathcal{F}_k].$$

was zeigt, dass Z_k ein Supermartingal ist. Nach dem Doob'schen Stoppsatz gilt nun $\mathbf{E}[Z_\tau] \leq \mathbf{E}[Z_0]$; insbesondere erhalten wir also $\mathbf{E}_x[Z_\tau] \leq \mathbf{E}_x[Z_0]$ mit $\mathbf{E}_x[\cdot] = \mathbf{E}[\cdot | X_0 = x]$, womit der erste Teil der Aussage bewiesen ist, denn

$$\mathbf{E}_x[f(\tau, X_\tau)] \leq \mathbf{E}_x[V(\tau, X_\tau)] \leq \mathbf{E}_x[V(0, X_0)] = V(0, x).$$

Um den zweiten Teil zu beweisen, betrachten wir den Fall

$$V(k, x) > f(k, x) \quad \forall x \in S.$$

Dann folgt aus (4.6), dass

$$\begin{aligned} V(k, x) &= \sum_{y \in S} V(k+1, y) p_{xy} \\ &= \mathbf{E}[V(k+1, X_{k+1}) | X_k = x], \end{aligned}$$

und somit ist $M_k = V(\tau^* \wedge k, X_{\tau^* \wedge k})$ nach dem Doob'schen Stoppsatz ein Martingal, woraus sich mit (4.7) und

$$\mathbf{E}_x[f(\tau^*, X_{\tau^*})] = \mathbf{E}_x[V(\tau^*, X_{\tau^*})] = \mathbf{E}_x[V(0, X_0)] = V(0, x).$$

die Behauptung ergibt. □

⁴² Richard Ernest Bellman (1920–1984), US-amerik. Mathematiker.

Anwendung: Das Sekretärinnenproblem

Das Sekretärinnenproblem – auch *Heiratsproblem* oder *Game of Googol* – ist ein klassisches Beispiel aus der Entscheidungstheorie und wahrscheinlich die bekannteste Anwendung der Theorie des optimalen Stoppens. Das Sekretärinnenproblem besteht darin, aus einer Reihe von nacheinander vorsprechenden, unterschiedlich gut qualifizierten Bewerbern auf eine Sekretariatsstelle den besten auszuwählen. Dabei ist

- a) kein Bewerber genauso gut wie ein anderer qualifiziert,
- b) nur die Anzahl der Bewerber bekannt, nicht aber ihre Qualifikation,
- c) die Ablehnung eines Bewerbers nicht revidierbar.

Die Reihenfolge der vorsprechenden Bewerber ist dabei beliebig (unabhängig und gleichverteilt). Die Aufgabe ist also, auf Grundlage der Informationen, die man durch die bereits befragten Bewerber gesammelt hat, eine optimale Entscheidung zu treffen, wobei “Information” so hier etwas wie die σ -Algebra des Vorstellungsprozesses darstellt und “Entscheidung” bedeutet, dass der Prozess gestoppt wird. Eine optimale Entscheidung kann dabei wegen der unvollständigen Informationen, die zur Verfügung stehen, und der Nichtrevidierbarkeit der Entscheidung nicht die Wahl des besten Bewerbers sein, sondern nur eine, die die Wahrscheinlichkeit maximiert, den besten zu wählen. Wenn die Bewerber absteigend nach Qualifikation geordnet werden können, gilt es, folgendes Stoppszeitenproblem zu lösen: Finde $n \in \{1, \dots, N\}$, so dass

$$J_N(n) = P(n\text{-ter Bewerber ist die Nummer 1}) \quad (4.8)$$

maximiert wird. Wie wollen nun einige mögliche Strategien im Hinblick auf ihre Erfolgswahrscheinlichkeit bei der optimalen Wahl eines Bewerbers diskutieren:

1. Wähle immer den ersten oder den letzten Bewerber aus. Da die Bewerber unabhängig gleichverteilt aus dem Pool der Kandidaten ausgewählt werden und es keine zwei Bewerber mit der gleichen Qualifikation gibt, ist das sicherlich eine suboptimale Strategie, denn die Erfolgswahrscheinlichkeit beträgt nur

$$P(1\text{-ter Bewerber ist die Nummer 1}) = P(\text{letzter Bewerber ist die Nummer 1}) = \frac{1}{N},$$

und für $N \rightarrow \infty$ geht diese Wahrscheinlichkeit gegen 0.

2. Da wir N kennen, können wir zunächst die Hälfte aller Kandidaten vorsprechen lassen, und dann den ersten wählen, der besser ist als diese Bewerber. In diesem Fall ist die Erfolgswahrscheinlichkeit gerade durch

$$P(\text{zweitbesten Kandidat wurde interviewt \& bester Kandidat kommt noch}) \approx \frac{1}{4}$$

gegeben, woraus wir schließen können, dass

$$\liminf_{N \rightarrow \infty} \max_{n \leq N} J_N(n) \geq \frac{1}{4}.$$

Doch auch diese Strategie ist suboptimal, denn es geht noch besser.

3. Es ist nicht einzusehen, warum man gerade zunächst die Hälfte der Bewerber ziehen lassen soll, um dann den nächstbesten zu nehmen. Deshalb betrachten wir nun den Fall, dass wir von N Bewerbern die ersten $k - 1$ interviewen und dann den ersten zu nehmen, der besser qualifiziert als die ersten $k - 1$ ist. Bezeichnet

nun p_N die Wahrscheinlichkeit, den besten der N Bewerber ausgewählt zu haben, so ist

$$\begin{aligned}
 p_N &= \sum_{i=1}^N P(i \text{ gewählt} \ \& \ i \text{ ist Nr. 1}) \\
 &= \sum_{i=1}^N P(i \text{ gewählt} \mid i \text{ ist Nr. 1})P(i \text{ ist Nr. 1}) \\
 &= \frac{1}{N} \sum_{i=1}^N P(i \text{ gewählt} \mid i \text{ ist Nr. 1}) \\
 &= \frac{1}{N} \sum_{i=k}^N P(i \text{ gewählt} \mid i \text{ ist Nr. 1}) \\
 &= \frac{1}{N} \sum_{i=k}^N P(\text{Nr. 2 ist unter den ersten } k-1 \text{ Kandidaten} \mid i \text{ ist Nummer 1}) \\
 &= \frac{1}{N} \sum_{i=k}^N \frac{k-1}{i-1}.
 \end{aligned}$$

Um herauszufinden, für welches k die Erfolgswahrscheinlichkeit p_N maximiert wird, wobei uns insbesondere der Fall $N \gg 1$ interessiert, betrachten wir nun den Grenzwert $k, N \rightarrow \infty$ mit $k/N \rightarrow r$. Die Frage ist also, welcher Prozentsatz von Bewerbern idealerweise vorsprechen muss, bevor eine Auswahl getroffen wird. Eine grobe Approximation von p_N für große N finden wir durch

$$\begin{aligned}
 p_N &= \frac{k-1}{N} \sum_{i=k}^N \frac{1}{i-1} \\
 &\approx r \sum_{i=\lceil rN \rceil}^N \frac{N}{N(i-1)} \\
 &\approx r \int_r^1 \frac{ds}{s}
 \end{aligned}$$

Der letzte Ausdruck liefert $p_N \approx -r \log r$, und das wird maximal für $x^* = 1/e \approx 1/3$; das ist die berühmte Drittel-Regel, die besagt, dass man etwa zunächst ein Drittel aller Bewerber befragen soll, bevor man den nächsten Bewerber nimmt, dessen Qualifikation über der des ersten Drittels liegt (vgl. [6]).

Dynamische Programmierung, Forts.

Wir werden nun das Sekretärinnenproblem mit Hilfe der Bellman-Gleichung (4.6) lösen, um zu zeigen, dass die Drittel-Regel tatsächlich für große N annähernd optimal ist. Dazu betrachten wir eine Markovkette $(X_n)_{n \geq 1}$ mit 2 Zuständen 0 und 1, wobei $X_n = 1$ bedeutet, dass der n -te Bewerber besser qualifiziert ist als die ersten $n-1$ Bewerber und $X_n = 0$ bedeutet, dass der n -te Bewerber nicht besser qualifiziert ist. Die zeitabhängigen Übergangswahrscheinlichkeiten sind für $1 \leq n \leq N-1$ entsprechend durch

$$P(X_{n+1} = 1 \mid X_n) = P(X_{n+1} = 1) = \frac{1}{n+1}$$

und

$$P(X_{n+1} = 0 \mid X_n) = 1 - P(X_{n+1} = 1) = \frac{n}{n+1}$$

gegeben. Ist $X_n = 1$, so ist die Wahrscheinlichkeit, die beste Wahl getroffen zu haben gerade n/N . Damit haben wir unser Sekretärinnenproblem auf ein Stoppzeitenproblem für eine Folge unabhängiger Bernoulli-verteilter Zufallsvariablen X_1, \dots, X_N mit $X_n \in \{0, 1\}$ mit

$$P(X_n = 0) = \frac{n}{n+1}, \quad P(X_n = 1) = \frac{1}{n+1}$$

und der Gewinnfunktion

$$f(k, x) = \begin{cases} 0 & x = 0 \\ k/N & x = 1 \end{cases}$$

zurückgeführt. Damit ist $J(n) = \mathbf{E}[f(n, X_n)]$ gerade die Wahrscheinlichkeit mit dem n -ten Bewerber die beste Wahl zu treffen; vgl. (4.8). Die Bellman-Gleichung für $V(k, x) = \mathbf{E}[f(n, X_n) | X_k = x]$ ist

$$\begin{aligned} V(k, 0) &= \frac{1}{k+1}V(k+1, 1) + \frac{k}{k+1}V(k+1, 0) \\ V(k, 1) &= \max\{f(k+1, 1), V(k, 0)\} \end{aligned} \quad (4.9)$$

wobei wir in der zweiten Gleichung den Erwartungswert von $V(k+1, X_{k+1})$ durch $V(k, 0)$ ersetzt haben. Nun ist es klar, dass die Suche weitergeht, so lange $X_k = 0$ ist; ist dagegen $X_k = 1$, so sollte die Suche weitergehen, bis $k/N \geq V(k, 0)$. Nun ist $V(k, 1) \geq V(k, 0)$ für alle $k \in \{1, \dots, N\}$ und folglich $V(k, 0) \geq V(k+1, 0)$, d.h., $V(\cdot, 0)$ ist monoton fallend. Damit gibt es eine optimale Stoppzeit $n^* \in \{1, \dots, N\}$, so dass

$$\frac{k}{N} \geq V(k, 0) \quad \forall k \geq n^* \quad \text{und} \quad \frac{k}{N} < V(k, 0) \quad \forall k < n^* .$$

Nach (4.7) besteht also die optimale Strategie tatsächlich darin, die ersten $n^* - 1$ Bewerber ziehen zu lassen und dann den nächsten Bewerber zu nehmen, der besser qualifiziert als die ersten $n^* - 1$ Kandidaten ist. Nun lässt sich direkt aus der Bellman-Gleichung (4.9) ablesen, dass

$$V(k, 0) = \frac{1}{k+1} \frac{k+1}{N} + \frac{k}{k+1} V(k+1, 0)$$

bzw.

$$\frac{V(k, 0)}{k} - \frac{V(k+1, 0)}{k+1} = \frac{1}{kN} .$$

Zusammen mit der Randbedingung $V(N, 0) = 0$ ergibt sich dann

$$V(k, 0) = \frac{k}{N} \sum_{i=k}^{N-1} \frac{1}{i} \quad (4.10)$$

und somit ist

$$n^* = \inf \left\{ k : \frac{1}{N} \sum_{i=k}^{N-1} \frac{1}{i} \geq \frac{1}{N} \right\} \sim \frac{N}{e} \quad (N \rightarrow \infty) . \quad (4.11)$$

Das ist gerade die zuvor hergeleitete Drittel-Regel, die folglich nach Satz 4.41 für $N \rightarrow \infty$ optimal ist.

Literatur

- [1] L. Arnold. *Stochastische Differentialgleichungen: Theorie und Anwendung*. Wiley & Sons, 1973.
- [2] P. Billingsley. *Convergence of Probability Measures*. Wiley & Sons, 1999.
- [3] P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer, 2010.
- [4] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Comput. Netw. ISDN Syst.*, **30**, 107–117, 1998.
- [5] K.L. Chung. *Markov Chains with Stationary Transition Probabilities*. Springer, 1960.
- [6] E.B. Dynkin. The optimal choice of the instant for stopping a Markov process. *Doklady Akad., Nauk.* **4**, 627–629, 1963.
- [7] J. Elstrodt. *Maß- und Integrationstheorie*. Springer, 2005.
- [8] S. Gerschgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk. UdSSR Otd. Fiz.-Mat. Nauk.* **7**(4), 749–754, 1931.
- [9] A. Klenke. *Wahrscheinlichkeitstheorie*. Springer, 2006.
- [10] D. Meintrup, S. Schäffler. *Stochastik*. Springer, 2005.
- [11] E.L. Melnick, A. Tenenbein. Misspecifications of the Normal Distribution. *Amer. Statist.* **36**(4), 372–373, 1982.
- [12] N. Metropolis. The beginning of the Monte Carlo method. *Los Alamos Science* **15**(584), 125–130, 1987.
- [13] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, London, 1993.
- [14] B. Øksendal. *Stochastic Differential Equations: An Introduction with Applications*. Springer, 2003.
- [15] H. Thorisson. *Coupling, Stationarity and Regeneration*. Springer, 2000.
- [16] S.R.S. Varadhan. *Probability Theory*. AMS, 2001.
- [17] D. Werner. *Funktionalanalysis*. Springer, 2005.