

Stochastische Prozesse

Andreas Eberle

9. August 2014

Inhaltsverzeichnis

Inhaltsverzeichnis	2
I Bedingen und stochastische Prozesse	6
1 Bedingte Erwartungen	7
1.1 Bedingen auf diskrete Zufallsvariablen	7
1.1.1 Bedingte Erwartungen als Zufallsvariablen	7
1.1.2 Formel von der totalen Wahrscheinlichkeit	9
1.1.3 Bedingte Varianz	10
1.1.4 Anwendung auf zufällige Summen	11
1.1.5 Charakterisierende Eigenschaften der bedingten Erwartung	12
1.2 Erzeugende Funktionen, Verzweigungsprozesse, und Erneuerungen	13
1.2.1 Erzeugende Funktionen von ganzzahligen Zufallsvariablen	13
1.2.2 Erzeugende Funktionen zufälliger Summen	14
1.2.3 Galton-Watson-Verzweigungsprozesse	15
1.2.4 Rekurrente Ereignisse und Erneuerungsgleichung	18
1.3 Bedingen auf allgemeine Zufallsvariablen	21
1.3.1 Das Faktorisierungslemma	22
1.3.2 Definition allgemeiner bedingter Erwartungen	23
1.3.3 Diskreter und absolutstetiger Fall	26
1.3.4 Reguläre bedingte Verteilungen	28
1.4 Rechnen mit bedingten Erwartungen; Poissonprozess	31
1.4.1 Eigenschaften der bedingten Erwartung	32
1.4.2 Poissonprozesse	35
1.4.3 Poissonscher Punktprozess	39
1.5 Bedingte Erwartung als beste L^2 -Approximation	42

1.5.1	Jensensche Ungleichung	43
1.5.2	Bedingte Erwartung als beste L^2 -Prognose	44
1.5.3	Existenz der bedingten Erwartung	46
2	Markovketten und Differenzgleichungen	48
2.1	Grundlagen	48
2.1.1	Zufällige dynamische Systeme als Markovketten	49
2.1.2	Beispiele	51
2.1.3	Endlichdimensionale Randverteilungen	54
2.2	Gleichgewichte	56
2.2.1	Definitionen und Beispiele	56
2.2.2	Eigenwerte und Konvergenz ins Gleichgewicht	60
2.2.3	Kopplungen und Kontraktivität	64
2.3	Markov-Eigenschaften	68
2.3.1	Kanonisches Modell einer Markovkette	68
2.3.2	Markov-Eigenschaft	70
2.3.3	Starke Markov-Eigenschaft	73
2.4	Zusammenhang mit Differenzgleichungen	75
2.4.1	Differenzgleichungen für Markovketten	78
2.4.2	Dirichletproblem und Austrittsverteilung	82
2.4.3	Beispiele harmonischer Funktionen	84
2.4.4	Mittlere Aufenthaltszeiten und Greenfunktion	87
3	Martingale	89
3.1	Definitions and examples	89
3.1.1	Martingales and supermartingales	90
3.1.2	Some fundamental examples	91
3.2	Doob Decomposition and Martingale Problem	97
3.2.1	Doob Decomposition	97
3.2.2	Conditional Variance Process	98
3.2.3	Martingale problem	101
3.3	Gambling strategies and stopping times	103
3.3.1	Martingale transforms	103
3.3.2	Stopped Martingales	107
3.3.3	Optional Stopping Theorems	111

3.3.4	Wald's identity for random sums	113
3.4	Almost sure convergence of supermartingales	114
3.4.1	Doob's upcrossing inequality	115
3.4.2	Proof of Doob's Convergence Theorem	117
3.4.3	Examples and first applications	117
3.4.4	Generalized Borel-Cantelli Lemma	121
4	Stochastische Prozesse in stetiger Zeit	123
4.1	From Random Walks to Brownian Motion	123
4.1.1	Central Limit Theorem	125
4.1.2	Brownian motion as a Lévy process.	129
4.1.3	Brownian motion as a Markov process.	129
4.1.4	Wiener Measure	131
4.2	Gaussian Processes	133
4.2.1	Multivariate normals	133
4.2.2	Gaussian processes	136
4.2.3	Brownian Motion as a Gaussian Process	137
4.3	Markovprozesse in stetiger Zeit	142
4.3.1	Übergangskerne und Markovprozesse	142
4.3.2	Zeitstetige Markovketten	145
4.4	Vorwärts- und Rückwärtsgleichungen	149
4.4.1	Vorwärts- und Rückwärtsgleichungen für Markovketten	149
4.4.2	Gleichgewichte zeitstetiger Markovketten	151
4.4.3	Vorwärts- und Rückwärtsgleichung für die Brownsche Bewegung	153
II	Pfadeigenschaften von stochastischen Prozessen	156
5	Pfadeigenschaften der Brownschen Bewegung	157
5.1	The Wiener-Lévy Construction	157
5.1.1	A first attempt	158
5.1.2	The Wiener-Lévy representation of Brownian motion	159
5.1.3	Lévy's construction of Brownian motion	165
5.2	Regularity of Brownian paths	169
5.2.1	Typical Brownian sample paths are nowhere differentiable	169
5.2.2	Hölder continuity	171

5.2.3	Law of the iterated logarithm	172
5.2.4	Level sets of Brownian paths	173
5.3	Strong Markov property and reflection principle	175
5.3.1	Running maximum of Brownian motion	176
5.3.2	Strong Markov property for Brownian motion	178
5.3.3	A rigorous reflection principle	181
6	Langzeitasymptotik von Markovketten	183
6.1	Rekurrenz und Transienz von Markovketten	183
6.1.1	Rekurrenz und Transienz von d -dimensionalen Random Walks	184
6.1.2	Rekurrenz und Transienz von einzelnen Zuständen	186
6.1.3	Kommunikationsklassen und globale Rekurrenz	189
6.2	Stationäre stochastische Prozesse	193
6.2.1	Stationarität und Reversibilität	193
6.2.2	Rekurrenz von stationären Prozessen	195
6.2.3	Anwendung auf Markovketten	197
6.3	Ergodizität	198
6.3.1	Positive Rekurrenz und Gleichgewichte	199
6.3.2	Ein Gesetz der großen Zahlen für Markovketten	200
6.3.3	Allgemeinere Ergodensätze	204

Teil I

Bedingen und stochastische Prozesse

Kapitel 1

Bedingte Erwartungen

Zur Analyse von stochastischen Modellen mit Abhängigkeiten verwendet man in der Regel bedingte Wahrscheinlichkeiten und Erwartungswerte gegeben die Werte von Zufallsvariablen. Beispielsweise beschreibt man einen stochastischen Prozess $(X_n)_{n \in \mathbb{Z}_+}$ durch die bedingten Verteilungen des nächsten Zustands X_{n+1} gegeben den Verlauf $X_{0:n} = (X_0, X_1, \dots, X_n)$ bis zur Zeit n .

1.1 Bedingen auf diskrete Zufallsvariablen

Wir betrachten zunächst das Bedingen auf den Ausgang einer diskreten Zufallsvariable $Y : \Omega \rightarrow S$, S abzählbar. In diesem Fall können wir die *bedingte Wahrscheinlichkeitsverteilung*

$$P[A | Y = z] = \frac{P[A \cap \{Y = z\}]}{P[Y = z]}, \quad A \in \mathcal{A},$$

und die *bedingten Erwartungswerte*

$$E[X | Y = z] = \frac{E[X; Y = z]}{P[Y = z]}, \quad X \in \mathcal{L}^1(\Omega, \mathcal{A}, P),$$

für alle $z \in S$ mit $P[Y = z] > 0$ auf elementare Weise wie in Abschnitt ?? definieren. Für $z \in S$ mit $P[Y = z] = 0$ sind die bedingten Wahrscheinlichkeiten nicht definiert.

1.1.1 Bedingte Erwartungen als Zufallsvariablen

Es wird sich als praktisch erweisen, die bedingten Wahrscheinlichkeiten und Erwartungswerte nicht als Funktion des Ausgangs z , sondern als Funktion der Zufallsvariable Y zu interpretieren. Die bedingten Wahrscheinlichkeiten und Erwartungswerte sind dann selbst Zufallsvariablen:

Definition (Bedingte Erwartung gegeben eine diskrete Zufallsvariable). Sei $X : \Omega \rightarrow \mathbb{R}$ eine Zufallsvariable mit $E[X^-] < \infty$, und $Y : \Omega \rightarrow S$ eine diskrete Zufallsvariable. Die durch

$$E[X | Y] := g(Y) = \sum_{z \in S} g(z) \cdot I_{\{Y=z\}}$$

mit

$$g(z) := \begin{cases} E[X | Y = z] & \text{falls } P[Y = z] > 0 \\ \text{beliebig} & \text{falls } P[Y = z] = 0 \end{cases}$$

P -fast sicher eindeutig definierte Zufallsvariable $E[X | Y]$ heißt (**Version der**) **bedingte(n) Erwartung von X gegeben Y** . Für ein Ereignis $A \in \mathcal{A}$ heißt die Zufallsvariable

$$P[A | Y] := E[I_A | Y]$$

(**Version der**) **bedingte(n) Wahrscheinlichkeit von A gegeben Y** .

Die bedingte Erwartung $E[X | Y]$ und die bedingte Wahrscheinlichkeit $P[A | Y]$ sind also Zufallsvariablen mit den Werten $E[X | Y = z]$ bzw. $P[A | Y = z]$ auf den Mengen $\{Y = z\}$, $z \in S$ mit $P[Y = z] > 0$. Auf jeder der Nullmengen $\{Y = z\}$, $z \in S$ mit $P[Y = z] = 0$, wird der bedingten Erwartung ein willkürlicher konstanter Wert zugewiesen, d.h. die Definition ist nur P -fast überall eindeutig. Wir fassen zunächst einige elementare Eigenschaften der so definierten bedingten Erwartung zusammen:

Lemma 1.1 (Eigenschaften der bedingten Erwartung).

- (1). Die Abbildung $X \mapsto E[X | Y]$ ist P -fast sicher linear und monoton.
- (2). Sind X und Y unabhängig, dann gilt $E[X | Y] = E[X]$ P -fast sicher.
- (3). Herausziehen, was bekannt ist:

Für alle $f : S \rightarrow \mathbb{R}$ mit $f(Y) \cdot X \geq 0$ bzw. $f(Y) \cdot X \in \mathcal{L}^1$ gilt

$$E[f(Y) \cdot X | Y] = f(Y) \cdot E[X | Y] \quad P\text{-fast sicher.}$$

Insbesondere gilt

$$E[f(Y) | Y] = f(Y) \quad P\text{-fast sicher.}$$

Beweis. (2). Sind X und Y unabhängig, dann gilt

$$E[X | Y = z] = \frac{E[X \cdot I_{\{Y=z\}}]}{P[Y = z]} = E[X]$$

für alle $z \in S$ mit $P[Y = z] > 0$, also $E[X | Y] = E[X]$ P -fast sicher. Die ebenso elementaren Beweise von (1) und (3) werden dem Leser als Übung überlassen. □

Anschaulich können wir die zweite Aussage folgendermaßen interpretieren: Sind X und Y unabhängig, dann liefert die Kenntnis des Wertes $Y(\omega)$ keine zusätzlichen Informationen über $X(\omega)$. Daher ist die beste L^2 -Prognose für $X(\omega)$ wie im unbedingten Fall durch den Erwartungswert $E[X]$ gegeben.

1.1.2 Formel von der totalen Wahrscheinlichkeit

Die aus Satz ?? bekannte Formel von der totalen Wahrscheinlichkeit können wir mithilfe der obigen Definition in kompakter Weise schreiben.

Satz 1.2 (Formel von der totalen Wahrscheinlichkeit). Sei $Y : \Omega \rightarrow S$ eine diskrete Zufallsvariable mit Verteilung $\mu(z) = P[Y = z]$. Für alle messbaren $X : \Omega \rightarrow \mathbb{R}_+$ gilt:

$$E[X] = \sum_{z: \mu(z) \neq 0} E[X | Y = z] \mu(z) = E[E[X | Y]]$$

Insbesondere gilt

$$P[A] = E[P[A | Y]] \quad \text{für alle } A \in \mathcal{A}.$$

Beweis. Wegen $\Omega = \bigcup_{z \in S} \{Y = z\}$ gilt nach dem Transformationssatz

$$\begin{aligned} E[X] &= \sum_{z \in S} E[X; Y = z] = \sum_{z: \mu(z) \neq 0} E[X; Y = z] \\ &= \sum_{z: \mu(z) \neq 0} E[X | Y = z] \cdot \mu(z) = \sum_{z: \mu(z) \neq 0} g(z) \cdot \mu(z) \\ &= E[g(Y)], \end{aligned}$$

wobei $g : S \rightarrow \mathbb{R}$ eine beliebige Funktion mit $g(z) = E[X | Y = z]$ für alle $z \in S$ mit $\mu(z) \neq 0$ ist. Die Aussage folgt wegen $g(Y) = E[X | Y]$ P -fast sicher. \square

Bemerkung. Für $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ folgt aus der Monotonie der bedingten Erwartung

$$|E[X | Y]| \leq E[|X| | Y]$$

und damit die Ungleichung

$$E[|E[X | Y]|] \leq E[E[|X| | Y]] = E[|X|].$$

Die Abbildung $X \mapsto E[X | Y]$ ist also eine Kontraktion auf $L^1(\Omega, \mathcal{A}, P)$. Die Aussage von Satz 1.2 gilt entsprechend auch für $X \in \mathcal{L}^1$.

1.1.3 Bedingte Varianz

Sei nun $X : \Omega \rightarrow \overline{\mathbb{R}}$ eine bzgl. P integrierbare Zufallsvariable

Definition (Bedingte Varianz). Die Zufallsvariable

$$\text{Var}[X | Y] := E[(X - E[X | Y])^2 | Y]$$

heißt *bedingte Varianz* von X gegeben Y .

Ist X quadratintegrierbar, dann gelten die folgenden Aussagen:

Lemma 1.3. Für $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ gilt:

(1). *L^2 -Kontraktivität:* $E[|E[X | Y]|^2] \leq E[X^2]$.

(2). $\text{Var}[X | Y] = E[X^2 | Y] - E[X | Y]^2$ *P -fast sicher.*

Insbesondere folgt für $z \in S$ mit $\mu(z) \neq 0$:

$$\text{Var}[X | Y] = \text{Var}[X | Y = z] \quad \text{auf } \{Y = z\}. \quad (1.1.1)$$

Beweis. (1). folgt aus Satz 1.2, da für alle $z \in S$ mit $P[Y = z] \neq 0$ auf $\{Y = z\}$ gilt:

$$|E[X | Y]|^2 = |E[X | Y = z]|^2 \leq E[X^2 | Y = z] = E[X^2 | Y].$$

(2). Nach Lemma 1.1, (1) und (3), ergibt sich dann ähnlich wie für die unbedingte Varianz:

$$\begin{aligned} \text{Var}[X | Y] &= E[X^2 | Y] - 2 \cdot E[X \cdot E[X | Y] | Y] + E[E[X | Y]^2 | Y] \\ &= E[X^2 | Y] - E[X | Y]^2 \quad P\text{-fast sicher.} \end{aligned}$$

□

Die folgende Zerlegungsformel kann häufig verwendet werden, um Varianzen zu berechnen oder abzuschätzen:

Satz 1.4 (Formel von der bedingten Varianz). Für eine Zufallsvariable $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ gilt:

$$\begin{aligned} \text{Var}[X] &= E[\text{Var}[X | Y]] + \text{Var}[E[X | Y]] \\ &= \sum_{z:\mu(z) \neq 0} \text{Var}[X | Y = z] \cdot \mu(z) + \sum_{z:\mu(z) \neq 0} (E[X | Y = z] - E[X])^2 \cdot \mu(z). \end{aligned}$$

Beweis. Es gilt

$$\begin{aligned}\text{Var}[X] &= E[X^2] - E[X]^2 = E[E[X^2 | Y]] - E[E[X | Y]]^2 \\ &= E[E[X^2 | Y]] - E[E[X | Y]^2] + E[E[X | Y]^2] - E[E[X | Y]]^2 \\ &= E[\text{Var}[X | Y]] + \text{Var}[E[X | Y]].\end{aligned}$$

Der zweite Teil der Behauptung folgt nun aus (1.1.1) und der entsprechenden Eigenschaft für die bedingte Erwartung. \square

1.1.4 Anwendung auf zufällige Summen

Als erste Anwendung betrachten wir eine Summe

$$S_N(\omega) := \sum_{i=1}^{N(\omega)} X_i(\omega)$$

von unabhängigen, identisch verteilten Zufallsvariablen $X_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ mit zufälliger Anzahl N von Summanden. Hierbei sei $N : \Omega \rightarrow \{0, 1, 2, \dots\}$ eine von den X_i unabhängige Zufallsvariable. Seien $m = E[X_1]$ und $\sigma^2 = \text{Var}[X_1]$. Wir berechnen nun die verschiedenen Kenngrößen der Verteilung von S_N .

Berechnung des Erwartungswertes: Da S_k und N unabhängig sind, erhalten wir

$$E[S_N | N = k] = E[S_k | N = k] = E[S_k] = k \cdot m \quad \text{für alle } k \in \mathbb{N},$$

also $E[S_N | N] = N \cdot m$, und damit nach Satz 1.2:

$$E[S_N] = E[E[S_N | N]] = E[N] \cdot m.$$

Berechnung der Varianz: Erneut folgt wegen der Unabhängigkeit von S_k und N :

$$\text{Var}[S_N | N = k] = \text{Var}[S_k | N = k] = \text{Var}[S_k] = k \cdot \sigma^2,$$

also $\text{Var}[S_N | N] = N \cdot \sigma^2$, und damit nach Satz 1.4:

$$\text{Var}[S_N] = E[\text{Var}[S_N | N]] + \text{Var}[E[S_N | N]] = E[N] \cdot \sigma^2 + \text{Var}[N] \cdot m^2.$$

Berechnung der momentenerzeugenden Funktion: Für $t \in \mathbb{R}$ gilt

$$\begin{aligned}M_{S_N}(t) &= E[e^{tS_N}] = E[E[e^{tS_N} | N]] = E\left[\prod_{i=1}^N E[e^{tX_i}]\right] \\ &= E[E[e^{tX_1}]^N] = E[M_{X_1}(t)^N] = M_N(\log M_{X_1}(t)).\end{aligned}$$

Mithilfe von M_{S_N} kann man die Momente der zufälligen Summe S_N berechnen:

$$E[S_N^m] = M_{S_N}^{(m)}(0) \quad \text{für alle } m \in \mathbb{N}.$$

Im Prinzip erhält man die Verteilung von S_N durch Laplace-Inversion, was aber nicht immer praktikabel ist. Nehmen die Zufallsvariablen X_i nur nichtnegative ganzzahlige Werte an, kann man statt der momentenerzeugenden Funktion die erzeugende Funktion verwenden, und daraus die Verteilung berechnen. Wir gehen darauf im folgenden Abschnitt ein.

1.1.5 Charakterisierende Eigenschaften der bedingten Erwartung

Zum Abschluss dieses Abschnitts beweisen wir eine alternative Charakterisierung der bedingten Erwartung gegeben eine diskrete Zufallsvariable $Y : \Omega \rightarrow S$, S abzählbar. Diese Charakterisierung werden wir in Abschnitt 1.3 verwenden, um bedingte Erwartungen für allgemeine Bedingungen zu definieren. Sei $X : \Omega \rightarrow \mathbb{R}_+$ eine nichtnegative (bzw. integrierbare) Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

Satz 1.5. *Eine reellwertige Zufallsvariable $\bar{X} \geq 0$ (bzw. $\bar{X} \in \mathcal{L}^1$) auf (Ω, \mathcal{A}, P) ist genau dann eine Version der bedingten Erwartung $E[X | Y]$, wenn gilt:*

(I) $\bar{X} = g(Y)$ für eine Funktion $g : S \rightarrow \mathbb{R}$, und

(II) $E[\bar{X} \cdot f(Y)] = E[X \cdot f(Y)]$ für alle nichtnegativen bzw. beschränkten Funktionen $f : S \rightarrow \mathbb{R}$.

Beweis. Ist \bar{X} eine Version von $E[X | Y]$, dann gilt (I). Außerdem folgt nach Lemma 1.1 (3) und der Formel von der totalen Wahrscheinlichkeit:

$$E[\bar{X} \cdot f(Y)] = E[E[X | Y] \cdot f(Y)] = E[E[X \cdot f(Y) | Y]] = E[X \cdot f(Y)]$$

für jede nichtnegative bzw. beschränkte Funktion $f : S \rightarrow \mathbb{R}$.

Umgekehrt folgt aus (I), dass $\bar{X} = g(z)$ auf $\{Y = z\}$ gilt. Ist außerdem (II) erfüllt, dann folgt weiter

$$\begin{aligned} g(z) &= E[\bar{X} | Y = z] = \frac{E[\bar{X} \cdot I_{\{z\}}(Y)]}{P[Y = z]} \\ &= \frac{E[X \cdot I_{\{z\}}(Y)]}{P[Y = z]} = E[X | Y = z] \end{aligned}$$

für alle $z \in S$ mit $P[Y = z] > 0$, d.h. $\bar{X} = g(Y)$ ist eine Version der bedingten Erwartung $E[X | Y]$. \square

In einigen Fällen können die charakterisierenden Eigenschaften direkt überprüft werden, um bedingte Erwartungen zu identifizieren:

Beispiel (Summen austauschbarer Zufallsvariablen). Seien $X_1, X_2, \dots, X_n \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ integrierbare Zufallsvariablen, deren gemeinsame Verteilung invariant unter Koordinatenpermutationen ist, d.h. $(X_{\pi(1)}, X_{\pi(2)}, \dots, X_{\pi(n)}) \sim (X_1, X_2, \dots, X_n)$ für alle $\pi \in \mathcal{S}_n$. Zufallsvariablen mit dieser Eigenschaft heißen **austauschbar** – beispielsweise sind unabhängige identisch verteilte Zufallsvariablen austauschbar. Wir zeigen:

$$E[X_i | S_n] = \frac{1}{n} S_n \quad P\text{-fast sicher für alle } i = 1, \dots, n,$$

wobei $S_n = X_1 + \dots + X_n$. Zum Beweis überprüfen wir, dass $\bar{X}_i := \frac{1}{n} S_n$ die Bedingungen (I) und (II) aus Satz 1.5 für $Y = S_n$ erfüllt. (I) ist offensichtlich. Zudem gilt wegen der Austauschbarkeit für jede beschränkte messbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$:

$$E[X_i \cdot f(S_n)] = E[X_j \cdot f(S_n)] \quad \text{für alle } i, j = 1, \dots, n,$$

also

$$E\left[\frac{1}{n} S_n \cdot f(S_n)\right] = \frac{1}{n} \sum_{j=1}^n E[X_j \cdot f(S_n)] = E[X_i \cdot f(S_n)]$$

für alle $i = 1, \dots, n$, d.h. (II) ist auch erfüllt.

1.2 Erzeugende Funktionen, Verzweigungsprozesse, und Erneuerungen

Wir wollen die Methoden aus dem letzten Abschnitt nun verwenden, um Verzweigungs- und Erneuerungsprozesse zu untersuchen. Ein wichtiges Hilfsmittel sind in beiden Fällen erzeugende Funktionen:

1.2.1 Erzeugende Funktionen von ganzzahligen Zufallsvariablen

Sei $X : \Omega \rightarrow \{0, 1, 2, \dots\}$ eine auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definierte Zufallsvariable mit nichtnegativen *ganzzahligen* Werten.

Definition (Erzeugende Funktion einer Zufallsvariable mit Werten in \mathbb{Z}_+). Die durch

$$G(s) = E[s^X] = \sum_{k=0}^{\infty} P[X = k] s^k, \quad s \in [-1, 1],$$

definierte Funktion heißt **erzeugende Funktion** der Zufallsvariable X bzw. der Folge $\mu(k) = P[X = k]$ der Gewichte von X .

Durch Vergleich mit der geometrischen Reihe sieht man, dass der Konvergenzradius der Potenzreihe stets größer oder gleich 1 ist. Also ist die erzeugende Funktion analytisch auf $(-1, 1)$, und es gilt

$$P[X = k] = \frac{G^{(k)}(0)}{k!} \quad \text{für alle } k = 0, 1, 2, \dots$$

Kennen wir also die erzeugende Funktion explizit, dann können wir die Gewichte der Verteilung berechnen.

Durch zweimaliges Ableiten zeigt man zudem, dass G monoton und konvex auf $[0, 1]$ ist. Für $s \in (0, 1]$ gilt nach Definition $G(s) = M(\log s)$. Daher lassen sich aus der erzeugenden Funktion die Momente von X berechnen – beispielsweise gilt $E[X] = G'(1-)$ (linksseitige Ableitung von $G(s)$ bei $s = 1$), falls der Erwartungswert endlich ist.

Für die erzeugende Funktion einer Summe $X + Y$ von unabhängigen, nichtnegativen, ganzzahligen Zufallsvariablen X und Y gilt offensichtlich

$$G_{X+Y}(s) = G_X(s) \cdot G_Y(s) \quad \text{für alle } s \in [-1, 1].$$

Somit ist die erzeugende Funktion der Faltung

$$(\mu * \nu)(k) = \sum_{i=0}^k \mu(i)\nu(k-i) \quad (k = 0, 1, 2, \dots)$$

zweier Wahrscheinlichkeitsverteilungen μ und ν auf $\mathbb{N} \cup \{0\}$ das Produkt der einzelnen erzeugenden Funktionen.

Erzeugende Funktionen können in verschiedenen Situationen für explizite Berechnungen verwendet werden. Wir demonstrieren dies hier in einigen grundlegenden Beispielen. Viele weitere entsprechende Anwendungen finden sich in den Wahrscheinlichkeitstheorie-Lehrbüchern von Feller und Grimmett/Stirzacker.

1.2.2 Erzeugende Funktionen zufälliger Summen

Sind $N, X_1, X_2, \dots : \Omega \rightarrow \{0, 1, 2, \dots\}$ unabhängige Zufallsvariablen, dann erhalten wir für die Summe $S_N = \sum_{i=1}^N X_i$:

$$G_{S_N}(s) = E[s^{S_N}] = E[E[s^{S_N} | N]] = E[G(s)^N] = G_N(G(s)), \quad (1.2.1)$$

wobei G die erzeugende Funktion der Summanden X_i ist. Für die Verteilung von S_N ergibt sich

$$P[S_N = k] = \frac{1}{k!} (G_N \circ G)^{(k)}(0) \quad \text{für alle } k \geq 0.$$

Beispiel (Ausdünnungseigenschaft von Poissonverteilungen). Ein Huhn lege eine mit Parameter $\lambda > 0$ Poissonverteilte Anzahl N von Eiern, von denen aus jedem unabhängig voneinander und von N mit Wahrscheinlichkeit p ein Küken schlüpfe. Die erzeugende Funktion der Poissonverteilung ist

$$G_N(s) = E[s^N] = \sum_{k=0}^{\infty} s^k \cdot \frac{\lambda^k}{k!} e^{-\lambda} = e^{\lambda(s-1)}.$$

Die Anzahl der geschlüpften Küken ist $S_N = \sum_{i=1}^N X_i$, wobei die X_i untereinander und von N unabhängige, Bernoulli(p)-verteilte Zufallsvariablen sind. Wir erhalten also

$$G_{S_N}(s) = G_N(G_{X_1}(s)) = G_N(1 - p + p \cdot s) = e^{p\lambda \cdot (s-1)},$$

d.h. die Zahl der geschlüpften Küken ist wieder Poissonverteilt mit Parameter $p \cdot \lambda$. Eine ausgedünnte Poissonverteilung ist also wieder eine Poissonverteilung!

1.2.3 Galton-Watson-Verzweigungsprozesse

Wir betrachten das folgende Modell für ein zufälliges Populationswachstum: Alle Individuen der Population erzeugen unabhängig voneinander eine zufällige Anzahl von Nachkommen in der nächsten Generation mit Verteilung ν . Hierbei sei ν eine feste Wahrscheinlichkeitsverteilung auf $\{0, 1, 2, \dots\}$ mit $\nu[\{2, 3, \dots\}] \neq 0$. Dieses Modell wurde 1889 von Galton und Watson eingeführt, um die Aussterbewahrscheinlichkeit englischer Adelstitel zu untersuchen. Ähnlich wie beim Random Walk handelt es sich um ein fundamentales stochastisches Modell mit unzähligen Erweiterungen und Anwendungen, z.B. auf das Wachstum von Zellpopulationen, die Ausbreitung von Epidemien, die Zunahme der Neutronenzahl in einem Reaktor, oder auch die näherungsweise Berechnung von genetischen Abständen oder der Anzahl von Zuständen in einem großen zufälligen Graphen (z.B. dem Internet), die man in einer bestimmten Anzahl von Schritten erreichen kann. Die Nachkommensstruktur eines einzelnen Individuums bestimmt einen zufälligen verwurzelten Baum, s. Grafik 1.1. Dementsprechend spielen Verzweigungsprozesse auch eine zentrale Rolle bei der Analyse diverser stochastischer Modelle auf Bäumen, s. z.B. [Peres: Probability on trees].

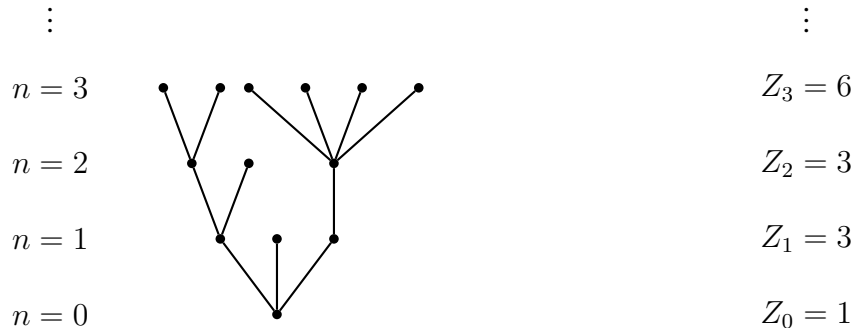


Abbildung 1.1: Beispiel für eine Realisierung eines Galton-Watson-Prozesses.

Wir beschreiben die Nachkommenszahlen der einzelnen Individuen in der $(n-1)$ -ten Generation eines Verzweigungsprozesses durch unabhängige Zufallsvariablen

$$N_i^n : \Omega \rightarrow \{0, 1, 2, \dots\}, \quad i, n = 1, 2, \dots,$$

mit Verteilung ν . Für die Gesamtzahl der Individuen in der n -ten Generation erhalten wir die folgende rekursive Darstellung:

$$Z_n = \sum_{i=1}^{Z_{n-1}} N_i^n \quad \text{für alle } n \geq 1.$$

Ohne wesentliche Einschränkungen nehmen wir $Z_0 = 1$ an. Enthält die Anfangspopulation stattdessen mehrere Individuen, dann erzeugen diese voneinander unabhängige, identisch verteilte Unterpopulationen. Da Z_{n-1} nur von den Zufallsvariablen N_i^k für $k \leq n-1$ abhängt, sind Z_{n-1} und N_i^n ($i \in \mathbb{N}$) unabhängige Zufallsvariablen. Durch Bedingen auf Z_{n-1} erhalten wir für die mittleren Populationsgrößen die Rekursion

$$E[Z_n] = E[Z_{n-1}] \cdot m,$$

wobei $m := \sum_{i=1}^{\infty} i \cdot \nu(i)$ die mittlere Nachkommenszahl eines Individuums ist. Wir unterscheiden die folgenden Fälle:

- $m > 1$: Exponentielles Wachstum der Erwartungswerte (superkritisch)
- $m = 1$: Erwartungswerte konstant (kritisch)
- $m < 1$: Exponentieller Abfall der Erwartungswerte (subkritisch)

Wir wollen nun untersuchen, mit welcher Wahrscheinlichkeit die Population in den einzelnen Fällen ausstirbt. Nach (1.2.1) gilt für die erzeugenden Funktionen der Populationsgrößen die Rekursionsformel

$$G_{Z_n}(s) = E \left[s^{\sum_{i=1}^{Z_{n-1}} N_i^n} \right] = G_{Z_{n-1}}(G(s)),$$

wobei G die erzeugende Funktion der Verteilung ν der Anzahl N_i^n der Kinder eines Individuums ist. Per Induktion folgt wegen $G_{Z_1}(s) = G(s)$:

$$G_{Z_n}(s) = \underbrace{G(G(\dots G(s)\dots))}_{n\text{-mal}} = G^n(s) \quad \text{für alle } n \in \mathbb{N} \text{ und } s \in [0, 1].$$

Für die Wahrscheinlichkeiten π_n , dass der Prozess zur Zeit n ausgestorben ist, erhalten wir die Rekursionsformel

$$\pi_n = P[Z_n = 0] = G_{Z_n}(0) = G^n(0) = G(\pi_{n-1}). \quad (1.2.2)$$

Sei nun π die Wahrscheinlichkeit, dass die Population in endlicher Zeit ausstirbt. Da die Ereignisse $\{Z_n = 0\}$ monoton wachsend sind, gilt

$$\pi = P\left[\bigcup_n \{Z_n = 0\}\right] = \lim_{n \rightarrow \infty} \pi_n.$$

Da G auf $[0, 1]$ stetig ist, folgt aus (1.2.2)

$$\pi = G(\pi),$$

d.h. die Aussterbewahrscheinlichkeit π ist ein Fixpunkt der erzeugenden Funktion. Wie oben bemerkt, ist die erzeugende Funktion $G : [0, 1] \rightarrow [0, 1]$ strikt konvex mit $G(1) = 1$ und $G'(1-) = E[N_i^n] = m$. Hieraus folgt, dass 1 im Fall $m \leq 1$ der einzige Fixpunkt von G in $[0, 1]$ ist, während im superkritischen Fall $m > 1$ ein weiterer Fixpunkt $\pi^* \in [0, 1)$ existiert, siehe auch Grafik 1.2. Aus den Skizzen erkennt man zudem, dass die Aussterbewahrscheinlichkeit $\pi = \lim \pi_n$ der kleinste Fixpunkt von G auf $[0, 1]$ ist. Also stirbt der Prozess im subkritischen bzw. kritischen Fall mit Wahrscheinlichkeit 1 aus, während er im superkritischen Fall mit einer strikt positiven Wahrscheinlichkeit überlebt.

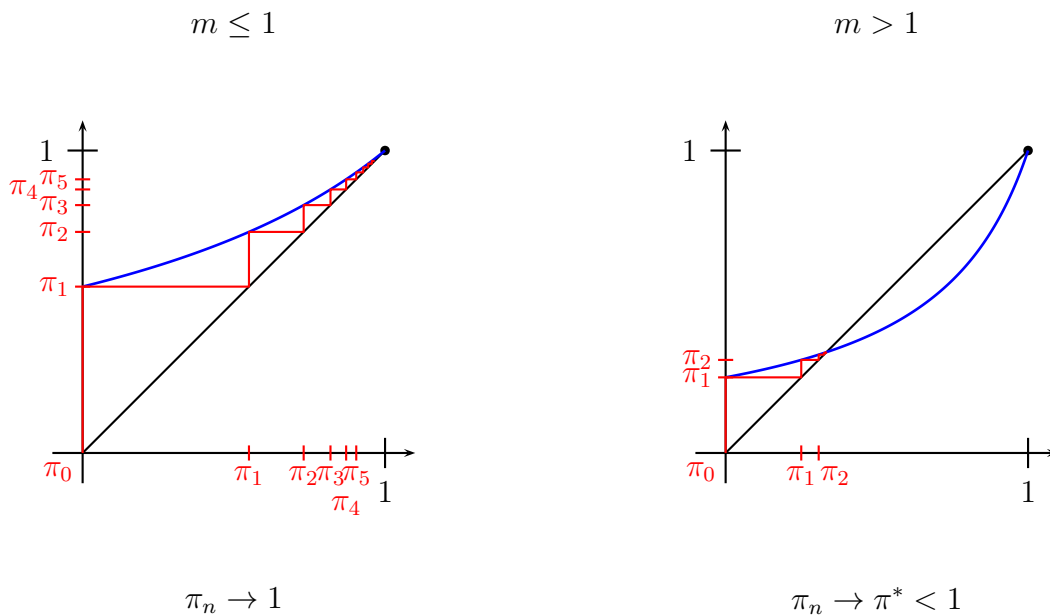


Abbildung 1.2: Erzeugendenfunktionen von Galton-Watson-Prozessen mit unterschiedlichen Verteilungen für die Anzahl der Nachkommen. In Rot: Fixpunktiteration

Beispiel (Geometrische Nachkommensverteilung). Ist die Verteilung

$$\nu(k) = p^k(1-p) \quad (k = 0, 1, 2, \dots)$$

der Anzahl der Nachkommen eine geometrische Verteilung mit Parameter $p \in (0, 1)$, dann ergibt sich

$$G(s) = \sum_{k=0}^{\infty} s^k p^k (1-p) = \frac{1-p}{1-ps} \quad \text{für alle } s \in [0, 1].$$

Fixpunkte dieser Funktion sind 1 und $\frac{1-p}{p}$. Für $1-p \geq p$ (subkritischer Fall) ist 1 der einzige Fixpunkt in $[0, 1]$, also stirbt die Population P -fast sicher aus. Im superkritischen Fall $1-p < p$ beträgt die Aussterbewahrscheinlichkeit π dagegen nur $\frac{1-p}{p}$.

1.2.4 Rekurrente Ereignisse und Erneuerungsgleichung

Als weitere Anwendung von erzeugenden Funktionen betrachten wir eine Folge von unvorhersehbaren Ereignissen, die zu diskreten Zeitpunkten $n \in \mathbb{N}$ eintreten. Die Ereignisse bezeichnen wir auch als „Erneuerungen“ (engl. renewals), und denken dabei z.B. an den wiederholten Ausfall und Austausch eines Verschleißteils in einer Maschine, oder das wiederholte Abarbeiten einer

Warteschlange. Wir beschreiben den Zeitpunkt, an dem die k -te Erneuerung stattfindet, durch eine Zufallsvariable

$$S_k = T_1 + T_2 + \dots + T_k.$$

T_1 ist also der Zeitpunkt der ersten Erneuerung, und für $k \geq 2$ ist T_k der zeitliche Abstand der $(k-1)$ -ten und der k -ten Erneuerung. In einem einfachen Modell nehmen wir an, dass $T_1, T_2, \dots : \Omega \rightarrow \mathbb{N}$ unabhängige Zufallsvariablen sind, und, dass T_2, T_3, \dots identisch verteilt sind (aber nicht T_1 !). Wir wollen nun die Wahrscheinlichkeiten p_n der Ereignisse

$$A_n = \{\exists k \in \mathbb{N} : S_k = n\} \quad \text{„Erneuerung zur Zeit } n\text{“}$$

aus den Verteilungen der Wartezeiten berechnen. Bedingen auf den Wert von T_1 liefert für $n \geq m$:

$$\begin{aligned} P[A_n | T_1 = m] &= P[\exists k \in \mathbb{N} : T_1 + \dots + T_k = n | T_1 = m] \\ &= P[\exists k \in \mathbb{N} : T_2 + \dots + T_k = n - m | T_1 = m] \\ &= P[\exists k \in \mathbb{N} : T_2 + \dots + T_k = n - m], \end{aligned}$$

und damit

$$P[A_n | T_1 = m] = P[A_{n-m+1} | T_1 = 1] = P[A_{n-m+1} | A_1].$$

Nach der Formel von der totalen Wahrscheinlichkeit erhalten wir für $n \in \mathbb{N}$:

$$p_n = \sum_{m=1}^n q_{n-m} \cdot P[T_1 = m] \tag{1.2.3}$$

mit $q_n := P[A_{n+1} | A_1]$. Um die bedingten Wahrscheinlichkeiten q_n zu berechnen, bedingen wir zusätzlich auf T_2 . Da T_2, T_3, \dots unabhängig und identisch verteilt sind, gilt für $n \geq m$:

$$\begin{aligned} P[A_{n+1} | A_1 \cap \{T_2 = m\}] &= P[\exists k \in \mathbb{N} : T_1 + \dots + T_k = n + 1 | T_1 = 1, T_2 = m] \\ &= P[\exists k \geq 2 : T_3 + \dots + T_k = n - m | T_1 = 1, T_2 = m] \\ &= P[\exists k \geq 2 : T_3 + \dots + T_k = n - m] \\ &= P[\exists k \geq 2 : T_2 + \dots + T_{k-1} = n - m] \\ &= P[A_{n-m+1} | A_1] = q_{n-m}. \end{aligned}$$

Wegen

$$q_n = P[A_{n+1} | A_1] = \sum_{m=1}^n P[A_{n+1} | A_1 \cap \{T_2 = m\}] \cdot P[T_2 = m]$$

erhalten wir

$$q_n = \sum_{m=1}^n q_{n-m} \cdot P[T_2 = m] \quad \text{für alle } n \geq 1. \quad (1.2.4)$$

Die Gleichungen (1.2.3) und (1.2.4) heißen *Erneuerungsgleichungen*. Auf den rechten Seiten dieser Gleichungen stehen (wegen $T_1, T_2 \geq 1$) die Faltungen der Folge $q_n, n \in \mathbb{N}$, mit der Folge der Gewichte der Wartezeiten T_1 bzw. T_2 . Daher ist es zweckmäßig, zu den erzeugenden Funktionen

$$G_p(s) = \sum_{n=1}^{\infty} p_n s^n$$

und

$$G_q(s) = \sum_{n=0}^{\infty} q_n s^n$$

überzugehen. Für $|s| < 1$ erhalten wir aus (1.2.3)

$$G_p(s) = G_q(s) \cdot G_{T_1}(s).$$

Aus (1.2.4) ergibt sich, da die rechte Seite für $n = 0$ verschwindet:

$$G_q(s) - 1 = \sum_{n=1}^{\infty} q_n s^n = G_q(s) \cdot G_{T_2}(s).$$

Es folgt $G_q(s) = (1 - G_{T_2}(s))^{-1}$, und damit

$$G_p(s) = \frac{G_{T_1}(s)}{1 - G_{T_2}(s)}. \quad (1.2.5)$$

(1.2.5) liefert den gesuchten Zusammenhang zwischen der Verteilung der Wartezeiten, und den Wahrscheinlichkeiten p_n , dass zur Zeit n eine Erneuerung stattfindet.

Sei nun die Verteilung der Lebensdauern T_2, T_3, \dots vorgegeben. Dann können wir untersuchen, welche Verteilung die Anfangswartezeit T_1 haben muss, damit die Wahrscheinlichkeiten p_n nicht von n abhängen (*Stationarität*). Für $\alpha \in [0, 1]$ gilt $p_n = \alpha$ für alle $n \in \mathbb{N}$ genau dann, wenn

$$G_p(s) = \sum_{n=1}^{\infty} p_n s^n = \frac{\alpha}{1-s} \quad \text{für alle } s \in (-1, 1),$$

d.h. wenn

$$G_{T_1}(s) = \alpha \cdot \frac{1 - G_{T_2}(s)}{1-s} \quad \text{für alle } s \in (-1, 1). \quad (1.2.6)$$

Da G_{T_1} und G_{T_2} erzeugende Funktionen von Wahrscheinlichkeitsverteilungen sind, muss dann gelten:

$$\begin{aligned} 1 &= G_{T_1}(1) = \lim_{s \uparrow 1} G_{T_1}(s) \\ &= \alpha \lim_{s \uparrow 1} \frac{G_{T_2}(s) - 1}{s - 1} = \alpha G'_{T_2}(1-) \\ &= \alpha \cdot E[T_2]. \end{aligned}$$

Also muss T_2 endlichen Erwartungswert haben, und

$$\alpha = 1/E[T_2] \quad (1.2.7)$$

gelten. Dies ist auch anschaulich plausibel: Im stationären Fall ist die Erneuerungswahrscheinlichkeit zu einem festen Zeitpunkt der Kehrwert des mittleren zeitlichen Abstandes zwischen zwei Erneuerungen. Gilt (1.2.7), dann ergibt sich aus (1.2.6) durch Koeffizientenvergleich:

$$P[T_1 = n] = \alpha \cdot \left(1 - \sum_{k=1}^n P[T_2 = k] \right) = \frac{P[T_2 > n]}{E[T_2]}. \quad (1.2.8)$$

Die Folge p_n der Erneuerungswahrscheinlichkeiten ist also genau dann konstant, wenn die Verteilung von T_1 durch (1.2.8) gegeben ist („stationärer Erneuerungsprozess“). Weiter kann man ausgehend von (1.2.6) zeigen, dass für *beliebige* Verteilungen der ersten Erneuerungszeit die Wahrscheinlichkeiten p_n für $n \rightarrow \infty$ gegen $1/E[T_2]$ konvergieren („asymptotische Stationarität“), falls der Erwartungswert endlich ist und keine *Periodizität* auftritt, d.h.

$$\text{ggT}(\{n | P[T_2 = n] > 0\}) = 1.$$

Den Beweis dieses **Erneuerungssatzes** über erzeugende Funktionen findet man im Klassiker von W.Feller (An Introduction to Probability Theory and its Applications, Vol. 1).

1.3 Bedingen auf allgemeine Zufallsvariablen

Ist Y eine reellwertige Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit stetiger Verteilungsfunktion, dann gilt $P[Y = z] = 0$ für alle $z \in \mathbb{R}$. Bedingte Wahrscheinlichkeiten gegeben $Y = z$ können daher nicht wie für diskrete Zufallsvariablen definiert werden. Alternativ könnte man versuchen, bedingte Wahrscheinlichkeiten gegeben Y als Grenzwert zu definieren:

$$P[A | Y = z] = \lim_{h \searrow 0} P[A | z - h \leq Y \leq z + h]. \quad (1.3.1)$$

Dies ist in bestimmten Fällen möglich, aber im allgemeinen ist die Existenz des Grenzwertes nicht gewährleistet.

Stattdessen definiert man bedingte Erwartungen gegeben allgemeine Zufallsvariablen Y mithilfe der Charakterisierung aus Satz 1.5. Bedingte Wahrscheinlichkeiten gegeben Y erhält man als Spezialfall bedingter Erwartungen:

$$P[A | Y] := E[I_A | Y]. \quad (1.3.2)$$

Bedingte Wahrscheinlichkeiten wie in (1.3.1) sind im Allgemeinen nicht im herkömmlichen Sinn definiert. Es ist allerdings ausgehend von (1.3.1) allgemein möglich, für ein festes Ereignis A die Abbildung $z \mapsto P[A \mid Y = z]$ bis auf Modifikation auf Nullmengen bzgl. der Verteilung von Y zu definieren.

1.3.1 Das Faktorisierungslemma

Wir beweisen zunächst eine wichtige maßtheoretische Aussage. Diese wird es uns unter Anderem ermöglichen, die charakterisierenden Eigenschaften bedingter Erwartungen aus Satz 1.5 noch etwas eleganter zu formulieren:

Satz 1.6 (Faktorisierungslemma). *Sei (S, \mathcal{S}) ein messbarer Raum und $Y : \Omega \rightarrow S$ eine Abbildung. Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ ist genau dann $\sigma(Y)$ -messbar, wenn*

$$X = f(Y) = f \circ Y$$

für eine \mathcal{S} -messbare Funktion $f : S \rightarrow \mathbb{R}$ gilt.

$$\begin{array}{c} X \\ \curvearrowright \\ (\Omega, \sigma(Y)) \xrightarrow{Y} (S, \mathcal{S}) \longrightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R})) \end{array}$$

Beweis. (1). Ist $X = f \circ Y$ für eine messbare Funktion f , dann gilt

$$X^{-1}(B) = Y^{-1}(f^{-1}(B)) \in \sigma(Y) \quad \text{für alle } B \in \mathcal{B}(\mathbb{R}),$$

da $f^{-1}(B) \in \mathcal{S}$. Daher ist X $\sigma(Y)$ -messbar.

(2). Für die umgekehrte Richtung müssen wir zeigen, dass aus der $\sigma(Y)$ -Messbarkeit von X folgt, dass X eine messbare Funktion von Y ist. Dazu gehen wir schrittweise vor („maßtheoretische Induktion“):

(a) Ist $X = I_A$ eine Indikatorfunktion mit $A \in \sigma(Y)$, dann gilt $A = Y^{-1}(B)$ mit $B \in \mathcal{S}$, und damit

$$X(\omega) = I_{Y^{-1}(B)}(\omega) = I_B(Y(\omega)) \quad \text{für alle } \omega \in \Omega.$$

(b) Für $X = \sum_{i=1}^n c_i I_{A_i}$ mit $A_i \in \sigma(Y)$ und $c_i \in \mathbb{R}$ gilt entsprechend

$$X = \sum_{i=1}^n c_i I_{B_i}(Y),$$

wobei B_i Mengen aus S mit $A_i = Y^{-1}(B_i)$ sind.

- (c) Für eine beliebige nichtnegative, $\sigma(Y)$ -messbare Abbildung $X : \Omega \rightarrow \mathbb{R}$ existiert eine Folge (X_n) von $\sigma(Y)$ -messbaren Elementarfunktionen mit $X_n \nearrow X$. Nach (b) gilt $X_n = f_n(Y)$ mit \mathcal{S} -messbaren Funktionen f_n . Damit folgt:

$$X = \sup X_n = \sup f_n(Y) = f(Y),$$

wobei $f = \sup f_n$ wieder \mathcal{S} -messbar ist.

- (d) Für eine allgemeine $\sigma(Y)$ -messbare Abbildung $X : \Omega \rightarrow \mathbb{R}$ sind sowohl X^+ als auch X^- messbare Funktionen von Y , also auch X selbst.

□

Mithilfe des Faktorisierungslemmas können wir die *charakterisierenden Eigenschaften* (I) und (II) bedingter Erwartungen gegeben eine diskrete Zufallsvariable Y aus Satz 1.5 wie folgt umformulieren:

\bar{X} ist genau dann eine Version von $E[X | Y]$, wenn gilt:

- (i) \bar{X} ist $\sigma(Y)$ -messbar,
- (ii) $E[\bar{X}; A] = E[X; A]$ für alle $A \in \sigma(Y)$.

Die Äquivalenz von (I) und (i) folgt aus dem Faktorisierungslemma, und die Äquivalenz von (II) und (ii) ergibt sich durch maßtheoretische Induktion, denn (ii) besagt gerade, dass

$$E[\bar{X} \cdot I_B(Y)] = E[X \cdot I_B(Y)] \quad \text{für alle } B \in \mathcal{S} \text{ gilt.}$$

1.3.2 Definition allgemeiner bedingter Erwartungen

Eine bemerkenswerte Konsequenz der Charakterisierung bedingter Erwartungen durch die Bedingungen (i) und (ii) ist, dass die *bedingte Erwartung* $E[X | Y]$ von der Zufallsvariablen Y nur über die von Y erzeugte σ -Algebra $\sigma(Y)$ abhängt! Sind zwei Zufallsvariablen Y und Z Funktionen voneinander, dann ist $\sigma(Y) = \sigma(Z)$, und damit stimmen auch die bedingten Erwartungen $E[X | Y]$ und $E[X | Z]$ überein (mit Wahrscheinlichkeit 1). Daher liegt es nahe, gleich von der bedingten Erwartung gegeben eine σ -Algebra zu sprechen. Die σ -Algebra (z.B. $\sigma(Y)$ oder $\sigma(Y_1, \dots, Y_n)$) beschreibt dann die zur Verfügung stehende „Information“, auf die bedingt wird.

Die Charakterisierung bedingter Erwartungen durch (i) und (ii) können wir sofort auf den Fall allgemeiner bedingter Erwartungen gegeben eine σ -Algebra oder gegeben beliebige Zufallsvariablen übertragen. Sei dazu $X : \Omega \rightarrow \overline{\mathbb{R}}$ eine nichtnegative (oder integrierbare) Zufallsvariable auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) .

Definition (Bedingte Erwartung, allgemein). (1). Sei $\mathcal{F} \subseteq \mathcal{A}$ eine σ -Algebra. Eine nicht-negative (bzw. integrierbare) Zufallsvariable $\bar{X} : \Omega \rightarrow \bar{\mathbb{R}}$ heißt **Version der bedingten Erwartung** $E[X | \mathcal{F}]$, falls gilt:

- (a) \bar{X} ist \mathcal{F} -messbar, und
 (b) $E[\bar{X} ; A] = E[X ; A]$ für alle $A \in \mathcal{F}$.

(2). Für beliebige Zufallsvariablen Y, Y_1, Y_2, \dots, Y_n auf (Ω, \mathcal{A}, P) definieren wir

$$\begin{aligned} E[X | Y] &:= E[X | \sigma(Y)], \\ E[X | Y_1, \dots, Y_n] &:= E[X | (Y_1, \dots, Y_n)] = E[X | \sigma(Y_1, \dots, Y_n)]. \end{aligned}$$

(3). Für ein Ereignis $A \in \mathcal{A}$ definieren wir

$$P[A | \mathcal{F}] := E[I_A | \mathcal{F}], \quad \text{und entsprechend} \quad P[A | Y] = E[I_A | Y].$$

Bemerkung. Durch maßtheoretische Induktion zeigt man, dass Bedingung (b) äquivalent ist zu:

- (b') $E[\bar{X} \cdot Z] = E[X \cdot Z]$ für alle nichtnegativen (bzw. beschränkten) \mathcal{F} -messbaren $Z : \Omega \rightarrow \mathbb{R}$.

Satz 1.7 (Existenz und Eindeutigkeit der bedingten Erwartung). Sei $X \geq 0$ oder $X \in \mathcal{L}^1$, und $\mathcal{F} \subseteq \mathcal{A}$ eine σ -Algebra. Dann gilt

- (1). Es existiert eine Version der bedingten Erwartung $E[X | \mathcal{F}]$.
 (2). Zwei Versionen stimmen P -fast sicher überein.

Beweis. Die Existenz kann man unmittelbar aus dem Satz von Radon-Nikodym folgern, s. z.B. [A.Klenke, Wahrscheinlichkeitstheorie]. Wir geben stattdessen am Ende von Abschnitt 1.4 einen Existenzbeweis, der mit elementaren Methoden auskommt.

Zum Beweis der Eindeutigkeit seien \bar{X} und \tilde{X} zwei Versionen der bedingten Erwartung $E[X | \mathcal{F}]$. Dann sind \bar{X} und \tilde{X} beide \mathcal{F} -messbar, und

$$E[\bar{X} ; A] = E[\tilde{X} ; A] \quad \text{für alle } A \in \mathcal{F}.$$

Hieraus folgt $\bar{X} = \tilde{X}$ P -fast sicher. □

Bemerkung (Probleme mit Ausnahmemengen). Man beachte, dass die bedingte Erwartung $E[X | \mathcal{F}]$ und damit auch die bedingte Wahrscheinlichkeit $P[A | \mathcal{F}]$ nur für jede *feste* Zufallsvariable X bzw. jedes *feste* Ereignis A bis auf Modifikation auf Nullmengen eindeutig definiert sind. Ein weiteres Problem ist, dass wir allgemein zwar bedingte Erwartungen gegeben eine Zufallsvariable Y definieren können, aber nicht solche gegeben das Ereignis $Y = z$ für festes z . In vielen Fällen kann man die beschriebenen Probleme durch Auswahl einer „regulären Version der bedingten Verteilung gegeben Y “ umgehen. Wir kommen darauf in Korollar 1.9 zurück.

Bemerkung ($E[X | Y = z]$). Obwohl $E[X | Y = z]$ für ein festes z im Allgemeinen nicht definiert ist, kann man die Funktion $z \mapsto E[X | Y = z]$ bis auf Modifikation auf Nullmengen bzgl. der Verteilung von Y sinnvoll definieren: Ist $Y : \Omega \rightarrow S$ eine Zufallsvariable mit Werten in einem messbaren Raum (S, \mathcal{S}) , dann ist jede Version der bedingten Erwartung $E[X | Y]$ nach Definition $\sigma(Y)$ -messbar. Also gilt nach dem Faktorisierungslemma:

$$E[X | Y] = g(Y) \quad \text{für eine messbare Funktion } g : S \rightarrow \mathbb{R}. \quad (1.3.3)$$

Da die Versionen der bedingten Erwartung bis auf Modifikation auf P -Nullmengen eindeutig festgelegt sind, ist die Funktion g bis auf Modifikation auf μ_Y -Nullmengen eindeutig festgelegt. In Anlehnung an den diskreten Fall setzt man manchmal:

$$E[X | Y = z] := g(z). \quad (1.3.4)$$

Genauer definieren wir für eine nichtnegative Zufallsvariable X :

Definition. Eine messbare Funktion $g : S \rightarrow \mathbb{R}^+$ heißt **Version der bedingten Erwartung** $z \mapsto E[X | Y = z]$ **von X gegeben $Y = z$** , wenn gilt:

$$E[X ; Y \in B] = \int_B g(z) \mu_Y(dz) \quad \text{für alle } B \in \mathcal{S}. \quad (1.3.5)$$

Die charakterisierende Bedingung (1.3.5) ist nichts anderes als eine allgemeine Variante der *Formel von der totalen Wahrscheinlichkeit*. Mithilfe des Transformationssatzes sieht man, dass g genau dann (1.2.3) erfüllt, wenn $g(Y)$ eine Version von $E[X | Y]$ ist.

WARNUNG: Bei der Definition ist zu beachten, dass $E[X | Y = z]$ für ein festes z im Allgemeinen nicht definiert ist, sondern nur die Funktion $z \mapsto E[X | Y = z]$ modulo Modifikation auf μ_Y -Nullmengen! Das formale Rechnen mit bedingten Erwartungen wie in (1.3.4) ist daher eine häufige Fehlerquelle.

Trotz dieser Gefahren ist die Notation $E[X | Y = z]$ oft nützlich, um Argumentationen transparenter zu machen, oder um anschauliche Überlegungen in mathematische Formeln zu übersetzen. Wir werden sie daher auch hier gelegentlich verwenden.

1.3.3 Diskreter und absolutstetiger Fall

In einigen Fällen kann man die Definition direkt anwenden, um bedingte Erwartungswerte zu berechnen. Wir betrachten zunächst noch einmal den Spezialfall eine *diskreten Bedingung*:

Gilt $\mathcal{F} = \sigma(\{H_i | i \in \mathbb{N}\})$ für eine disjunkte Zerlegung $\Omega = \bigcup_{i \in \mathbb{N}} H_i$ in abzählbar viele messbare Teilmengen („Hypothesen“) $H_i \in \mathcal{A}$, dann sind \mathcal{F} -messbare Zufallsvariablen konstant auf jeder der Mengen H_i . Aus der Definition der bedingten Erwartung folgt dann

$$E[X | \mathcal{F}] = E[X | H_i] \quad \text{auf } H_i$$

für alle $i \in \mathbb{N}$ mit $P[H_i] > 0$.

Beispiel (Unbedingte Erwartungen). Die bedingte Erwartung einer Zufallsvariable X gegeben die triviale σ -Algebra $\{\emptyset, \Omega\}$ ist der Erwartungswert von X .

Beispiel (Bedingen auf eine Partition). Ist $P = \mathcal{U}_{(0,1)}$ die Gleichverteilung auf $[0, 1)$, und $\mathcal{F} = \sigma(\{[t_{i-1}, t_i] | i = 1, \dots, n\})$ die von einer Partition $0 = t_0 < t_1 < t_2 < \dots < t_n = 1$ erzeugte σ -Algebra, dann ist die bedingte Erwartung $E[g | \mathcal{F}]$ einer integrierbaren Funktion $g : [0, 1) \rightarrow \mathbb{R}$ die durch

$$E[g | \mathcal{F}] = \frac{1}{t_i - t_{i-1}} \int_{t_{i-1}}^{t_i} g(u) du \quad \text{auf } [t_{i-1}, t_i)$$

definierte Funktion.

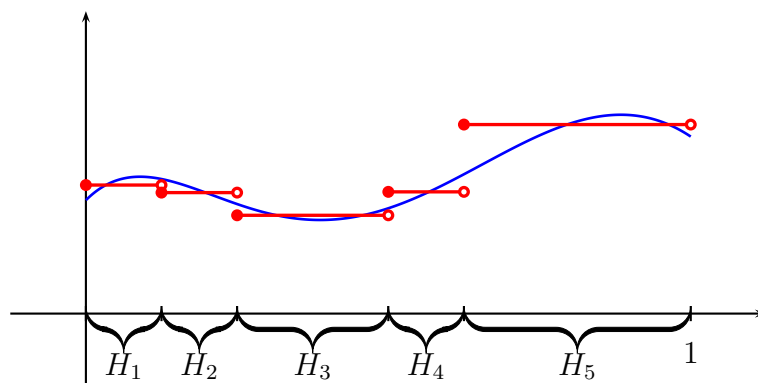


Abbildung 1.3: Die Funktion $g(\omega)$ ist hier blau dargestellt und $E[g | \mathcal{F}]$ in rot.

Ist die gemeinsame Verteilung aller relevanten Zufallsvariablen absolutstetig, dann kann man bedingte Erwartungen mithilfe von bedingten Dichten berechnen:

Satz 1.8 (Berechnung bedingter Erwartungen im absolutstetigen Fall). Seien $X : \Omega \rightarrow \mathbb{R}^n$ und $Y : \Omega \rightarrow \mathbb{R}^m$ Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , deren gemeinsame Verteilung $\mu_{X,Y}$ absolutstetig ist, und sei $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow [0, \infty]$ messbar. Dann ist

$$E[h(X, Y) | Y](\omega) = \int_{\mathbb{R}^n} h(x, Y(\omega)) f_{X|Y}(x|Y(\omega)) dx \quad (1.3.6)$$

eine Version der bedingten Erwartung von $h(X, Y)$ gegeben Y .

Beweis. Nach dem Satz von Fubini ist die rechte Seite von (1.3.6) eine messbare Funktion von $Y(\omega)$, und es gilt

$$\begin{aligned} E \left[g(Y) \cdot \int h(x, Y) f_{X|Y}(x|Y) dx \right] &= \int \int g(y) h(x, y) f_{X|Y}(x|y) f_Y(y) dx dy \\ &= E[g(Y) h(X, Y)] \end{aligned}$$

für jede messbare Funktion $g : \mathbb{R}^m \rightarrow [0, \infty]$. □

Mit der Notation aus (1.3.4) lautet die Aussage des Satzes:

$$E[h(X, Y) | Y = z] = \int_{\mathbb{R}^n} h(x, z) f_{X|Y}(x|z) dx \quad \text{für } \mu_Y\text{-fast alle } z \in S.$$

Um die bedingte Erwartung zu berechnen, müssen wir also den uns bekannten Wert von Y einsetzen, und die Funktion bzgl. der bedingten Dichte $f_{X|Y}$ nach x integrieren.

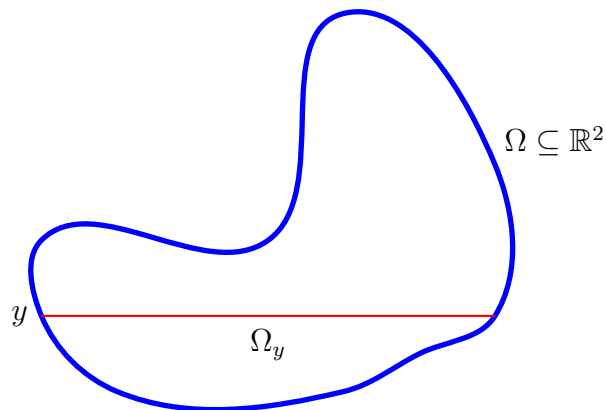
Beispiel (Bedingen auf eine Koordinate). Ist $P = \mathcal{U}_\Omega$ die Gleichverteilung auf einer beschränkten, messbaren Menge $\Omega \subseteq \mathbb{R}^2$, und ist

$$Y : \Omega \rightarrow \mathbb{R}, \quad Y(x, y) = y,$$

die Projektion auf die zweite Komponente, dann gilt

$$E[h|Y](x, y) = \frac{1}{\lambda(\Omega_y)} \int_{\Omega_y} h(x, y) dx \quad P\text{-fast sicher} \quad (1.3.7)$$

für jede integrierbare Funktion $h : \Omega \rightarrow \mathbb{R}$. Hierbei ist $\Omega_y = \{x \in \mathbb{R} | (x, y) \in \Omega\}$ der y -Schnitt von Ω . Bedingen auf Y entspricht hier also dem normierten „Herausintegrieren“ der komplementären Koordinate x .

Abbildung 1.4: In Rot: Der y -Schnitt der Menge Ω .

1.3.4 Reguläre bedingte Verteilungen

Beim Bedingen auf diskrete Zufallsvariablen konnten wir bedingte Wahrscheinlichkeitsverteilungen auf elementare Weise definieren. Für allgemeine Zufallsvariablen sind die bedingten Wahrscheinlichkeiten

$$P[X \in B | Y] = E[I_B(X) | Y]$$

für jede feste messbare Menge B nur bis auf Modifikation auf P -Nullmengen eindeutig definiert. Dies ist ein Nachteil, da die Ausnahmemenge von B abhängen kann, und im Allgemeinen überabzählbar viele messbare Mengen existieren. Die bedingte Verteilung von X gegeben Y ist daher zunächst nicht definiert. Im absolutstetigen Fall können wir das Problem umgehen, indem wir die über die bedingte Dichte gegebene Version

$$\mu_{X|Y}(y, dx) := f_{X|Y}(x|y)dx$$

der bedingten Verteilung verwenden. Aus Satz 1.8 folgt unmittelbar, dass wir bedingte Wahrscheinlichkeiten gegeben Y aus $\mu_{X|Y}$ berechnen können:

Korollar 1.9. *Ist die gemeinsame Verteilung der Zufallsvariablen $X : \Omega \rightarrow \mathbb{R}^n$ und $Y : \Omega \rightarrow \mathbb{R}^m$ absolutstetig, dann ist $\mu_{X|Y}$ eine **reguläre Version der bedingten Verteilung von X gegeben Y** , d.h.*

(1). $\mu_{X|Y}$ ist ein stochastischer Kern von \mathbb{R}^m nach \mathbb{R}^n .

(2). Für jedes $B \in \mathcal{B}(\mathbb{R}^n)$ ist

$$P[X \in B | Y] = \mu_{X|Y}(Y, B)$$

eine Version der bedingten Wahrscheinlichkeit von $\{X \in B\}$ gegeben Y .

Bemerkung (Existenz von regulären Versionen bedingter Verteilungen). Die Existenz von regulären Versionen von bedingten Verteilungen gegeben eine Zufallsvariable Y kann man allgemein beweisen, wenn Y Werte in einem vollständigen, separablen, metrischen Raum (kurz: polnischen Raum) annimmt, siehe z.B. [Breiman, Ch. 4.3.]. Eine explizite Berechnung über bedingte Dichten ist natürlich im Allgemeinen nicht möglich.

Wenn wir uns auf eine bestimmte reguläre Version $\mu_{X|Y}$ festlegen, dann können wir die bedingten Wahrscheinlichkeiten $P[X \in B | Y = z]$ durch

$$P[X \in B | Y = z] = \mu_{X|Y}(z, B)$$

für *alle* $z \in S$ definieren. Die Festlegung auf eine bestimmte reguläre Version der bedingten Verteilung ist im Allgemeinen willkürlich. Manchmal gibt es aber eine kanonische Version, die sich auszeichnet. Dies ist zum Beispiel der Fall, wenn die Dichte der gemeinsamen Verteilung von X und Y eine stetige Version hat.

Beispiel (Bivariate Normalverteilung). Ist (X, Y) bivariat normalverteilt mit Mittel $(0, 0)$ und Kovarianzmatrix $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $\rho \in (-1, 1)$, dann gilt

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \cdot \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right).$$

Für ein festes $x \in \mathbb{R}$ folgt

$$f_{Y|X}(y|x) \propto f_{X,Y}(x, y) \propto \exp\left(-\frac{(y - \rho x)^2}{2(1-\rho^2)}\right)$$

als Funktion von y . Also ist

$$\mu_{Y|X}(x, \bullet) = N(\rho x, 1 - \rho^2)$$

eine kanonische reguläre Version der bedingten Verteilung von Y gegeben X .

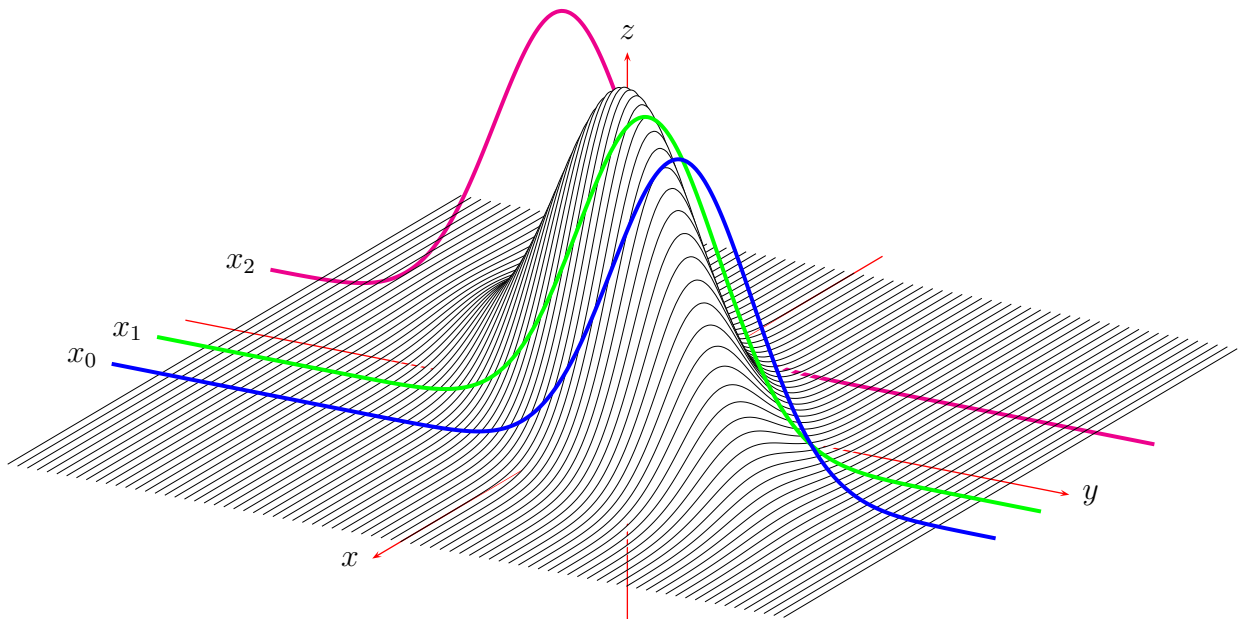


Abbildung 1.5: Die Dichte $f_{X,Y}(x,y)$ und in Blau, Grün und Magenta $f_{Y|X}(y|x_i)$ für $i \in \{0, 1, 2\}$. Man beachte, dass $f_{Y|X}(y|x_i) \propto f_{X,Y}(x_i, y)$ als Funktion von y .

Beispiel (Grenzen naiven Bedingens). Sei (X, Y) gleichverteilt auf dem Viertelkreis

$$S = \{(x, y) \in \mathbb{R}^2 | x > 0, y > 0, x^2 + y^2 < 1\}.$$

Wir versuchen auf zwei Arten eine „bedingte Verteilung von X gegeben $X = Y$ “ zu berechnen. Dazu betrachten wir die Zufallsvariablen $V = Y - X$ und $W = Y/X$. Wegen $f_{X,Y} \propto I_S$ erhalten wir mithilfe des Dichtetransformationssatzes für fast jedes v :

$$\begin{aligned} f_{X|V}(x|v) &\propto f_{X,V}(x,v) = f_{X,Y}(x, v+x) \cdot \left| \det \frac{\partial(x, v+x)}{\partial(x, v)} \right| \\ &\propto I_S(x, v+x), \end{aligned}$$

wobei „ \propto “ für „proportional als Funktion von x “ steht. Wählen wir die normierte rechte Seite als kanonische Version der bedingten Dichte, so ergibt sich

$$f_{X|V}(x|0) \propto I_S(x, x) = I_{(0, 1/\sqrt{2})}(x).$$

Gegeben $Y - X = 0$ ist X also gleichverteilt auf $(0, 1/\sqrt{2})$.

Andererseits erhalten wir für fast jedes w :

$$\begin{aligned} f_{X|W}(x|w) &\propto f_{X,W}(x,w) = f_{X,Y}(x, wx) \cdot \left| \det \frac{\partial(x, wx)}{\partial(x, w)} \right| \\ &\propto I_S(x, wx) \cdot x. \end{aligned}$$

Wählen wir wieder die rechte Seite als kanonische Version, so ergibt sich

$$f_{X|W}(x|1) \propto x \cdot I_S(x, x) = x \cdot I_{(0,1/\sqrt{2})}(x).$$

Die bedingte Verteilung von X gegeben $Y/X = 1$ unterscheidet sich also von der bedingten Verteilung von X gegeben $Y - X = 0$. Bedingte Wahrscheinlichkeiten gegeben $X = Y$ sind daher nicht wohldefiniert!

Eine anschauliche Erklärung für das Phänomen ist, dass wir in den beiden Fällen oben auf unterschiedliche infinitesimale Umgebungen der Diagonale $\{(x, y) \in S \mid x = y\}$ bedingen, wie die folgende Grafik veranschaulicht:

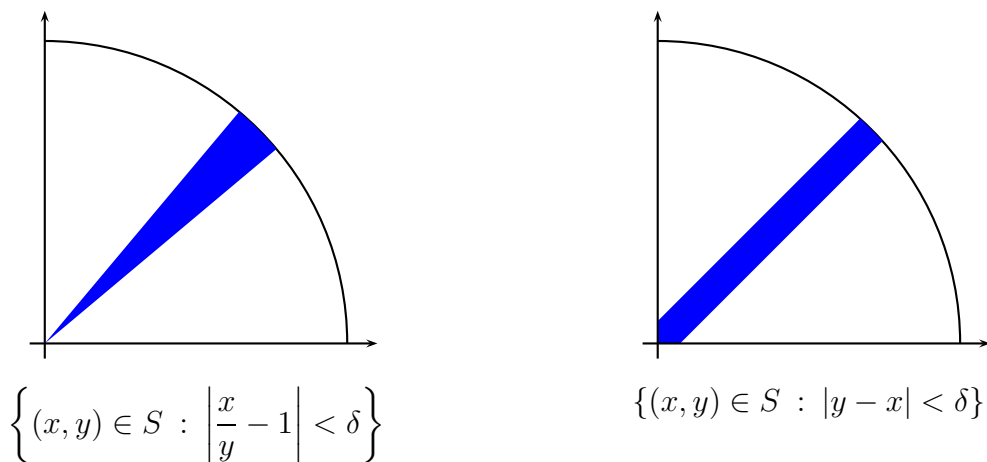


Abbildung 1.6: Zwei verschiedene Arten die Diagonale zu approximieren.

1.4 Rechnen mit bedingten Erwartungen; Poissonprozess

In vielen Fällen tritt eine Kombination bedingter Erwartungen bezüglich verschiedener Zufallsvariablen und/oder σ -Algebren auf. Die bedingten Erwartungswerte können dann meist nicht unmittelbar berechnet werden, lassen sich aber mithilfe grundlegender Eigenschaften und Rechenregeln schrittweise umformen und ggf. vereinfachen. Wir leiten nun aus der Definition einige fundamentale Eigenschaften bedingter Erwartungen her, die wir in diesem Zusammenhang häufig verwenden werden.

Als eine erste Anwendung untersuchen wir zeitliche und räumliche Poissonprozesse. Zeitliche Poissonprozesse sind die einfachsten Beispiele von zeitstetigen stochastischen Prozessen mit stationären unabhängigen Inkrementen, bzw. von zeitstetigen Markovketten. Räumliche Poissonprozesse (Poissonsche Punktprozesse) sind grundlegende Modelle für zufällige Punktmen-gen. Beide Arten von Prozessen spielen in etlichen Anwendungsbereichen eine wichtige Rolle

(z.B. Warteschlangen, Versicherungsmathematik, Materialwissenschaften, stochastische Geometrie etc.), und bilden die Basis für die Konstruktion vieler komplexerer stochastischer Modelle.

1.4.1 Eigenschaften der bedingten Erwartung

Wir leiten zunächst aus der Definition einige fundamentale Eigenschaften der bedingten Erwartung her, die wir häufig bei der Berechnung bedingter Erwartungswerte verwenden werden:

Satz 1.10. *Seien X, Y und X_n ($n \in \mathbb{N}$) nichtnegative oder integrierbare Zufallsvariablen auf (Ω, \mathcal{A}, P) , und seien $\mathcal{F}, \mathcal{G} \subseteq \mathcal{A}$ σ -Algebren.*

Es gelten folgende Aussagen:

(1). *Linearität:* $E[\lambda X + \mu Y | \mathcal{F}] = \lambda E[X | \mathcal{F}] + \mu E[Y | \mathcal{F}]$ *P -fast sicher für alle $\lambda, \mu \in \mathbb{R}$.*

(2). *Monotonie:* *Aus $X \geq 0$ P -fast sicher folgt $E[X | \mathcal{F}] \geq 0$ P -fast sicher.*

(3). *Aus $X = Y$ P -fast sicher folgt $E[X | \mathcal{F}] = E[Y | \mathcal{F}]$ P -fast sicher.*

(4). *Monotone Konvergenz:* *Ist (X_n) monoton wachsend mit $X_1 \geq 0$, dann gilt*

$$E[\sup X_n | \mathcal{F}] = \sup E[X_n | \mathcal{F}] \quad P\text{-fast sicher.}$$

(5). *Projektivität / Tower Property:* *Ist $\mathcal{G} \subseteq \mathcal{F}$, dann gilt*

$$E[E[X | \mathcal{F}] | \mathcal{G}] = E[X | \mathcal{G}] \quad P\text{-fast sicher.}$$

Insbesondere:

$$E[E[X | Y, Z] | Y] = E[X | Y] \quad P\text{-fast sicher.}$$

(6). *Herausziehen, was bekannt ist:* *Sei Y \mathcal{F} -messbar mit $Y \cdot X \in \mathcal{L}^1$ bzw. ≥ 0 . Dann gilt*

$$E[Y \cdot X | \mathcal{F}] = Y \cdot E[X | \mathcal{F}] \quad P\text{-fast sicher.}$$

(7). *Unabhängigkeit:* *Ist X unabhängig von \mathcal{F} , dann gilt $E[X | \mathcal{F}] = E[X]$ P -fast sicher.*

(8). *Seien (S, \mathcal{S}) und (T, \mathcal{T}) messbare Räume. Ist $Y : \Omega \rightarrow S$ \mathcal{F} -messbar, und $X : \Omega \rightarrow T$ unabhängig von \mathcal{F} , und $f : S \times T \rightarrow [0, \infty)$ eine produktmessbare Abbildung, dann gilt*

$$E[f(X, Y) | \mathcal{F}](\omega) = E[f(X, Y(\omega))] \quad \text{für } P\text{-fast alle } \omega.$$

Beweis. (1). Aus der Linearität des Erwartungswertes folgt, dass $\lambda E[X | \mathcal{F}] + \mu E[Y | \mathcal{F}]$ eine Version der bedingten Erwartung $E[\lambda X + \mu Y | \mathcal{F}]$ ist.

(2). Sei \bar{X} eine Version von $E[X | \mathcal{F}]$. Aus $X \geq 0$ P -fast sicher folgt wegen $\{\bar{X} < 0\} \in \mathcal{F}$:

$$E[\bar{X}; \bar{X} < 0] = E[X; \bar{X} < 0] \geq 0,$$

und damit $\bar{X} \geq 0$ P -fast sicher.

(3). Dies folgt unmittelbar aus (1) und (2).

(4). Ist $X_n \geq 0$ und monoton wachsend, dann ist $\sup E[X_n | \mathcal{F}]$ eine nichtnegative \mathcal{F} -messbare Zufallsvariable (mit Werten in $[0, \infty]$), und nach dem „klassischen“ Satz von der monotonen Konvergenz (siehe Satz ??) gilt:

$$E[\sup E[X_n | \mathcal{F}] \cdot Z] = \sup E[E[X_n | \mathcal{F}] \cdot Z] = \sup E[X_n \cdot Z] = E[\sup X_n \cdot Z]$$

für jede nichtnegative \mathcal{F} -messbare Zufallsvariable Z . Also ist $\sup E[X_n | \mathcal{F}]$ eine Version der bedingten Erwartung von $\sup X_n$ gegeben \mathcal{F} .

(5). Wir zeigen, dass jede Version von $E[X | \mathcal{G}]$ auch eine Version von $E[E[X | \mathcal{F}] | \mathcal{G}]$ ist, also die Eigenschaften (i) und (ii) aus der Definition der bedingten Erwartung erfüllt:

(i) $E[X | \mathcal{G}]$ ist nach Definition \mathcal{G} -messbar.

(ii) Für $A \in \mathcal{G}$ gilt auch $A \in \mathcal{F}$, und somit $E[E[X | \mathcal{G}]; A] = E[X; A] = E[E[X | \mathcal{F}]; A]$.

(6) und (7). Auf ähnliche Weise verifiziert man, dass die Zufallsvariablen, die auf der rechten Seite der Gleichungen in (6) und (7) stehen, die definierenden Eigenschaften der bedingten Erwartungen auf der linken Seite erfüllen (Übung).

(8). Dies folgt aus (6) und (7) in drei Schritten:

(a) Gilt $f(x, y) = g(x) \cdot h(y)$ mit messbaren Funktionen $g, h \geq 0$, dann folgt nach (6) und (7) P -fast sicher:

$$\begin{aligned} E[f(X, Y) | \mathcal{F}] &= E[g(X) \cdot h(Y) | \mathcal{F}] = h(Y) \cdot E[g(X) | \mathcal{F}] \\ &= h(Y) \cdot E[g(X)], \end{aligned}$$

und somit

$$E[f(X, Y) | \mathcal{F}](\omega) = E[g(X) \cdot h(Y(\omega))] = E[f(X, Y(\omega))] \quad \text{für } P\text{-fast alle } \omega.$$

- (b) Um die Behauptung für Indikatorfunktionen $f(x, y) = I_B(x, y)$ von produktmessbaren Mengen B zu zeigen, betrachten wir das Mengensystem

$$\mathcal{D} = \{B \in \mathcal{S} \otimes \mathcal{T} \mid \text{Behauptung gilt für } f = I_B\}.$$

\mathcal{D} ist ein Dynkinsystem, das nach (a) alle Produkte $B = B_1 \times B_2$ mit $B_1 \in \mathcal{S}$ und $B_2 \in \mathcal{T}$ enthält. Also gilt auch

$$\mathcal{D} \supseteq \sigma(\{B_1 \times B_2 \mid B_1 \in \mathcal{S}, B_2 \in \mathcal{T}\}) = \mathcal{S} \otimes \mathcal{T}.$$

- (c) Für beliebige produktmessbare Funktionen $f : S \times T \rightarrow \mathbb{R}_+$ folgt die Behauptung nun durch maßtheoretische Induktion. □

Bemerkung (Konvergenzsätze für bedingte Erwartungen). Aus dem Satz von der monotonen Konvergenz (Eigenschaft (4)) folgen auch Versionen des Lemmas von Fatou und des Satzes von der dominierten Konvergenz für bedingte Erwartungen. Der Beweis verläuft ähnlich wie im unbedingten Fall (Übung).

Die letzte Eigenschaft aus Satz 1.10 ist oft sehr nützlich. Für unabhängige Zufallsvariablen X und Y ergibt sich insbesondere

$$E[f(X, Y) \mid Y](\omega) = E[f(X, Y(\omega))] \quad \text{für } P\text{-fast alle } \omega, \quad (1.4.1)$$

d.h.

$$E[f(X, Y) \mid Y = z] = E[f(X, z)] \quad \text{für } \mu_Y\text{-fast alle } z. \quad (1.4.2)$$

Die Unabhängigkeit von X und Y ist wesentlich für (1.4.1) bzw. (1.4.2):

Beispiel. Ist $Y = X$, dann gilt offensichtlich

$$\begin{aligned} E[X \cdot Y \mid Y = z] &= E[Y^2 \mid Y = z] = z^2 && \text{für } \mu_Y\text{-fast alle } z, && \text{aber} \\ E[X \cdot z] &= z \cdot E[X] = z \cdot E[Y]. \end{aligned}$$

Das Anwenden der Formeln (1.4.1) und (1.4.2) ohne dass Unabhängigkeit vorliegt ist ein sehr häufiger Fehler beim Rechnen mit bedingten Erwartungen!

Beispiel (Summen von Wartezeiten). Für eine exponential-verteilte Zufallsvariable gilt

$$P[T > t + h \mid T > t] = P[T > h] \quad \text{für alle } t \geq 0 \text{ und } h \in \mathbb{R}.$$

Durch Bedingen können wir diese Aussage deutlich verallgemeinern:

Lemma 1.11 (Erweiterte Gedächtnislosigkeit). *Sind T und R unabhängige nichtnegative Zufallsvariablen, und ist T exponentialverteilt, dann gilt*

$$P[T + R > t + h \mid T > t] = P[T + R > h] \quad \text{für alle } t \geq 0 \text{ und } h \in \mathbb{R}.$$

Beweis. Durch Bedingen auf R erhalten wir nach (1.4.2) für $t \geq 0$:

$$\begin{aligned} P[T + R > t + h \text{ und } T > t \mid R = r] &\stackrel{(*)}{=} P[T + r > t + h \text{ und } T > t] \\ &= P[T > t + h - r \mid T > t] \cdot P[T > t] \\ &= P[T > h - r] \cdot P[T > t] \end{aligned}$$

für fast alle $r > 0$, also

$$\begin{aligned} P[T + R > t + h \text{ und } T > t] &= \int P[T + R > t + h, T > t \mid R = r] \mu_R(dr) \\ &= \int P[T > h - r] \mu_R(dr) \cdot P[T > t] \\ &\stackrel{(**)}{=} P[T + R > h] \cdot P[T > t]. \end{aligned}$$

Hierbei haben wir in $(*)$ und $(**)$ wesentlich benutzt, dass T und R unabhängig sind. \square

Das Lemma zeigt, dass für Summen von unabhängigen Wartezeiten eine Gedächtnislosigkeitseigenschaft gilt, sofern der erste Summand exponentialverteilt ist. Diese Tatsache ist von grundlegender Bedeutung um nachzuweisen, dass die zukünftige Weiterentwicklung von zeitstetigen Markovketten nicht vom Verlauf in der Vergangenheit, sondern nur vom gegenwärtigen Zustand abhängt. Wir betrachten zunächst exemplarisch den einfachsten Fall einer solchen zeitstetigen Markovkette - den Poissonprozess.

1.4.2 Poissonprozesse

Ein Poissonprozess mit Intensität $\lambda > 0$ ist ein zeitstetiger stochastischer Prozess, d.h. eine Kollektion $N_t, t \in [0, \infty)$, von Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , mit nichtnegativen ganzzahligen Werten. Der Prozess wartet jeweils eine $\text{Exp}(\lambda)$ -verteilte Zeit ab, und springt dann um eine Einheit nach oben. Naheliegende Anwendungen sind z.B. die Modellierung einer Warteschlange, oder der Anzahl der bei einer Versicherung auflaufenden Schadensfälle.

Um einen Poissonprozess zu konstruieren, wählen wir unabhängige exponentialverteilte Zufallsvariablen $T_1, T_2, \dots \geq 0$ mit festem Parameter $\lambda > 0$ auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , und setzen

$$S_n = T_1 + T_2 + \dots + T_n, \quad n \in \mathbb{N}, \quad \text{und}$$

$$N_t = \#\{n \in \mathbb{N} \mid S_n \leq t\}, \quad t \in [0, \infty).$$

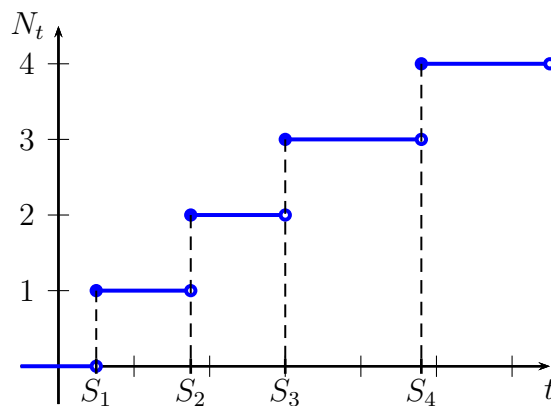


Abbildung 1.7: Darstellung von $N_t(\omega)$.

Dann ist $t \mapsto N_t(\omega)$ für alle ω monoton wachsend mit ganzzahligen Werten und $N_0(\omega) = 0$. Die Wartezeit S_n bis zum n -ten Sprung ist $\Gamma(\lambda, n)$ -verteilt, s. Lemma ???. Durch Bedingen können wir die Verteilungen des Prozesses $(N_t)_{t \geq 0}$ auf elegante Weise berechnen. Beispielsweise folgt aus der erweiterten Gedächtnislosigkeit (Lemma 1.11) für $t, h \geq 0$ unmittelbar

$$\begin{aligned} P[N_{t+h} < k \mid N_t = 0] &= P[S_k > t + h \mid S_1 > t] \\ &= P[T_1 + T_2 + \dots + T_k > t + h \mid T_1 > t] \\ &= P[T_1 + T_2 + \dots + T_k > h] \\ &= P[N_h < k] \quad \text{für alle } k \in \mathbb{N}, \end{aligned}$$

d.h. die bedingte Verteilung von N_{t+h} gegeben $N_t = 0$ stimmt mit der Verteilung von N_h überein. Allgemeiner erhalten wir:

Satz 1.12 (Eigenschaften des Poissonprozesses). Für $t, h \geq 0$ gilt:

- (1). $N_t \sim \text{Poisson}(\lambda t)$
- (2). *Stationarität:* $N_{t+h} - N_t \sim N_h$
- (3). *Unabhängige Inkremente:* $N_{t+h} - N_t \perp\!\!\!\perp \sigma(N_s \mid 0 \leq s \leq t)$.

Beweis. (1). *Verteilung von N_t :* Da $S_k = T_1 + \dots + T_k$ unabhängig von T_{k+1} und $\Gamma(\lambda, k)$ -verteilt ist, erhalten wir für $k \in \mathbb{N}$ nach (1.4.2):

$$\begin{aligned}
 P[N_t = k] &= P[S_k \leq t < S_{k+1}] \\
 &= \int P[S_k \leq t < S_k + T_{k+1} \mid S_k = u] : \mu_{S_k}(du) \\
 &= \int I_{(0,t]}(u) \cdot P[t < u + T_{k+1}] \mu_{S_k}(du) \\
 &= \int_0^t e^{-\lambda(t-u)} \cdot \frac{1}{(k-1)!} \lambda^k u^{k-1} e^{-\lambda u} du \\
 &= \frac{(\lambda t)^k}{k!} e^{-\lambda t}.
 \end{aligned}$$

Also ist N_t Poisson-verteilt zum Parameter λt .

(2). *Gemeinsame Verteilung von N_t und N_{t+h} :* Seien $k, l \geq 0$. Wegen $S_k = T_1 + \dots + T_k$ und $S_{k+l} = S_k + T_{k+1} + \dots + T_{k+l}$ erhalten wir nach (1.4.1) aufgrund der Unabhängigkeit der T_i :

$$\begin{aligned}
 &P[N_{t+h} < k+l, N_t = k \mid T_1, \dots, T_k](\omega) \\
 &= P[S_{k+l} > t+h, S_k \leq t < S_{k+1} \mid T_1, \dots, T_k](\omega) \\
 &= P[S_k(\omega) + T_{k+1} + \dots + T_{k+l} > t+h, S_k(\omega) \leq t < S_k(\omega) + T_{k+1}] \quad (1.4.3) \\
 &= P[T_{k+1} + \dots + T_{k+l} > h] \cdot P[T_{k+1} > t - S_k(\omega)] \cdot I_{\{S_k \leq t\}}(\omega) \\
 &= P[N_h < l] \cdot P[N_t = k \mid T_1, \dots, T_k](\omega)
 \end{aligned}$$

für P -fast alle ω . Hierbei haben wir im vorletzten Schritt Lemma 1.11 verwendet. Aus (a) folgt:

$$\begin{aligned}
 P[N_{t+h} - N_t < l, N_t = k] &= E[P[N_{t+h} < k+l, N_t = k \mid T_1, \dots, T_k]] \\
 &= P[N_h < l] \cdot P[N_t = k], \quad (1.4.4)
 \end{aligned}$$

d.h.

$$P[N_{t+h} - N_t < l \mid N_t = k] = P[N_h < l] \quad \text{für alle } k, l \geq 0.$$

Also ist das Inkrement $N_{t+h} - N_t$ unabhängig von N_t mit Verteilung

$$P \circ (N_{t-h} - N_t)^{-1} = P \circ N_h^{-1} = \text{Poisson}(\lambda h).$$

- (3). *Unabhängigkeit von $N_{t+h} - N_t$ und $\sigma(N_s | 0 \leq s \leq t)$* : Wir bemerken zunächst, dass für jedes Ereignis $A \in \sigma(N_s | 0 \leq s \leq t)$ und $k \geq 0$ ein Ereignis $A_k \in \sigma(T_1, \dots, T_k)$ existiert mit

$$A \cap \{N_t = k\} = A_k \cap \{N_t = k\}. \quad (1.4.5)$$

Zum Beweis kann man sich auf Ereignisse der Form $A = \{N_s = l\}$ mit $s \in [0, t]$ und $l \geq 0$ beschränken, da diese die σ -Algebra $\sigma(N_s | 0 \leq s \leq t)$ erzeugen. Für solche Ereignisse A gilt in der Tat

$$A \cap \{N_t = k\} = \{N_s = l, N_t = k\} = \{S_l \leq s < S_{l+1}, S_k \leq t < S_{k+1}\} = A_k \cap \{N_t = k\}$$

wobei

$$A_k := \begin{cases} \emptyset & \text{falls } l > k, \\ \{S_l \leq s\} & \text{falls } l = k, \\ \{S_l \leq s < S_{l+1}\} & \text{falls } l < k, \end{cases}$$

ein Ereignis ist, das nur von T_1, \dots, T_k abhängt.

Nach (1.4.5) erhalten wir für $A \in \sigma(N_s | 0 \leq s \leq t)$ und $k, l \geq 0$ analog zu (1.4.4):

$$\begin{aligned} & P[\{N_{t+h} - N_t < l\} \cap A \cap \{N_t = k\}] \\ &= E[P[N_{t+h} - N_t < l, N_t = k | T_1, \dots, T_k]; A_k] \\ &= P[N_h < l] \cdot P[A_k \cap \{N_t = k\}] \\ &= P[N_{t+h} - N_t < l] \cdot P[A \cap \{N_t = k\}]. \end{aligned}$$

Durch Summieren über k folgt die Unabhängigkeit von $N_{t+h} - N_t$ und A .

□

Aus Satz 1.12 folgt, dass für jede Partition $t_0 < t_1 < \dots < t_k$ die Inkremente $N_{t_1} - N_{t_0}, N_{t_2} - N_{t_1}, \dots, N_{t_k} - N_{t_{k-1}}$ unabhängige Zufallsvariablen mit Verteilung

$$N_t - N_s \sim \text{Poisson}(\lambda \cdot (t - s)), \quad 0 \leq s \leq t, \quad (1.4.6)$$

sind. Insbesondere sind die Inkremente **stationär**, d.h. die Verteilung von $N_t - N_s$ hängt nur von $t - s$ ab.

Definition (Lévy-Prozess). (1). Ein stochastischer Prozess $(N_t)_{t \geq 0}$ auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) heißt **Lévy-Prozess**, falls

- (a) die Inkremente $N_t - N_s, 0 \leq s \leq t$, stationär sind, und

(b) Inkremente über disjunkten Intervallen unabhängig sind.

(2). Ein Lévy-Prozess heißt **Poissonprozess mit Intensität** $\lambda > 0$, falls (1.4.6) gilt.

Weitere wichtige Beispiele von Lévy-Prozessen sind Brownsche Bewegungen und α -stabile Prozesse. Eine *Brownsche Bewegung* $(B_t)_{t \geq 0}$ ist ein Lévy-Prozess mit normalverteilten Inkrementen $B_t - B_s \sim N(0, t - s)$, $0 \leq s \leq t$, dessen Pfade $t \mapsto B_t(\omega)$ für P -fast alle ω stetig sind.

Prozesse in diskreter Zeit mit unabhängigen stationären Inkrementen sind Random Walks. Lévy-Prozesse kann man aus Random Walks durch Grenzübergänge mit unterschiedlichen Skalierungen erhalten (Poissonapproximation, zentraler Grenzwertsatz, Grenzwertsatz für Inkremente mit heavy tails etc.). Den Poissonprozess erhält man beispielsweise als Grenzwert für $k \rightarrow \infty$ der reskalierten Random Walks $N_t^{(k)} = S_{\lfloor kt \rfloor}^{(k)}$,

$$S_n^{(k)} = \sum_{i=1}^n X_i^{(k)}, \quad X_i^{(k)} \text{ unabhängig, } \sim \text{Bernoulli}(\lambda/k).$$

Die Simulation in Abbildung ?? deutet an, wie andere Lévyprozesse als Skalierungslimiten von Random Walks auftreten.

Ein weiteres Beispiel für Lévy-Prozesse sind zusammengesetzte (compound) Poissonprozesse:

Beispiel (Compound Poisson-Prozess). Sei μ eine Wahrscheinlichkeitsverteilung auf \mathbb{R}^d und $\lambda > 0$. Dann heißt der stochastische Prozess

$$S_t = \sum_{i=1}^{N_t} X_i, \quad t \geq 0,$$

mit unabhängigen Zufallsvariablen X_i mit Verteilung μ und einem von den X_i unabhängigen Poissonprozess $(N_t)_{t \geq 0}$ mit Intensität λ , **Compound-Poisson-Prozess mit Sprungverteilung μ und Intensität λ** . Der Compound-Poisson-Prozess ist eine zeitstetige Version des Random Walks mit Inkrementen X_i . Er wartet jeweils eine $\text{Exp}(\lambda)$ -verteilte Zeit ab, und macht dann einen Sprung gemäß der Verteilung μ . Entsprechende Prozesse werden u. A. in der Versicherungsmathematik zur Modellierung der akkumulierten Schadenshöhe bis zur Zeit t verwendet. Die Verteilung S_t für ein festes $t \geq 0$ kann man mit den oben eingeführten Methoden für zufällige Summen berechnen. Zudem kann man beweisen, dass $(S_t)_{t \geq 0}$ in der Tat ein Prozess mit stationären unabhängigen Inkrementen ist.

1.4.3 Poissonscher Punktprozess

Die Sprungzeitpunkte eines Poissonprozesses in einem endlichen Zeitintervall $(s, t]$ kann man auch auf andere Weise konstruieren: Ist Z eine Poisson-verteilte Zufallsvariable mit Parame-

ter $\lambda \cdot (t - s)$, und sind U_1, U_2, \dots unabhängig voneinander und von Z , und gleichverteilt auf $(s, t]$, dann sind U_1, \dots, U_Z die Sprungzeiten eines Poissonprozesses mit Parameter λ (s. Korollar 1.14). Allgemeiner sei nun ν ein endliches Maß auf einem messbaren Raum (S, \mathcal{S}) . Wir wollen eine zufällige „Punkt看ke“ in S mit Intensität ν konstruieren. Dazu wählen wir unabhängige Zufallsvariablen $X_1, X_2, \dots : \Omega \rightarrow S$ mit Verteilung $\mu = \frac{\nu}{\nu(S)}$, und setzen für $A \subseteq S$:

$$N(A) = \sum_{i=1}^Z \delta_{X_i}[A] = \#\{1 \leq i \leq Z \mid X_i \in A\}, \quad (1.4.7)$$

wobei Z (Gesamtzahl der Punkte) unabhängig von den X_i und Poisson-verteilt mit Parameter $\nu(S)$ ist. Die Abbildung $A \mapsto N(A)$ ist die Häufigkeitsverteilung der Punkte X_1, \dots, X_Z , und damit ein zufälliges Maß. Hat das Intensitätsmaß ν keine Atome (d.h. gilt $\nu[\{x\}] = 0$ für alle $x \in S$), dann sind die Punkte X_i mit Wahrscheinlichkeit 1 alle verschieden, und wir können N P -fast sicher mit der zufälligen Punktmenge $\{X_1, X_2, \dots, X_Z\} \subseteq S$ identifizieren.

Satz 1.13 (Konstruktion von Poissonschen Punktprozessen). *Das durch (1.4.7) definierte zufällige Maß N ist ein **Poissonscher Punktprozess mit Intensitätsmaß** ν , d.h. für beliebige $k \in \mathbb{N}$ und disjunkte Teilmenge $A_1, \dots, A_k \subseteq S$, sind die Zufallsvariablen $N(A_1), \dots, N(A_k)$ unabhängig mit Verteilung*

$$N(A_i) \sim \text{Poisson}(\nu(A_i)).$$

Zum Beweis benötigen wir die erzeugende Funktion der gemeinsamen Verteilung mehrerer Zufallsvariablen:

Definition (Erzeugende Funktion und gemeinsame Verteilung). *Seien $N_1, \dots, N_k : \Omega \rightarrow \{0, 1, 2, \dots\}$ nichtnegative ganzzahlige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , und sei*

$$\nu(n_1, \dots, n_k) = P[N_1 = n_1, \dots, N_k = n_k].$$

*Die **erzeugende Funktion** des Zufallsvektors (N_1, \dots, N_n) bzw. der Wahrscheinlichkeitsverteilung ν auf $\{0, 1, 2, \dots\}^k$ ist die durch*

$$G(s_1, \dots, s_k) = E[s_1^{N_1} s_2^{N_2} \cdot \dots \cdot s_k^{N_k}] = \sum_{n_1, \dots, n_k=0}^{\infty} \nu(n_1, \dots, n_k) \cdot s_1^{n_1} s_2^{n_2} \cdot \dots \cdot s_k^{n_k}$$

definierte Funktion $G : [0, 1]^k \rightarrow [0, 1]$.

Die gemeinsame Verteilung ν ist ähnlich wie im eindimensionalen Fall eindeutig durch die erzeugende Funktion festgelegt, denn für $n_1, \dots, n_k \in \{0, 1, 2, \dots\}$ gilt:

$$\nu(n_1, \dots, n_k) = \frac{1}{n_1! \cdot \dots \cdot n_k!} \cdot \frac{\partial^{n_1+n_2+\dots+n_k}}{\partial s_1^{n_1} \cdot \dots \cdot \partial s_k^{n_k}}(0, \dots, 0).$$

Beweis. O.B.d.A. können wir $S = \bigcup_{i=1}^k A_i$ annehmen. Wir berechnen für diesen Fall die erzeugende Funktion der gemeinsamen Verteilung von $N(A_1), \dots, N(A_k)$. Für $s_1, \dots, s_k \in [0, 1)$ gilt

$$\prod_{j=1}^k s_j^{N(A_j)} = \prod_{i=1}^Z \prod_{j=1}^k s_j^{I_{A_j}(X_i)},$$

also wegen der Unabhängigkeit von Z und den X_i :

$$E \left[\prod_{j=1}^k s_j^{N(A_j)} \middle| Z \right] = \prod_{i=1}^Z E \left[\prod_{j=1}^k s_j^{I_{A_j}(X_i)} \right] = \left(\sum_{j=1}^k s_j \cdot \mu[A_j] \right)^Z.$$

Hierbei haben wir im letzten Schritt verwendet, dass das Produkt über j gleich s_j ist, falls X_i in der Menge A_j liegt. Da Z Poisson-verteilt ist mit Parameter $\nu(S)$, erhalten wir

$$\begin{aligned} E \left[\prod_{j=1}^k s_j^{N(A_j)} \right] &= G_Z \left(\sum_{j=1}^k s_j \cdot \mu[A_j] \right) \\ &= \exp \left(\nu(S) \cdot \left(\sum_{j=1}^k s_j \cdot \mu[A_j] - 1 \right) \right) \\ &= \prod_{j=1}^k \exp(\nu(A_j) \cdot (s_j - 1)), \end{aligned}$$

d.h. die erzeugende Funktion von $(N(A_1), \dots, N(A_k))$ ist das Produkt der erzeugenden Funktionen von Poissonverteilungen mit Parametern $\nu(A_j)$. Hieraus folgt, dass die gemeinsame Verteilung der Zufallsvariablen $N(A_1), \dots, N(A_k)$ das Produkt dieser Poissonverteilungen ist. \square

Poissonsche Punktprozesse bezeichnet man auch synonym als *räumliche Poissonprozesse*, *Poissonsche Zufallsmaße*, oder *Poissonsche Felder*. Sie spielen eine wichtige Rolle bei der Modellierung zufälliger räumlicher Strukturen, z.B. in der stochastischen Geometrie. Satz 1.13 liefert uns einen einfachen Algorithmus zur Simulation Poissonscher Punktprozesse. Graphik ?? wurde mit diesem Algorithmus erzeugt. Als eindimensionalen Spezialfall von Satz 1.13 erhalten wir eine alternative Konstruktion von zeitlichen Poissonprozessen:

Korollar 1.14. Seien $\lambda, a \in (0, \infty)$. Sind Z, U_1, U_2, \dots unabhängige Zufallsvariablen mit Verteilungen $Z \sim \text{Poisson}(\lambda \cdot a)$ und $U_1, U_2, \dots \sim \text{Unif}_{(0,a)}$, dann ist

$$N_t := \sum_{i=1}^Z I_{[0,t]}(U_i), \quad 0 \leq t \leq a,$$

ein Poissonprozess mit Intensität λ .

Beweis. Es gilt $N_t = \overline{N}([0, t])$, wobei \overline{N} der wie in (1.4.7) definierte Poissonsche Punktprozess auf $S = [0, a]$ mit homogenem Intensitätsmaß $\lambda \cdot dt$ ist. Nach Satz 1.13 folgt, dass für jede Partition $0 \leq t_0 < t_1 < \dots < t_k \leq a$ die Inkremente

$$N_{t_j} - N_{t_{j-1}} = \overline{N}((t_{j-1}, t_j]), \quad 1 \leq j \leq k,$$

unabhängig und $\text{Poisson}(\lambda \cdot (t_j - t_{j-1}))$ -verteilt sind. \square

Poissonsche Punktprozesse lassen sich durch verschiedene Transformationen wieder in Poissonsche Punktprozesse überführen. Bildet man beispielsweise die Punkte $X_i, 1 \leq i \leq Z$, eines Poissonschen Punktprozesses N mit Intensitätsmaß ν mit einer (messbaren) Abbildung ϕ ab, dann erhält man einen Poissonschen Punktprozess

$$\tilde{N}(A) := \sum_{i=1}^Z \delta_{\phi(X_i)}[A] = \sum_{i=1}^Z I_A(\phi(X_i)) = \sum_{i=1}^Z I_{\phi^{-1}(A)}(X_i)$$

mit Intensitätsmaß $\tilde{\nu} = \nu \circ \phi^{-1}$. Zudem gilt eine Ausdünnungseigenschaft:

Seien $Z, X_1, X_2, \dots, U_1, U_2, \dots$ unabhängige Zufallsvariablen mit Verteilungen

$$Z \sim \text{Poisson}(\nu(S)), \quad X_i \sim \frac{\nu}{\nu(S)}, \quad U_i \sim \text{Unif}_{(0,1)},$$

und sei $\alpha : S \rightarrow [0, 1]$ eine messbare Funktion (Akzeptanzwahrscheinlichkeit). Wir konstruieren einen ausgedünnten Punktprozess N_α , indem wir einen Punkt X_i nur mit Wahrscheinlichkeit $\alpha(X_i)$ berücksichtigen:

$$N_\alpha := \sum_{i=1}^Z I_{\{U_i \leq \alpha(X_i)\}} \delta_{X_i}.$$

Satz 1.15 (Färbungssatz, Ausdünnungseigenschaft). N_α ist ein Poissonscher Punktprozess mit Intensitätsmaß $\alpha(x)\nu(dx)$.

Der Beweis wird dem Leser als Übung überlassen. Bemerkenswert ist unter Anderem, dass die beschriebene Konstruktion eine Kopplung von Poissonprozessen mit verschiedenen Intensitätsmaßen, d.h. eine simultane Konstruktion dieser Prozesse auf einem gemeinsamen Wahrscheinlichkeitsraum ermöglicht.

1.5 Bedingte Erwartung als beste L^2 -Approximation

In diesem Abschnitt zeigen wir, dass sich die bedingte Erwartung einer quadratintegrierbaren Zufallsvariable X gegeben eine σ -Algebra \mathcal{F} charakterisieren lässt als beste Approximation von

X im Unterraum der \mathcal{F} -messbaren quadratintegrierbaren Zufallsvariablen, bzw. als orthogonale Projektion von X auf diesen Unterraum. Neben naheliegenden Anwendungen auf nichtlineare Prognosen liefert uns dies auch einen einfachen Existenzbeweis für die bedingte Erwartung.

1.5.1 Jensensche Ungleichung

Die Jensensche Ungleichung gilt auch für bedingte Erwartungen.

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ eine integrierbare Zufallsvariable und $\mathcal{F} \subseteq \mathcal{A}$ eine σ -Algebra.

Satz 1.16 (Jensen). *Ist $u : \mathbb{R} \rightarrow \mathbb{R}$ eine konvexe Funktion mit $u(X) \in \mathcal{L}^1$ oder $u \geq 0$, dann gilt*

$$E[u(X) | \mathcal{F}] \geq u(E[X | \mathcal{F}]) \quad P\text{-fast sicher.}$$

Beweis. Jede konvexe Funktion u lässt sich als Supremum von abzählbar vielen affinen Funktionen darstellen, d.h. es gibt $a_n, b_n \in \mathbb{R}$ mit

$$u(x) = \sup_{n \in \mathbb{N}} (a_n x + b_n) \quad \text{für alle } x \in \mathbb{R}.$$

Zum Beweis betrachtet man die Stützgeraden an allen Stellen einer abzählbaren dichten Teilmenge von \mathbb{R} , siehe z.B. [Williams: Probability with martingales, 6.6]. Wegen der Monotonie und Linearität der bedingten Erwartung folgt

$$E[u(X) | \mathcal{F}] \geq E[a_n X + b_n | \mathcal{F}] = a_n \cdot E[X | \mathcal{F}] + b_n$$

P -fast sicher für alle $n \in \mathbb{N}$, also auch

$$E[u(X) | \mathcal{F}] \geq \sup_{n \in \mathbb{N}} (a_n \cdot E[X | \mathcal{F}] + b_n) \quad P\text{-fast sicher.}$$

□

Korollar 1.17 (L^p -Kontraktivität). *Die Abbildung $X \mapsto E[X | \mathcal{F}]$ ist eine Kontraktion auf $\mathcal{L}^p(\Omega, \mathcal{A}, P)$ für alle $p \geq 1$, d.h.*

$$E[|E[X | \mathcal{F}]|^p] \leq E[|X|^p] \quad \text{für alle } X \in \mathcal{L}^1(\Omega, \mathcal{A}, P).$$

Beweis. Nach der Jensenschen Ungleichung gilt:

$$|E[X | \mathcal{F}]|^p \leq E[|X|^p | \mathcal{F}] \quad P\text{-fast sicher.}$$

Die Behauptung folgt durch Bilden des Erwartungswertes. □

Im Beweis des Korollars haben wir insbesondere gezeigt, dass für eine Zufallsvariable $X \in \mathcal{L}^p$ auch die bedingte Erwartung $E[X | \mathcal{F}]$ in \mathcal{L}^p enthalten ist. Wir beschränken uns nun auf den Fall $p = 2$.

1.5.2 Bedingte Erwartung als beste L^2 -Prognose

Der Raum $L^2(\Omega, \mathcal{A}, P) = \mathcal{L}^2(\Omega, \mathcal{A}, P) / \sim$ der Äquivalenzklassen von quadratintegrierbaren Zufallsvariablen ist ein Hilbertraum mit Skalarprodukt $(X, Y)_{L^2} = E[XY]$. Ist $\mathcal{F} \subseteq \mathcal{A}$ eine Unter- σ -Algebra, dann ist $L^2(\Omega, \mathcal{F}, P)$ ein **abgeschlossener Unterraum** von $L^2(\Omega, \mathcal{A}, P)$, denn Grenzwerte von \mathcal{F} -messbaren Zufallsvariablen sind wieder \mathcal{F} -messbar. Nach der Jensenschen Ungleichung ist für $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ jede Version der bedingten Erwartung $E[X | \mathcal{F}]$ im Unterraum $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ der \mathcal{F} -messbaren quadratintegrierbaren Zufallsvariablen enthalten. Außerdem respektiert die bedingte Erwartung Äquivalenzklassen, s. Satz 1.7. Die Zuordnung $X \mapsto E[X | \mathcal{F}]$ definiert also eine lineare Abbildung vom Hilbertraum $L^2(\Omega, \mathcal{A}, P)$ der Äquivalenzklassen auf den Unterraum $L^2(\Omega, \mathcal{F}, P)$.

Satz 1.18 (Charakterisierung der bedingten Erwartung als beste L^2 Approximation und als orthogonale Projektion). Für $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ sind äquivalent:

- (1). Y ist eine Version der bedingten Erwartung $E[X | \mathcal{F}]$.
- (2). Y ist eine „**beste Approximation**“ von X im Unterraum $\mathcal{L}^2(\Omega, \mathcal{F}, P)$, d.h.

$$E[(X - Y)^2] \leq E[(X - Z)^2] \quad \text{für alle } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P).$$

- (3). Y ist eine Version der **orthogonalen Projektion** von X auf den Unterraum $L^2(\Omega, \mathcal{F}, P) \subseteq L^2(\Omega, \mathcal{A}, P)$, d.h.

$$E[(X - Y) \cdot Z] = 0 \quad \text{für alle } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P).$$

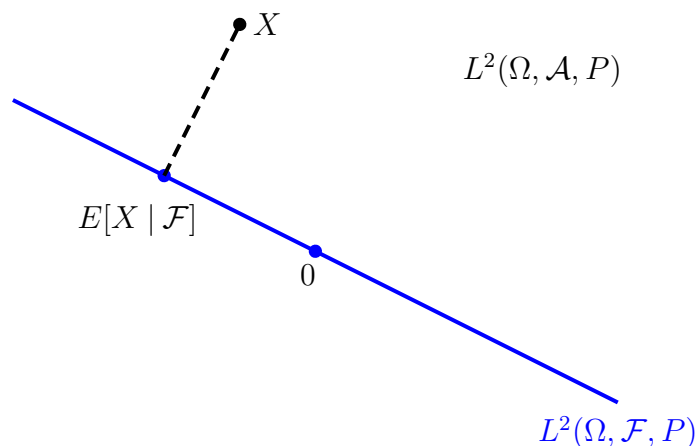


Abbildung 1.8: Darstellung von $X \mapsto E[X | \mathcal{F}]$ als orthogonale Projektion auf den Unterraum $L^2(\Omega, \mathcal{F}, P)$.

Beweis. **(1) \iff (3):** Für $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ gilt:

$$\begin{aligned} & Y \text{ ist eine Version von } E[X \mid \mathcal{F}] \\ \iff & E[Y \cdot I_A] = E[X \cdot I_A] \quad \text{für alle } A \in \mathcal{F} \\ \iff & E[Y \cdot Z] = E[X \cdot Z] \quad \text{für alle } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P) \\ \iff & E[(X - Y) \cdot Z] = 0 \quad \text{für alle } Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P) \end{aligned}$$

Hierbei zeigt man die zweite Äquivalenz mit den üblichen Fortsetzungsverfahren (maßtheoretische Induktion).

(3) \implies (2): Sei Y eine Version der orthogonalen Projektion von X auf $L^2(\Omega, \mathcal{F}, P)$. Dann gilt für alle $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$:

$$\begin{aligned} E[(X - Z)^2] &= E[((X - Y) + (Y - Z))^2] \\ &= E[(X - Y)^2] + E[(Y - Z)^2] + 2E[(X - Y) \underbrace{(Y - Z)}_{\in \mathcal{L}^2(\Omega, \mathcal{F}, P)}] \\ &\geq E[(X - Y)^2] \end{aligned}$$

Hierbei haben wir im letzten Schritt verwendet, dass $Y - Z$ im Unterraum $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ enthalten, also orthogonal zu $X - Y$ ist.

(2) \implies (3): Ist umgekehrt Y eine beste Approximation von X in $\mathcal{L}^2(\Omega, \mathcal{F}, P)$ und $Z \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$, dann gilt

$$\begin{aligned} E[(X - Y)^2] &\leq E[(X - Y + tZ)^2] \\ &= E[(X - Y)^2] + 2tE[(X - Y)Z] + t^2E[Z^2] \end{aligned}$$

für alle $t \in \mathbb{R}$, also $E[(X - Y) \cdot Z] = 0$.

□

Die Äquivalenz von (2) und (3) ist eine bekannte funktionalanalytische Aussage: die beste Approximation eines Vektors in einem abgeschlossenen Unterraum eines Hilbertraums ist die orthogonale Projektion des Vektors auf diesen Unterraum. Die dahinterstehende geometrische Intuition verdeutlicht man sich leicht anhand von Abbildung 1.8.

Satz 1.18 rechtfertigt die Verwendung der bedingten Erwartung als Prognoseverfahren. Beispielsweise ist $E[X \mid Y]$ nach dem Faktorisierungslemma die beste L^2 -Prognose für X unter allen Funktionen vom Typ $g(Y)$, $g: \mathbb{R} \rightarrow \mathbb{R}$ messbar.

Beispiel (Nichtlineare Prognose). Seien $S, T : \Omega \rightarrow \mathbb{R}_+$ unabhängige Zufallsvariablen, die zum Beispiel die Ausfallzeiten zweier Komponenten eines Systems beschreiben. S sei exponentialverteilt mit Parameter $\lambda > 0$ - die Verteilung von T ist beliebig. Angenommen, wir können nur den Ausfall der einen Komponente (mit Ausfallzeit T) beobachten, und wir möchten den Wert der ersten Ausfallzeit

$$X = \min(T, S)$$

aufgrund des beobachteten Wertes $T(\omega)$ prognostizieren. Nach Satz 1.18 ist der beste Prognosewert für X bzgl. des mittleren quadratischen Fehlers durch

$$\hat{X}(\omega) = E[X | T](\omega)$$

gegeben. Explizit erhalten wir wegen der Unabhängigkeit von T und S :

$$\begin{aligned} E[X | T](\omega) &= E[\min(T(\omega), S)] \\ &= \int_0^{\infty} \min(T(\omega), s) \lambda e^{-\lambda s} ds \\ &= \int_0^{T(\omega)} s \lambda e^{-\lambda s} ds + \int_{T(\omega)}^{\infty} T(\omega) \lambda e^{-\lambda s} ds \\ &= \frac{1}{\lambda} (1 - e^{-\lambda T(\omega)}) \quad \text{für } P\text{-fast alle } \omega. \end{aligned}$$

Die beste Prognose im quadratischen Mittel hängt also in diesem Fall *nichtlinear* von T ab. Sie unterscheidet sich damit von der *besten linearen Prognose* (Regressionsgerade), die wie in Abschnitt ?? gezeigt durch

$$\hat{X}_{\text{lin}} = aT + b \quad \text{mit} \quad a = \frac{\text{Cov}[X, T]}{\text{Var}[T]}, \quad b = E[X] - aE[T]$$

gegeben ist. Dass sich \hat{X} und \hat{X}_{lin} unterscheiden ist die Regel. Eine wichtige Ausnahme ergibt sich, wenn die gemeinsame Verteilung von X und T eine Gaußverteilung ist - in diesem Fall ist die beste L^2 Prognose $E[X | T]$ stets eine affine Funktion von T .

1.5.3 Existenz der bedingten Erwartung

Durch die Charakterisierung der bedingten Erwartung als beste L^2 -Approximation ergibt sich die Existenz der bedingten Erwartung einer quadratintegrierbaren Zufallsvariable unmittelbar aus der Existenz der Bestapproximation eines Vektors in einem abgeschlossenen Unterraum eines Hilbertraums. Durch monotone Approximation folgt hieraus die Existenz der bedingten Erwartung auch für beliebige nichtnegative bzw. integrierbare Zufallsvariablen:

Satz 1.19 (Existenz der bedingten Erwartung). Für jede Zufallsvariable $X \geq 0$ bzw. $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ und jede σ -Algebra $\mathcal{F} \subseteq \mathcal{A}$ existiert eine Version der bedingten Erwartung $E[X | \mathcal{F}]$.

Beweis. (1). Wir betrachten zunächst den Fall $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$. Wie eben bemerkt, ist der Raum $L^2(\Omega, \mathcal{F}, P)$ ein abgeschlossener Unterraum des Hilbertraums $L^2(\Omega, \mathcal{A}, P)$. Sei $d = \inf\{\|Z - X\|_{L^2} \mid Z \in L^2(\Omega, \mathcal{F}, P)\}$ der Abstand von X zu diesem Unterraum. Um zu zeigen, dass eine beste Approximation von X in $L^2(\Omega, \mathcal{F}, P)$ existiert, wählen wir eine Folge (X_n) aus diesem Unterraum mit $\|X_n - X\|_{L^2} \rightarrow d$. Mithilfe der Parallelogramm-Identität folgt für $n, m \in \mathbb{N}$:

$$\begin{aligned} \|X_n - X_m\|_{L^2}^2 &= \|(X_n - X) - (X_m - X)\|_{L^2}^2 \\ &= 2 \cdot \|X_n - X\|_{L^2}^2 + 2 \cdot \|X_m - X\|_{L^2}^2 - \|(X_n - X) + (X_m - X)\|_{L^2}^2 \\ &= 2 \cdot \underbrace{\|X_n - X\|_{L^2}^2}_{\rightarrow d^2} + 2 \cdot \underbrace{\|X_m - X\|_{L^2}^2}_{\rightarrow d^2} - 4 \underbrace{\left\| \frac{X_n + X_m}{2} - X \right\|_{L^2}^2}_{\leq d^2}, \end{aligned}$$

und damit

$$\limsup_{n, m \rightarrow \infty} \|X_n - X_m\|_{L^2}^2 \leq 0.$$

Also ist die Minimalfolge (X_n) eine Cauchyfolge in dem vollständigen Raum $L^2(\Omega, \mathcal{F}, P)$, d.h. es existiert ein $Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ mit

$$\|X_n - Y\|_{L^2} \rightarrow 0.$$

Für Y gilt

$$\|Y - X\|_{L^2} = \left\| \lim_{n \rightarrow \infty} X_n - X \right\|_{L^2} \leq \liminf_{n \rightarrow \infty} \|X_n - X\|_{L^2} \leq d,$$

d.h. Y ist die gesuchte Bestapproximation, und damit eine Version der bedingten Erwartung $E[X | \mathcal{F}]$.

- (2). Für eine beliebige nichtnegative Zufallsvariable X auf (Ω, \mathcal{A}, P) existiert eine monoton wachsende Folge (X_n) nichtnegativer quadratintegrierbarer Zufallsvariablen mit $X = \sup X_n$. Man verifiziert leicht, dass $\sup_n E[X_n | \mathcal{F}]$ eine Version von $E[X | \mathcal{F}]$ ist.
- (3). Entsprechend verifiziert man, dass für allgemeine $X \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ durch $E[X | \mathcal{F}] = E[X^+ | \mathcal{F}] - E[X^- | \mathcal{F}]$ eine Version der bedingten Erwartung gegeben ist.

□

Kapitel 2

Markovketten und Differenzgleichungen

In diesem Kapitel werden wir Markovketten genauer untersuchen. Ein wichtiges Hilfsmittel dabei ist der Zusammenhang von Markovketten und Differenzgleichungen.

2.1 Grundlagen

Sei (S, \mathcal{S}) ein messbarer Raum. Ein **(zeitdiskreter) stochastischer Prozess mit Zustandsraum S** ist eine Folge X_0, X_1, \dots von auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definierten Zufallsvariablen $X_n : \Omega \rightarrow S$. Den Index „ n “ interpretieren wir entsprechend als „Zeit.“ Für $m \leq n$ setzen wir:

$$X_{m:n} := (X_m, X_{m+1}, \dots, X_n).$$

Seien nun $p_n(x, dy), n = 1, 2, 3, \dots$, stochastische Kerne auf (S, \mathcal{S}) . Wir verwenden die Notation

$$(p_n f)(x) := \int p_n(x, dy) f(y)$$

für den Erwartungswert einer messbaren Funktion $f : S \rightarrow \mathbb{R}$ bzgl. der Wahrscheinlichkeitsverteilung $p_n(x, \bullet)$. Insbesondere gilt

$$(p_n I_A)(x) = p_n(x, A) \quad \text{für alle } A \in \mathcal{S}.$$

Definition (Markovkette). Ein stochastischer Prozess (X_n) mit Zustandsraum S heißt **Markovkette mit Übergangswahrscheinlichkeiten** $p_n(x, dy)$, falls gilt:

$$P[X_{n+1} \in A \mid X_{0:n}] = p_{n+1}(X_n, A) \quad P\text{-f.s. für alle } A \in \mathcal{S} \text{ und } n \geq 0, \quad (2.1.1)$$

bzw. dazu äquivalent

$$E[f(X_{n+1}) \mid X_{0:n}] = (p_{n+1} f)(X_n) \quad P\text{-f.s. für alle } \mathcal{S}\text{-messbaren } f : S \rightarrow \mathbb{R}_+ \text{ und } n \geq 0. \quad (2.1.2)$$

Die Markovkette heißt **zeitlich homogen**, falls p_n nicht von n abhängt. Die Verteilung von X_0 heißt **Startverteilung** der Markovkette. Gilt $P \circ X_0^{-1} = \delta_x$, dann sagen wir, die **Markovkette startet im Zustand x** .

Die Äquivalenz von (2.1.1) und (2.1.2) ergibt sich durch maßtheoretische Induktion. Die definierende Eigenschaft (2.1.1) besagt, dass, bedingt auf X_n , der nächste Zustand X_{n+1} unabhängig von X_0, \dots, X_{n-1} mit Verteilung $p_{n+1}(X_n, \bullet)$ ist. Eine Markovkette „vergisst“ also den vorherigen Verlauf bis zur Zeit $n - 1$, und startet in jedem Schritt neu im gegenwärtigen Zustand X_n .

Bemerkung (Allgemeine Definition einer Markovkette; Existenz von Übergangskernen). Allgemeiner heißt ein stochastischer Prozess (X_n) Markovkette, falls

$$P[X_{n+1} \in A \mid X_{0:n}] = P[X_{n+1} \in A \mid X_n] \quad P\text{-f.s. für alle } A \in \mathcal{S} \text{ und } n \geq 0 \quad (2.1.3)$$

gilt. Die Existenz eines Übergangskerns folgt aus (2.1.3) unter Regularitätsvoraussetzungen an (S, \mathcal{S}) . Hinreichend ist zum Beispiel, dass S ein polnischer (d.h. vollständiger separabler metrischer) Raum mit Borelscher σ -Algebra $\mathcal{S} = \mathcal{B}(S)$ ist.

Beispiel (Diskreter Zustandsraum). Ist S abzählbar, dann können wir einen stochastischen Kern p_n auf S mit der stochastischen Matrix $p_n(x, y) = p_n(x, \{y\})$ identifizieren. Ein stochastischer Prozess (X_n) ist genau dann eine Markovkette mit Übergangsmatrizen $p_n(x, y)$, wenn

$$P[X_{n+1} = x_{n+1} \mid X_{0:n} = x_{0:n}] = p_{n+1}(x_n, x_{n+1})$$

für alle $x_0, \dots, x_{n+1} \in S$ mit $P[X_{0:n} = x_{0:n}] \neq 0$ gilt.

2.1.1 Zufällige dynamische Systeme als Markovketten

Markovketten erhält man insbesondere als zufällige Störungen dynamischer Systeme:

Sei (T, \mathcal{T}) ein messbarer Raum. Wir betrachten einen stochastischen Prozess (X_n) mit Zustandsraum S , der rekursiv durch

$$X_{n+1} = \Phi_{n+1}(X_n, W_{n+1}), \quad n = 0, 1, 2, \dots,$$

definiert ist, wobei $X_0 : \Omega \rightarrow S$ und $W_1, W_2, \dots; \Omega \rightarrow T$ unabhängige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , und $\Phi : S \times T \rightarrow S, n \in \mathbb{N}$, messbare Abbildungen sind. Die Abbildungen Φ_n beschreiben das Bewegungsgesetz des dynamischen Systems, und die Zufallsvariablen W_n die zufälligen Einflussfaktoren (Rauschen, noise).

Satz 2.1. (1). (X_n) ist eine Markovkette mit Übergangswahrscheinlichkeiten

$$p_n(x, A) = P[\Phi_n(x, W_n) \in A], \quad x \in S, A \in \mathcal{S}.$$

(2). Hängen die Abbildungen Φ_n nicht von n ab, und sind die Zufallsvariablen W_n identisch verteilt, dann ist die Markovkette (X_n) zeitlich homogen.

Beweis. (1). Für $n \geq 0$ ist $X_{0:n}$ eine Funktion von $X_0, W_1, W_2, \dots, W_n$. Also ist W_{n+1} unabhängig von $X_{0:n}$, und für $A \in \mathcal{S}$ folgt

$$\begin{aligned} P[X_{n+1} \in A \mid X_{0:n}](\omega) &= P[\Phi_{n+1}(X_n, W_{n+1}) \in A \mid X_{0:n}](\omega) \\ &= P[\Phi_{n+1}(X_n(\omega), W_{n+1}) \in A] = p_{n+1}(X_n(\omega), A) \end{aligned}$$

für P -fast alle $\omega \in \Omega$.

(2). Hängen Φ_n und die Verteilung von W_n nicht von n ab, dann hängt auch p_n nicht von n ab, d.h. die Markovkette ist zeitlich homogen. □

Umgekehrt kann man Markovketten unter schwachen Voraussetzungen an den Zustandsraum stets als zufällige dynamische Systeme darstellen. Wir zeigen dies für reellwertige Markovketten, und erhalten so ein explizites Konstruktionsverfahren für diese Prozesse:

Seien μ eine Wahrscheinlichkeitsverteilung und $p_n, n \in \mathbb{N}$, stochastische Kerne auf $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Wir betrachten die linksstetigen Inversen

$$\begin{aligned} \underline{G}_0(u) &= \inf\{c \in \mathbb{R} : F_0(c) \geq u\} \quad \text{und} \\ \underline{G}_n(x, u) &= \inf\{c \in \mathbb{R} : F_n(x, c) \geq u\} \end{aligned}$$

der Verteilungsfunktionen $F_0(c) = \mu[(-\infty, c]]$ und $F_n(x, c) = p_n(x, (-\infty, c])$ der Wahrscheinlichkeitsverteilungen μ und $p_n(x, \bullet)$. Aus Satz 2.1 und Satz ?? folgt unmittelbar:

Korollar 2.2 (Existenzsatz und Konstruktionsverfahren für Markovketten). Sei U_0, U_1, U_2, \dots eine Folge von unabhängigen, auf $(0, 1)$ gleichverteilten Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Dann ist der durch

$$X_0 = \underline{G}_0(U_0), \quad X_{n+1} = \underline{G}_{n+1}(X_n, U_{n+1}),$$

definierte stochastische Prozess eine Markovkette mit Startverteilung μ und Übergangskernen p_n .

Bemerkung. Auch auf anderen Zustandsräumen kann man Markovketten oft auf ähnliche Weise explizit konstruieren.

Die obige Konstruktion liefert unmittelbar einen Algorithmus zur Simulation der Markovkette:

Algorithmus (Simulation einer reellwertigen Markovkette).

```

erzeuge  $u_0 \sim \text{Unif}(0, 1)$ ; setze  $y_0 := \underline{G}_0(u_0)$ 
for  $n := 1, 2, \dots$  do
    erzeuge  $u_n \sim \text{Unif}(0, 1)$ ; setze  $y_n := \underline{G}_n(y_{n-1}, u_n)$ 
end for

```

2.1.2 Beispiele

Wir betrachten eine Reihe von grundlegenden Beispielen von Markovketten, die wir explizit als zufällige dynamische System darstellen:

Random Walks auf \mathbb{Z}^d bzw. \mathbb{R}^d . Sind die Zufallsvariablen W_n unabhängig und identisch verteilt mit Werten in \mathbb{Z}^d oder \mathbb{R}^d , dann wird durch

$$X_0 = x, \quad X_{n+1} = X_n + W_{n+1},$$

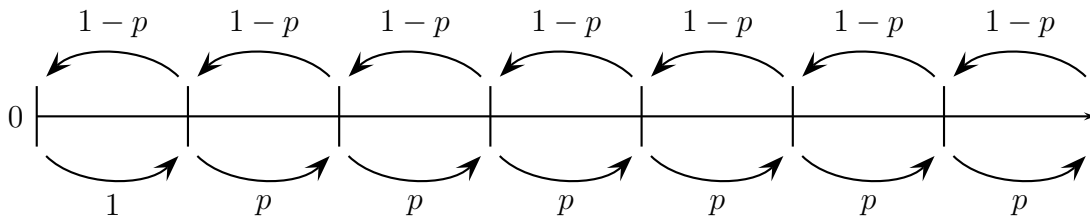
ein d -dimensionaler Random Walk definiert. Der Prozess (X_n) ist eine zeitlich homogene Markovkette mit Start in x und Übergangskern $p(x, \bullet) = \mu \circ \tau_x^{-1}$, wobei μ die Verteilung von W_n und $\tau_x(y) = y + x$ die Translation um x ist.

Random Walk auf $\{0, 1, 2, \dots\}$ mit Reflexion bzw. Absorption bei 0. Durch

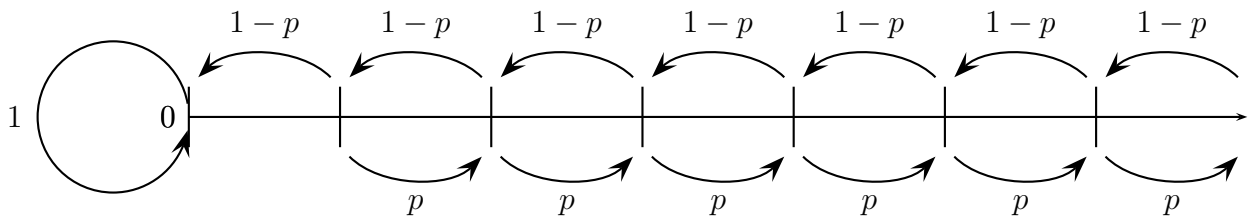
$$X_{n+1} = \begin{cases} X_n + W_{n+1} & \text{falls } X_n > 0, \\ 1 \text{ bzw. } 0 & \text{falls } X_n = 0, \end{cases}$$

mit unabhängigen Zufallsvariablen W_n mit $P[W_n = 1] = p$ und $P[W_n = -1] = 1 - p$, $p \in [0, 1]$, wird ein Random Walk auf \mathbb{Z}_+ definiert, der bei 0 reflektiert bzw. absorbiert wird. Der Prozess (X_n) ist eine zeitlich homogene Markovkette mit Übergangswahrscheinlichkeiten wie in Abbildung 2.1 dargestellt.

Random Walk mit Reflexion bei 0.



Random Walk mit Absorption bei 0.

Abbildung 2.1: Übergangswahrscheinlichkeiten von Random Walks auf $\{0, 1, 2, \dots\}$.

Warteschlange mit einem Server. In einem vereinfachten Warteschlangenmodell wird pro Zeiteinheit ein Kunde bedient, während A_n neue Kunden ankommen. Die Anzahlen A_n der Ankünfte in einer Bedienzeit sind unabhängige Zufallsvariablen mit Werten in $\{0, 1, 2, \dots\}$. Die Zahl X_n der wartenden Kunden ist dann eine Markovkette mit Übergangsmechanismus

$$X_{n+1} = (X_n - 1 + A_{n+1})_+.$$

Autoregressive Prozesse. Ein $AR(p)$ -Prozess mit Parametern $\varepsilon, \alpha_1, \dots, \alpha_p \in \mathbb{R}$ ist durch die Rekursionsformel

$$X_n = \sum_{i=1}^p \alpha_i X_{n-i} + \varepsilon \cdot W_n, \quad n \geq p,$$

mit unabhängigen, standardnormalverteilten Zufallsvariablen W_n gegeben. Für $p = 1$ ergibt sich eine zeithomogene Markovkette mit Übergangskern

$$p(x, \cdot) = N(\alpha_1 x, \varepsilon^2).$$

Für $p \geq 2$ und $\alpha_i, \varepsilon \neq 0$ ist der $AR(p)$ -Prozess dagegen keine Markovkette, da der nächste Zustand nicht nur vom gegenwärtigen Zustand, sondern auch vom vorherigen Verlauf abhängt.

Wir können jedoch eine Markovkette erhalten, indem wir statt X_n die aus den letzten p Zuständen gebildeten Vektoren

$$\bar{X}_n = (X_n, X_{n-1}, \dots, X_{n-p+1}), \quad n = p-1, p, p+1, \dots,$$

betrachten. (\bar{X}_n) ist eine zeithomogene Markovkette mit Zustandsraum S^p , denn für $n \geq p$ gilt

$$\bar{X}_n = \begin{pmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_p \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix} \bar{X}_{n-1} + \varepsilon \begin{pmatrix} W_n \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Galton-Watson-Verzweigungsprozesse. Der in Abschnitt 1.2 eingeführte Galton-Watson-Prozess ist eine zeithomogene Markovkette auf $S = \{0, 1, 2, \dots\}$, denn für $n \geq 0$ gilt

$$Z_n = \sum_{i=1}^{Z_{n-1}} N_i^n$$

mit unabhängigen, identisch verteilten Zufallsvariablen N_i^n ($i, n \in \mathbb{N}$). Als Übergangskern ergibt sich

$$p(k, \bullet) = P \circ \left(\sum_{i=1}^k N_i^n \right)^{-1} = \nu^{*k},$$

wobei ν^{*k} die k -fache Faltung der Nachkommensverteilung $\nu = P \circ (N_i^n)^{-1}$ ist.

Wright'sches Evolutionsmodell. In diesem Modell besteht die Population zu jedem Zeitpunkt n aus einer festen Anzahl m von Individuen, von denen jedes genau eines der Merkmale aus einer endlichen Menge T besitzt. Die Merkmale werden gemäß folgendem Mechanismus von einer Generation zur nächsten vererbt:

Algorithmus (Multinomiales Resampling).

for $i := 1, \dots, m$ **do**

erzeuge $w \sim \text{Unif}\{1, \dots, m\}$, setze $x_{n+1}^{(i)} := x_n^{(w)}$

end for

Jedes Individuum der Nachkommengeneration sucht sich also zufällig und unabhängig voneinander einen Vorfahren in der Elterngeneration, und nimmt dessen Merkmalsausprägung an. Durch den Algorithmus wird eine Markovkette (X_n) mit Zustandsraum T^m und Übergangskern

$$p(x, \bullet) = \bigotimes_{i=1}^m \hat{\mu}(x)$$

definiert, wobei $\hat{\mu}(x) = \frac{1}{m} \sum_{i=1}^m \delta_{x^{(i)}}$ die empirische Verteilung der Merkmalsausprägungen $x = (x^{(1)}, \dots, x^{(m)})$ in der vorherigen Population ist.

Anstatt die Merkmalsausprägungen $X_n^{(i)}$ aller Individuen einer Generation n zu betrachten („mikroskopische Beschreibung“), genügt es die Häufigkeiten

$$H_n(a) = |\{i \in \{1, \dots, m\} : X_n^{(i)} = a\}|, \quad a \in T,$$

aller möglichen Merkmalsausprägungen a zu notieren („makroskopische Beschreibung“). Die Histogrammvektoren $H_n = (H_n(a))_{a \in T}$ bilden eine zeithomogene Markovkette mit Werten im Raum $\text{Hist}(m, T)$ der Histogramme von m Beobachtungswerten aus T . Der Übergangskern ist durch

$$p(h, \bullet) = \text{Mult} \left(h \left/ \sum_{a \in T} h(a) \right. \right), \quad h \in \text{Hist}(m, T),$$

gegeben, d.h. der Histogrammvektor im nächsten Schritt ist multinomialverteilt mit Ergebniswahrscheinlichkeiten der Merkmalsausprägungen $a \in T$ proportional zu den Häufigkeiten $h(a)$ im letzten Schritt. Dies erklärt auch die Bezeichnung „Multinomiales Resampling.“ Multinomiale Resamplingschritte werden u.a. in genetischen Algorithmen und sequentiellen Monte-Carlo Verfahren eingesetzt, siehe zum Beispiel [Cappé, Moulines, Ryden: Inference in Hidden Markov Models].

2.1.3 Endlichdimensionale Randverteilungen

Wir wollen nun erste Verteilungen von Markovketten berechnen. Sei (X_n) ein auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definierter stochastischer Prozess mit Zustandsraum (S, \mathcal{S}) .

Satz 2.3. *Es sind äquivalent:*

- (1). (X_n) ist eine Markovkette mit Übergangswahrscheinlichkeiten p_n und Startverteilung μ .
- (2). Für jedes $n \geq 0$ hat (X_0, X_1, \dots, X_n) die Verteilung

$$\mu(dx_0) p_1(x_0, dx_1) p_2(x_1, dx_2) \cdot \dots \cdot p_n(x_{n-1}, dx_n),$$

d.h. für alle messbaren Funktionen $f : S^{n+1} \rightarrow \mathbb{R}_+$ gilt

$$E[f(X_0, \dots, X_n)] = \int \mu(dx_0) \int p_1(x_0, dx_1) \cdots \int p_n(x_{n-1}, dx_n) f(x_0, \dots, x_n). \quad (2.1.4)$$

Beweis. „(1) \Rightarrow (2)“: Ist (X_n) eine Markovkette mit Startverteilung μ und Übergangskernen p_n , dann gilt für $n \in \mathbb{N}$ und $B_0, \dots, B_n \in \mathcal{S}$:

$$\begin{aligned} P[X_{0:n} \in B_0 \times \dots \times B_n] &= E[P[X_n \in B_n \mid X_{0:n-1}]; X_{0:n-1} \in B_0 \times \dots \times B_{n-1}] \\ &= \int_{B_0 \times \dots \times B_{n-1}} p(x_{n-1}, B_n) \mu_{X_{0:n-1}}(dx_{0:n-1}). \end{aligned}$$

Durch Induktion nach n folgt

$$P[X_{0:n} \in B_0 \times \dots \times B_n] = \int_{B_n} \dots \int_{B_1} \int_{B_0} \mu(dx_0) p_1(x_0, dx_1) \cdot \dots \cdot p_n(x_{n-1}, dx_n)$$

für alle $n \geq 0$ und $B_i \in \mathcal{S}$. Also gilt $X_{0:n} \sim \mu \otimes p_1 \otimes \dots \otimes p_n$, und damit (2.1.4).

„(2) \Rightarrow (1)“: Gilt (2.1.4), dann hat X_0 die Verteilung μ , und $(p_{n+1}f)(X_n)$ ist für alle messbaren Funktionen $f : S \rightarrow [0, \infty)$ eine Version der bedingten Erwartung $E[f(X_{n+1}) \mid X_{0:n}]$. Zum Beweis überprüfen wir die definierenden Eigenschaften der bedingten Erwartung: $(p_{n+1}f)(X_n)$ ist eine messbare Funktion von $X_{0:n}$, und es gilt

$$\begin{aligned} &E[f(X_{n+1}) \cdot g(X_{0:n})] \\ &= \int \mu(dx_0) \int p_1(x_0, dx_1) \cdot \dots \cdot \int p_n(x_{n-1}, dx_n) g(x_{0:n}) \int p_{n+1}(x_n, dx_{n+1}) f(x_{n+1}) \\ &= \int \mu(dx_0) \int p_1(x_0, dx_1) \cdot \dots \cdot \int p_n(x_{n-1}, dx_n) g(x_{0:n}) (p_{n+1}f)(x_n) \\ &= E[(p_{n+1}f)(X_n) \cdot g(X_{0:n})] \end{aligned}$$

für alle messbaren Funktionen $g : S^{n+1} \rightarrow [0, \infty)$. □

Anwendung von stochastischen Kernen auf Maße und Funktionen. Zum Abschluss dieses Abschnitts führen wir Notationen und Rechenregeln für stochastische Kerne auf allgemeinen Zustandsräumen ein, die mit der Notation im diskreten Fall konsistent sind. Seien μ eine Wahrscheinlichkeitsverteilung, p, q, r stochastische Kerne, und f eine messbare nicht-negative Funktion auf (S, \mathcal{S}) . Wir bezeichnen mit

$$(\mu p)(dy) = \int \mu(dx) p(x, dy)$$

die Verteilung der 2. Komponente unter dem Maß $\mu \otimes p$, und mit

$$(pq)(x, dz) = \int p(x, dy) q(y, dz)$$

den stochastischen Kern, der durch Hintereinanderausführen von p und q entsteht. Aus dem Satz von Fubini ergeben sich die folgenden *Rechenregeln für stochastische Kerne*:

$$\int f d(\mu p) = \int \int \mu(dx) p(x, dy) f(y) = \int (pf) d\mu \quad (2.1.5)$$

$$p(qf) = (pq)f \quad (2.1.6)$$

$$(\mu p)q = \mu(pq) \quad (2.1.7)$$

$$p(qr) = (pq)r \quad (2.1.8)$$

Als Verteilung der Markovkette zur Zeit n erhalten wir dementsprechend

$$P \circ X_n^{-1} = \mu p_1 p_2 \cdot \dots \cdot p_n, \quad (2.1.9)$$

wobei das Produkt wegen (2.1.7) und (2.1.8) nicht von der Klammerung abhängt.

Ist der Zustandsraum S abzählbar, dann gelten die folgenden Identifikationen:

μ	\leftrightarrow	$(\mu(x) : x \in S)$	Zeilenvektor
f	\leftrightarrow	$(f(x) : x \in S)$	Spaltenvektor
p	\leftrightarrow	$(p(x, y) : x, y \in S)$	stochastische Matrix
$(\mu p)(y)$	$=$	$\sum_x \mu(x) p(x, y)$	Multiplikation mit Zeilenvektor von links
$(pf)(x)$	$=$	$\sum_y p(x, y) f(y)$	Multiplikation mit Spaltenvektor von rechts
$(pq)(x, z)$	$=$	$\sum_y p(x, y) q(y, z)$	Matrizenprodukt.

2.2 Gleichgewichte

Die Verteilung vieler zeithomogener Markovketten nähert sich für große Zeiten einem Gleichgewicht an, d.h. einer Wahrscheinlichkeitsverteilung, die bei Anwendung des Übergangskerns invariant ist. In diesem Abschnitt zeigen wir, wie man Gleichgewichte in einfachen Fällen berechnen kann, und stellen zwei grundlegende Verfahren zum Nachweis der Konvergenz ins Gleichgewicht vor.

2.2.1 Definitionen und Beispiele

Definition (Gleichgewicht, Detailed Balance Bedingung). Sei p ein stochastischer Kern auf einem meßbaren Raum (S, \mathcal{S}) .

- (1). Ein Wahrscheinlichkeitsmaß μ auf (S, \mathcal{S}) heißt **stationäre Verteilung** oder **Gleichgewicht** von p falls $\mu p = \mu$ gilt, d.h. falls

$$\int \mu(dx) p(x, B) = \mu[B] \quad \text{für alle } B \in \mathcal{S}. \quad (2.2.1)$$

- (2). μ erfüllt die **Detailed Balance Bedingung** bzgl. p , falls $\mu(dx)p(x, dy) = \mu(dy)p(y, dx)$ gilt, d.h. falls

$$\int_A \mu(dx) p(x, B) = \int_B \mu(dy) p(y, A) \quad \text{für alle } A, B \in \mathcal{S}. \quad (2.2.2)$$

Offensichtlich folgt aus der Detailed Balance Bedingung, dass μ ein Gleichgewicht von p ist. Zeithomogene Markovketten mit Übergangskern p haben stationäre Randverteilungen, wenn sie in einem Gleichgewicht des Übergangskerns starten:

Korollar 2.4 (Stationarität von Markovketten). Für eine zeithomogene Markovkette $(X_n)_{n \in \mathbb{Z}_+}$ mit Startverteilung μ und Übergangskern p sind die folgenden Aussagen äquivalent:

- (1). μ ist ein Gleichgewicht von p .
- (2). $X_1 \sim \mu$.
- (3). $X_n \sim \mu$ für alle $n \geq 0$.

Beweis. Das Wahrscheinlichkeitsmaß μ ist genau dann ein Gleichgewicht von p , wenn $\mu p = \mu$ bzw. $\mu p^n = \mu$ für alle $n \geq 0$ gilt. Die Aussage folgt, da μp^n die Verteilung von X_n ist. \square

Auf ähnliche Weise zeigt man, dass die Detailed Balance Bedingung genau dann erfüllt ist, wenn $(X_0, X_1) \sim (X_1, X_0)$ bzw. $(X_0, X_1, \dots, X_n) \sim (X_n, X_{n-1}, \dots, X_0)$ für alle $n \geq 0$ gilt.

Bemerkung (Absolutstetiger Fall). In vielen praktisch relevanten Fällen haben das Maß μ und der Übergangskern p Dichten bzgl. eines geeigneten Referenzmaßes ν , z.B. bzgl. des Lebesgue-Maßes oder des Zählmaßes:

$$\mu(dx) = \mu(x) \nu(dx), \quad p(x, dy) = p(x, y) \nu(dy).$$

Dann lauten die Gleichgewichts- und die Detailed Balance Bedingung:

$$\int \mu(x) p(x, y) \nu(dx) = \mu(y) \quad \text{für } \nu\text{-fast alle } y, \quad (2.2.3)$$

$$\mu(x) p(x, y) = \mu(y) p(y, x) \quad \text{für } \nu\text{-fast alle } x, y. \quad (2.2.4)$$

Im zeitdiskreten Fall erhalten wir die uns schon bekannte Gleichgewichtsbedingung

$$\sum_x \mu(x) p(x, y) = \mu(y) \quad \text{für alle } y. \quad (2.2.5)$$

Für Anwendungen ergeben sich zwei zentrale Fragen:

- Wie berechnet man die stationären Verteilungen zu einem gegebenen Übergangskern ?
- Wie findet man zu einem gegebenen Wahrscheinlichkeitsmaß Übergangskerne, für die das Maß ein Gleichgewicht ist ?

Während sich die zweite (z.B. für Markov Chain Monte Carlo Verfahren zentrale) Frage oft leicht mithilfe der Detailed Balance Bedingung beantworten lässt, ist die Berechnung der stationären Verteilungen in der Regel nicht einfach. Wir betrachten jeweils ein Beispiel für beide Fragestellungen:

Beispiel (Eindimensionale Markovkette, Birth-Death-Prozess). Seien $p_x, q_x, r_x > 0$ mit $p_x + q_x + r_x = 1, q_0 = 0$, und $p_x, q_x > 0$ für alle $x \geq 1$. Ein (zeitdiskreter) Birth-Death-Prozess ist eine zeithomogene Markovkette auf $S = \{0, 1, 2, \dots\}$ mit Übergangswahrscheinlichkeiten

$$p(x, x+1) = p_x, \quad p(x, x-1) = q_x, \quad p(x, x) = r_x,$$



Das Gleichungssystem für eine Gleichgewichtsverteilung μ lautet

$$\begin{aligned} \mu(0) \cdot r_0 + \mu(1) \cdot q_1 &= \mu(0), \\ \mu(x-1) \cdot p_{x-1} + \mu(x) \cdot r_x + \mu(x+1) \cdot q_{x+1} &= \mu(x) \quad \text{für } x \in \mathbb{N}. \end{aligned}$$

Da die Lösung sich rekursiv aus $\mu(0)$ berechnen lässt, ist der Lösungsvektorraum des linearen Gleichungssystems eindimensional. Aus der hinreichenden Detailed-Balance-Bedingung

$$\mu(x-1) \cdot p_{x-1} = \mu(x) \cdot q_x \quad \text{für alle } x \in \mathbb{N} \quad (2.2.6)$$

erhalten wir daher in diesem Fall bereits die allgemeine Lösung

$$\mu(x) = \mu(0) \cdot \frac{p_0 \cdot p_1 \cdot \dots \cdot p_{x-1}}{q_1 \cdot q_2 \cdot \dots \cdot q_x}. \quad (2.2.7)$$

Sei

$$Z = \sum_{x=0}^{\infty} \frac{p_0 \cdot p_1 \cdot \dots \cdot p_{x-1}}{q_1 \cdot q_2 \cdot \dots \cdot q_x}.$$

Gilt $Z < \infty$, dann ist durch (2.2.7) mit $\mu(0) = 1/Z$ das eindeutige Gleichgewicht der Markovkette gegeben. Die Bedingung $Z < \infty$ bedeutet, dass die Wachstumswahrscheinlichkeiten $p(x-1, x)$ nicht zu groß im Vergleich zu den Abfallwahrscheinlichkeiten $p(x, x-1)$ sind. Gilt dagegen $Z = \infty$, dann existiert keine Gleichgewichtsverteilung.

Beispielsweise können wir ein vereinfachtes **Warteschlangenmodell**, bei dem in jeder Zeiteinheit mit Wahrscheinlichkeit a genau ein Kunde kommt, und mit Wahrscheinlichkeit b genau ein Kunde bedient wird, unter Unabhängigkeitsannahmen als Birth-Death-Prozess mit Übergangswahrscheinlichkeiten $q_x = (1-a)b$ und $p_x = (1-b)a$ für $x \geq 1$, sowie $p_0 = a$ darstellen. Für $a < b$ ergibt sich dann als eindeutiges Gleichgewicht eine geometrische Verteilung mit Parameter $\frac{(1-b)a}{(1-a)b}$, während für $a \geq b$ kein Gleichgewicht existiert.

Bemerkung (Gleichgewichtsbedingung als Differenzgleichung). Im letzten Beispiel lässt sich die Gleichgewichtsbedingung (2.2.5) auch als lineare Differenzgleichung schreiben. Zum Beispiel ist sie im Warteschlangenmodell äquivalent zu

$$q \cdot (\mu(x+1) - \mu(x)) + p \cdot (\mu(x-1) - \mu(x)) = 0, \quad \text{bzw.}$$

$$p \cdot [(\mu(x+1) - \mu(x)) - (\mu(x) - \mu(x-1))] + (q-p) \cdot (\mu(x+1) - \mu(x)) = 0.$$

Beispiel (Random Walk Metropolis). Sei $\mu(dx)$ ein absolutstetiges Wahrscheinlichkeitsmaß auf \mathbb{R}^d . Wir bezeichnen die Dichte bzgl. des Lebesgue-Maßes ebenfalls mit $\mu(x)$. Um näherungsweise Stichproben von μ zu erzeugen, konstruieren wir eine Markovkette mit Gleichgewicht μ durch Modifikation eines Random Walks auf \mathbb{R}^d . Seien

$$q(x, dy) = q(x, y) \lambda^d(dy)$$

die absolutstetigen Übergangswahrscheinlichkeiten des Random Walks. Wir nehmen an, dass

$$q(x, y) = q(y, x) \quad \text{für alle } x, y \in \mathbb{R}^d$$

gilt. Akzeptieren wir die Übergänge des Random Walks von x nach y mit Wahrscheinlichkeit $\alpha(x, y) = \min\left(\frac{\mu(y)}{\mu(x)}, 1\right)$, und verwerfen wir die Übergänge anderenfalls, dann erhalten wir eine Markovkette mit Übergangskern

$$p(x, dy) = q(x, y) \lambda^d(dy) + r(x) \delta_x(dy),$$

$$p(x, y) = q(x, y) \cdot \min\left(\frac{\mu(y)}{\mu(x)}, 1\right), \quad \text{und} \quad r(x) = 1 - \int p(x, y) \lambda^d(dy).$$

Diese Markovkette erfüllt die Detailed Balance Bedingung bzgl. μ , da

$$\mu(x) p(x, y) = \min(\mu(x), \mu(y)) q(x, y) = \mu(y) p(y, x)$$

für alle $x, y \in \mathbb{R}^d$ mit $x \neq y$ gilt.

2.2.2 Eigenwerte und Konvergenz ins Gleichgewicht

Wir betrachten nun einen endlichen Zustandsraum S mit k Elementen, und eine stochastische Matrix p , die nicht von n abhängt. Die Verteilung zur Zeit n einer zeithomogenen Markovkette mit Startverteilung μ und Übergangsmatrix p ist dann

$$P \circ X_n^{-1} = \mu p^n.$$

Um die Verteilung und deren Asymptotik zu berechnen, können wir die Spektraldarstellung der Übergangsmatrix verwenden. Seien $\lambda_1, \dots, \lambda_k \in \mathbb{C}$ die Eigenwerte von p , d.h. die Nullstellen des charakteristischen Polynoms $\chi(\lambda) = \det(p - \lambda I)$. Da p eine stochastische Matrix ist, gilt Folgendes:

- (1). $|\lambda_j| \leq 1$ für alle j . (Dies folgt wegen $\|pf\|_\infty = \max_x \left| \sum_y p(x, y) f(y) \right| \leq \|f\|_\infty$ für alle f)
- (2). $\lambda_1 = 1$ ist Eigenwert mit Rechtseigenvektor $f_1 = (1, \dots, 1)^T$.
- (3). Nichtreelle Eigenwerte treten in Paaren $\lambda, \bar{\lambda}$ auf.

Wir nehmen nun der Einfachheit halber an, dass alle Eigenwerte einfach sind, d.h. $\lambda_i \neq \lambda_j$ für $i \neq j$. In diesem Fall existieren Rechts- und Linkseigenvektoren f_j, ν_j ($1 \leq j \leq k$) mit

$$pf_j = \lambda_j f_j, \quad \nu_j p = \lambda_j \nu_j, \quad \text{und} \quad \langle \nu_i, f_j \rangle = \sum_{x \in S} \nu_i(x) f_j(x) = \delta_{ij}.$$

Mithilfe der aus den Rechts- und Linkseigenvektoren gebildeten Matrizen

$$U = (f_1, \dots, f_k), \quad V = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_k \end{pmatrix}, \quad V \cdot U = I,$$

erhalten wir die Spektraldarstellung

$$p = \sum_{j=1}^k \lambda_j f_j \otimes \nu_j = U \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_k \end{pmatrix} V,$$

für die Übergangsmatrix p , und damit auch für p^n :

$$p^n = \sum_{j=1}^k \lambda_j^n f_j \otimes \nu_j = U \begin{pmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_k^n \end{pmatrix} V.$$

Für die Verteilung der Markovkette zur Zeit n ergibt sich

$$P \circ X_n^{-1} = \mu p^n = \sum_{j=1}^k \lambda_j^n \langle \mu, f_j \rangle \nu_j. \quad (2.2.8)$$

Insbesondere folgt:

Satz 2.5 (Exponentielle Konvergenz ins Gleichgewicht). *Sind die Eigenwerte $\lambda_1, \dots, \lambda_k$ einer stochastischen Matrix $p \in \mathbb{R}^{k \times k}$ einfach, und gilt $|\lambda_j| < 1$ für alle $j \neq 1$, dann existiert eine Gleichgewichtsverteilung ν von p , und für jede Startverteilung μ gilt für $n \rightarrow \infty$:*

$$\mu p^n = \nu + O\left(\max_{j \neq 1} |\lambda_j|^n\right).$$

Beweis. Nach (2.2.8) gilt

$$\mu p^n = \langle \mu, f_1 \rangle \nu_1 + \sum_{j=2}^k \lambda_j^n \langle \mu, f_j \rangle \nu_j \quad \text{für alle } n \geq 0.$$

Aus $\langle \mu, f_1 \rangle = \langle \mu, (1, \dots, 1)^T \rangle = \sum \mu(x) = 1$ folgt

$$\mu p^n = \nu_1 + O\left(\max_{j \neq 1} |\lambda_j|^n\right)$$

Insbesondere ist $\nu_1 = \lim \mu p^n$ eine Wahrscheinlichkeitsverteilung mit $\nu_1 p = \nu_1$, also ein Gleichgewicht von p . Ist umgekehrt μ ein beliebiges Gleichgewicht von p , dann gilt $\mu p^n = \mu$ für alle $n \geq 0$, und damit

$$\mu = \lim_{n \rightarrow \infty} \mu p^n = \nu_1.$$

□

Bemerkung. (1). Sind die Eigenwerte nicht einfach, dann folgt eine ähnliche Aussage über die Jordansche Normalformdarstellung der Übergangsmatrix p . Als Konvergenzgeschwindigkeit ergibt sich in diesem Fall $O(n^{m-1} \max_{j \neq 1} |\lambda_j|^n)$, wobei m die größte Multiplizität des betragsmäßig zweitgrößten Eigenwertes ist (Satz von Perron-Frobenius).

(2). Entscheidend für die exponentielle Konvergenzrate ist die „Spektrallücke“ zwischen dem Eigenwert 1 und dem maximalen Betrag der übrigen Eigenwerte. Eine entsprechende Aussage kann man auch auf allgemeinen Zustandsräumen mithilfe des Spektralsatzes für selbstadjungierte Operatoren zeigen, falls die Gleichgewichtsverteilung die Detailed Balance Bedingung erfüllt.

Beispiele. (1). Die Übergangsmatrix der Markovkette aus Abbildung 2.2 ist

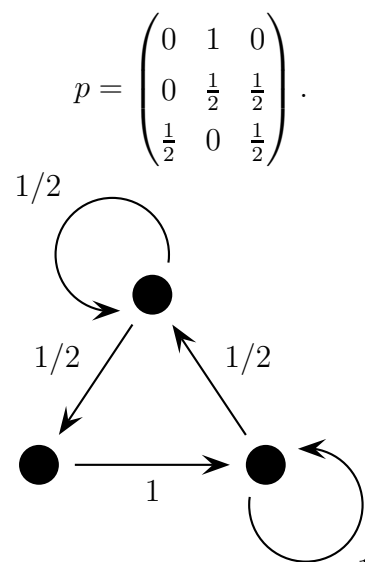


Abbildung 2.2: Übergangswahrscheinlichkeiten

Eigenwerte sind $\lambda_1 = 1$, $\lambda_2 = i/2$ und $\lambda_3 = -i/2$. Es folgt:

$$p^n = A + B \cdot \left(\frac{i}{2}\right)^n + C \cdot \left(-\frac{i}{2}\right)^n$$

mit Matrizen $A, B, C \in \mathbb{C}^{3 \times 3}$. Wegen $p^0(x_1, x_1) = 1$ und $p^1(x_1, x_1) = p^2(x_1, x_1) = 0$ folgt daraus

$$p^n(x_1, x_1) = \frac{1}{5} + \left(\frac{1}{2}\right)^n \cdot \left(\frac{4}{5} \cos \frac{n\pi}{2} - \frac{2}{5} \sin \frac{n\pi}{2}\right) \quad \text{für alle } n \geq 0.$$

Der Wert $1/5$ ist die erste Komponente des Gleichgewichtsvektors $\nu_1 = (1/5, 2/5, 2/5)$. Für $n \rightarrow \infty$ konvergieren die Übergangswahrscheinlichkeiten mit Rate $O(2^{-n})$ gegen ν_1 .

(2). Die Übergangsmatrix einer deterministischen Rotation auf dem diskreten Kreis $\mathbb{Z}/k\mathbb{Z}$ ist

$$p = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Das charakteristische Polynom ist $\chi(\lambda) = (-1)^k \cdot (\lambda^k - 1)$, und die Eigenwerte von p sind dementsprechend die k -ten Einheitswurzeln $\lambda_j = \exp(2\pi i \cdot (j - 1)/k)$, $j = 1, \dots, k$. Da alle Eigenwerte Betrag 1 haben, gilt keine exponentielle Konvergenz ins Gleichgewicht. Tatsächlich ist die Markovkette mit Übergangsmatrix p periodisch: Es gilt $X_{n+mk} = X_n$ P -fast sicher für alle $n, m \geq 0$.

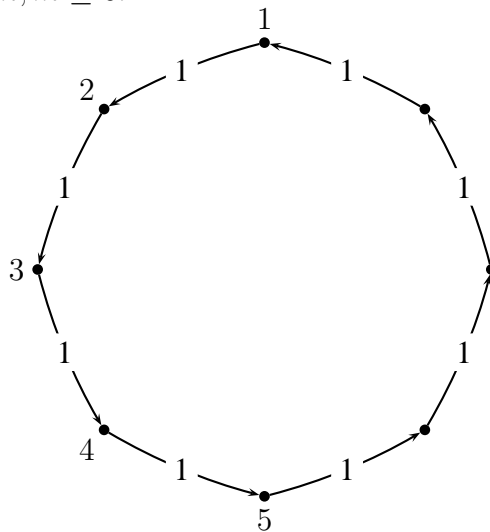


Abbildung 2.3: Deterministische Rotation auf $\mathbb{Z}/k\mathbb{Z}$.

(3). Für den symmetrischen Random Walk auf $\mathbb{Z}/k\mathbb{Z}$ gilt $p(x, x + 1) = p(x, x - 1) = 1/2$. Das lineare Gleichungssystem

$$\frac{1}{2}f(x + 1) + \frac{1}{2}f(x - 1) = \lambda f(x), \quad x \in \mathbb{Z}/k\mathbb{Z},$$

für die Rechtseigenvektoren löst man mit dem Ansatz $f(x) = \exp(itx)$. Für $t = 2\pi n/k$ mit $n = 0, 1, \dots, k - 1$ ist f periodisch mit Periode k . Wir erhalten so n linear unabhängige Eigenvektoren f_n zu den Eigenwerten

$$\lambda_n = \cos(2\pi n/k), \quad n = 0, 1, \dots, k - 1.$$

Der konstante Eigenvektor zum Eigenwert 1 ist nun f_0 . Ist k gerade, dann ist $f_{k/2}$ ein Eigenvektor zum Eigenwert -1 . In diesem Fall gilt keine exponentielle Konvergenz ins Gleichgewicht. In der Tat ist der Random Walk für gerade k periodisch mit Periode 2, und das Gleichgewicht ist nicht eindeutig. Ist k dagegen ungerade, dann gilt

$$|\lambda_n| \leq \cos(\pi/k) < 1 \quad \text{für alle } n \neq 0.$$

In diesem Fall folgt exponentielle Konvergenz ins Gleichgewicht (Gleichverteilung). Wegen $\cos(\pi/k) \sim 1 - \pi^2/k^2$ fällt die exponentielle Abfallrate für $k \rightarrow \infty$ mit der Ordnung $O(k^{-2})$ ab. Diese Ordnung entspricht einem diffusiven Verhalten.

2.2.3 Kopplungen und Kontraktivität

Ein anderer Ansatz um Konvergenz ins Gleichgewicht zu zeigen und zu quantifizieren basiert auf Kontraktionseigenschaften der Übergangskerne bezüglich geeigneter Metriken auf dem Raum der Wahrscheinlichkeitsmaße. Diese weist man in der Regel mithilfe von Kopplungsmethoden nach.

Wir setzen in diesem Abschnitt voraus, dass (S, d) ein vollständiger separabler metrischer Raum ist, den wir mit der Borelschen σ -Algebra $\mathcal{S} = \mathcal{B}(S)$ versehen. Sei $x_0 \in S$ ein fester Punkt. Wir bezeichnen die Menge aller Wahrscheinlichkeitsmaße auf $\mathcal{B}(S)$ mit $\mathcal{P}(S)$, und setzen

$$\mathcal{P}^1(S) = \left\{ \mu \in \mathcal{P}(S) : \int d(x_0, y) \mu(dy) < \infty \right\}.$$

Diese Definition hängt nicht von x_0 ab, falls der Raum S zusammenhängend ist.

Definition (Kopplung, Kantorovich-Metrik). Seien $\mu, \nu \in \mathcal{P}(S)$.

- (1). Eine **Kopplung** von μ und ν ist gegeben durch ein Wahrscheinlichkeitsmaß $\eta \in \mathcal{P}(S \times S)$ mit Randverteilungen μ und ν , bzw. durch Zufallsvariablen X, Y auf einem gemeinsamen Wahrscheinlichkeitsraum mit $X \sim \mu$ und $Y \sim \nu$.
- (2). Die **Kantorovich-Metrik (Transportmetrik, L^1 Wasserstein-Metrik)** von μ und ν ist definiert als

$$\mathcal{W}^1(\mu, \nu) = \inf_{\eta \in \Pi(\mu, \nu)} \int d(x, y) \eta(dx dy) = \inf_{X \sim \mu, Y \sim \nu} E[d(X, Y)],$$

wobei $\Pi(\mu, \nu)$ die Menge aller Kopplungen von μ und ν bezeichnet, und das zweite Infimum über alle Zufallsvariablen (X, Y) geht, welche die Kopplung realisieren.

Bemerkung (Eigenschaften der Kantorovich-Metrik). (1). Ist S zusammenhängend, dann gilt

$$\mathcal{W}^1(\mu, \nu) \leq \mathcal{W}^1(\mu, \delta_{x_0}) + \mathcal{W}^1(\delta_{x_0}, \nu) \quad \text{für alle } \mu, \nu \in \mathcal{P}^1(S).$$

(2). Die Kantorovich-Metrik $\mathcal{W}^1(\delta_x, \delta_y)$ zweier Dirac-Maße stimmt mit dem Abstand $d(x, y)$ der Basispunkte überein. In diesem Sinn ist S isometrisch in den Raum $\mathcal{P}^1(S)$ mit Kantorovich-Metrik eingebettet.

(3). Als Spezialfall der Kantorovich-Metrik ergibt sich die totale Variationsdistanz, wenn die Metrik auf dem Grundraum durch $d(x, y) = I_{S \setminus \{x\}}(y)$ gegeben ist.

Die **Kantorovich-Rubinstein-Dualität** besagt, dass sich die Kantorovich-Metrik als Supremum der Differenzen von Integralen 1-Lipschitzstetiger Funktionen auf S darstellen lässt:

$$\mathcal{W}^1(\mu, \nu) = \sup \left\{ \int f d\mu - \int f d\nu : f : S \rightarrow \mathbb{R} \text{ mit } |f(x) - f(y)| \leq d(x, y) \right\}.$$

Den Beweis dieser Aussage und des folgenden Satzes findet man zum Beispiel im Buch von C. Villani [Villani: Optimal Transport - old and new].

Satz (Eigenschaften der Kantorovich-Metrik). *Ist S ein zusammenhängender vollständiger separabler metrischer Raum, dann gilt:*

- (1). *Der Raum $(\mathcal{P}^1(S), \mathcal{W}^1)$ ist wieder ein vollständiger, separabler metrischer Raum.*
- (2). *Eine Folge (μ_k) von Wahrscheinlichkeitsmaßen in $\mathcal{P}^1(S)$ konvergiert genau dann bezüglich der Kantorovich-Metrik gegen ein Maß $\mu \in \mathcal{P}^1(S)$, wenn*

$$\lim_{k \rightarrow \infty} \int f d\mu_k = \int f d\mu$$

für jede stetige Funktion $f : S \rightarrow \mathbb{R}$ mit $f(x) \leq \text{const.} \cdot (1 + d(x, x_0))$ gilt.

Bevor wir ein allgemeines Resultat formulieren, zeigen wir anhand des AR(1) Modells, wie Kopplungsverfahren verwendet werden können, um die Konvergenzgeschwindigkeit ins Gleichgewicht zu quantifizieren.

Beispiel (AR(1) Modell). Seien $\alpha, \varepsilon \in \mathbb{R}$ und $\mu \in \mathcal{P}^1(\mathbb{R})$. Wir betrachten die durch

$$X_n = \alpha X_{n-1} + \varepsilon W_n \tag{2.2.9}$$

mit unabhängigen Zufallsvariablen $X_0 \sim \mu$, $W_n \sim N(0, 1)$, definierte Markovkette auf \mathbb{R} . Ist $X_{n-1} \sim N(m, v)$, dann folgt $X_n \sim N(\alpha m, \alpha^2 v + \varepsilon^2)$. Somit ist $\bar{\mu} = N(0, \varepsilon^2/(1 - \alpha^2))$

ein Gleichgewicht. Um Konvergenz ins Gleichgewicht zu zeigen, realisieren wir auf demselben Wahrscheinlichkeitsraum eine von den W_n unabhängige Zufallsvariable $\bar{X}_0 \sim \bar{\mu}$ (ggf. müssen wir dazu den Wahrscheinlichkeitsraum erweitern), und erhalten daraus über (2.2.9) eine Realisierung (\bar{X}_n) der Markovkette mit Startverteilung $\bar{\mu}$. Man beachte, dass (X_n) und (\bar{X}_n) mit denselben Zufallsvariablen W_n definiert sind. Da $\bar{\mu}$ ein Gleichgewicht ist, gilt $\bar{X}_n \sim \bar{\mu}$ für alle n , d.h. (X_n, \bar{X}_n) ist eine Kopplung der Maße μp^n und $\bar{\mu}$. Durch Induktion erhalten wir für $n \in \mathbb{N}$:

$$X_n - \bar{X}_n = \alpha^n (X_0 - \bar{X}_0),$$

und damit

$$\mathcal{W}^1(\mu p^n \bar{\mu}) \leq E[|X_n - \bar{X}_n|] = \alpha^n E[|X_0 - \bar{X}_0|].$$

Satz 2.6 (Kontraktivität und Konvergenz ins Gleichgewicht in der Kantorovich-Metrik).

Sei $\alpha \in (0, 1)$. Für alle $\mu, \nu \in \mathcal{P}^1(S)$ und für jede Kopplung (X, Y) von μ und ν existiere eine Kopplung (X', Y') von μp und νp mit

$$E[d(X', Y')] \leq \alpha E[d(X, Y)]. \quad (2.2.10)$$

Dann hat der Übergangskern p ein eindeutiges Gleichgewicht $\bar{\mu} \in \mathcal{P}^1(S)$, und

$$\mathcal{W}^1(\mu p^n, \bar{\mu}) \leq \alpha^n \mathcal{W}^1(\mu, \bar{\mu}) \quad \text{für alle } \mu \in \mathcal{P}^1(S).$$

Beweis. Seien $\mu, \nu \in \mathcal{P}^1(S)$, und $(X, Y), (X', Y')$ wie in der Voraussetzung. Dann gilt

$$\mathcal{W}^1(\mu p, \nu p) \leq E[d(X', Y')] \leq \alpha E[d(X, Y)].$$

Optimieren über alle Kopplungen (X, Y) von μ und ν liefert

$$\mathcal{W}^1(\mu p, \nu p) \leq \alpha \mathcal{W}^1(\mu, \nu),$$

und, durch Induktion,

$$\mathcal{W}^1(\mu p^n, \nu p^n) \leq \alpha^n \mathcal{W}^1(\mu, \nu) \quad \text{für alle } n \in \mathbb{N}.$$

Mithilfe der Dreiecksungleichung folgt für eine beliebige Startverteilung $\mu \in \mathcal{P}^1(S)$:

$$\mathcal{W}^1(\mu p^n, \mu p^{n+m}) \leq \alpha^n \mathcal{W}^1(\mu, \mu p^m) \leq \alpha^n \sum_{i=0}^{m-1} \alpha^i \mathcal{W}^1(\mu, \mu p).$$

Da die rechte Seite für $n \rightarrow \infty$ gleichmäßig in m gegen Null geht, und der Raum $\mathcal{P}^1(S)$ bzgl. der Kantorovich-Metrik vollständig ist, folgt die Existenz eines Grenzwertes $\bar{\mu} = \lim \mu p^n$ in diesem Raum. Offensichtlich gilt $\bar{\mu} = \bar{\mu} p$. Zudem folgt $\tilde{\mu} = \bar{\mu}$ für jedes Gleichgewicht $\tilde{\mu} \in \mathcal{P}^1(S)$ wegen

$$\mathcal{W}^1(\tilde{\mu}, \bar{\mu}) = \mathcal{W}^1(\tilde{\mu} p^n, \bar{\mu} p^n) \leq \alpha^n \mathcal{W}^1(\tilde{\mu}, \bar{\mu}) \rightarrow 0 \quad \text{für } n \rightarrow \infty.$$

□

Die Anwendung von Satz 2.6 bezüglich der totalen Variationsdistanz (d.h. bezüglich der zugrundeliegenden Metrik $d(x, y) = I_{\{x \neq y\}}$ auf S) liefert ein klassisches Resultat von Doeblin:

Korollar 2.7 (Konvergenz ins Gleichgewicht bei Minorisierungsbedingung). *Es existiere ein $\varepsilon > 0$ sowie ein Wahrscheinlichkeitsmaß $\pi \in \mathcal{P}(S)$ mit*

$$p(x, B) \geq \varepsilon \pi[B] \quad \text{für alle } x \in S \text{ und } B \in \mathcal{S}. \quad (2.2.11)$$

Dann hat p genau ein Gleichgewicht $\bar{\mu}$, und

$$\lim_{n \rightarrow \infty} d_{TV}(\mu p^n, \bar{\mu}) = 0 \quad \text{für alle } \mu \in \mathcal{P}(S).$$

Beweis. Aufgrund der Minorisierungsbedingung existiert ein stochastischer Kern q auf (S, \mathcal{S}) mit

$$p(x, B) = \varepsilon \pi[B] + (1 - \varepsilon) q(x, B) \quad \text{für alle } B \in \mathcal{S}.$$

Für $\mu, \nu \in \mathcal{P}(S)$ folgt dann

$$\mu p = \varepsilon \pi + (1 - \varepsilon) \mu q.$$

Damit erhalten wir aus einer Kopplung (V, W) von μq und νq eine Kopplung von μp und νp , indem wir

$$(X', Y') := \begin{cases} (U, U) & \text{falls } B = 1, \\ (V, W) & \text{falls } B = 0, \end{cases}$$

mit unabhängigen Zufallsvariablen $U \sim \pi$ und $B \sim \text{Bernoulli}(\varepsilon)$ setzen. Für die Variationsdistanz ergibt sich die Abschätzung

$$d_{TV}(\mu p, \nu p) = P[X' \neq Y'] = P[V \neq W \text{ und } B = 0] = (1 - \varepsilon) P[V \neq W].$$

Durch Minimieren über alle Kopplungen (V, W) von μq und νq folgt

$$d_{TV}(\mu p, \nu p) \leq d_{TV}(\mu q, \nu q) \leq d_{TV}(\mu, \nu),$$

wobei wir im letzten Schritt verwendet haben, dass stochastische Kerne Kontraktionen bzgl. der Variationsdistanz sind. Die Behauptung folgt nun wie in Satz 2.6. \square

Beispiel (Kartenhaus). Der Kartenhaus-Prozess ist eine Markovkette auf \mathbb{Z}_+ mit Übergangswahrscheinlichkeiten $p(x, x+1) = a$, $p(x, 0) = 1 - a$, $a \in (0, 1)$. Es gilt $p(x, \cdot) \geq (1 - a) \cdot \delta_0$ für alle $x \in S$. Also gibt es nach dem Korollar ein eindeutiges Gleichgewicht, und die Verteilung der Kette konvergiert in Variationsdistanz gegen die Gleichgewichtsverteilung.

2.3 Markov-Eigenschaften

In diesem Abschnitt werden wir verschiedene Versionen der charakterisierenden Eigenschaft einer Markovkette betrachten. Wir führen zunächst das kanonische Modell einer Markovkette $(X_n)_{n \geq 0}$ ein.

2.3.1 Kanonisches Modell einer Markovkette

In Satz 2.3 haben wir die endlich-dimensionalen Verteilungen $P \circ (X_0, X_1, \dots, X_n)^{-1}$ einer Markovkette berechnet. Viele relevante Ereignisse hängen aber von unendlich vielen der Zufallsvariablen X_n ab. Die gemeinsame Verteilung aller dieser Zufallsvariablen ist eine Wahrscheinlichkeitsverteilung auf dem unendlichen Produktraum

$$S^\infty := S^{\{0,1,2,\dots\}} = \{x = (x_0, x_1, x_2, \dots) : x_i \in S\}$$

aller diskreten Pfade (Folgen) mit Werten in S . Wir versehen die Menge S^∞ wie üblich mit der von den Koordinatenabbildungen

$$\pi_k : S^\infty \rightarrow S, \quad \pi_k(x) = x_k,$$

erzeugten Produkt- σ -Algebra

$$\mathcal{F}_\infty = \sigma(\pi_k : k \geq 0) = \bigotimes_{k \geq 0} \mathcal{S}.$$

Einen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definierten stochastischen Prozess $(X_n)_{n \geq 0}$ können wir auch als Abbildung

$$X = (X_n) : \Omega \rightarrow S^\infty$$

auffassen. Die Abbildung X ist eine S^∞ -wertige Zufallsvariable, also messbar bzgl. der σ -Algebren $\mathcal{A}/\mathcal{F}_\infty$, denn \mathcal{F}_∞ wird von den Koordinatenabbildungen π_k erzeugt, und $\pi_k(X) = X_k$ ist für alle $k \geq 0$ messbar. Wir können daher die Verteilung

$$\mu_X[A] = P[(X_n) \in A], \quad A \in \mathcal{F}_\infty,$$

des stochastischen Prozesses (X_n) auf dem Pfadraum $(S^\infty, \mathcal{F}_\infty)$ betrachten.

Wir beschränken uns nun wieder auf Markovketten. Seien p_1, p_2, \dots stochastische Kerne, und μ eine Wahrscheinlichkeitsverteilung auf (S, \mathcal{S}) .

Satz 2.8 (Existenz und Eindeutigkeit in Verteilung von Markovketten). (1). Es existiert genau eine Wahrscheinlichkeitsverteilung P_μ auf dem unendlichen Produktraum $(S^\infty, \mathcal{F}_\infty)$, bzgl. der die Folge $(\pi_n)_{n \geq 0}$ der Koordinatenabbildungen eine Markovkette mit Startverteilung $\mu(dx)$ und Übergangskern $p_n(x, dy)$ ist.

(2). Ist $(X_n)_{n \geq 0}$ auf (Ω, \mathcal{A}, P) eine beliebige Markovkette mit Startverteilung μ und Übergangswahrscheinlichkeiten p_n , dann gilt

$$P[(X_n) \in A] = P_\mu[A] \quad \text{für alle } A \in \mathcal{F}_\infty,$$

d.h. P_μ ist die Verteilung von (X_n) auf $(S^\infty, \mathcal{F}_\infty)$.

Bemerkung (Unendliches mehrstufiges Modell). Die Verteilung P_μ der Markovkette entspricht einem mehrstufigen Modell auf dem unendlichen Produktraum $S^\infty = S^{\{0,1,2,\dots\}}$:

$$P_\mu(dx) = \mu(dx_0)p_1(x_0, dx_1)p_2(x_1, dx_2) \cdot \dots$$

Beweis. Nach Satz 2.3 ist ein stochastischer Prozess (X_n) genau dann eine Markovkette zu μ und p_n , wenn (X_0, \dots, X_n) für jedes $n \geq 0$ die Verteilung

$$\mu_{0:n}(dx_{0:n}) := \mu(dx_0)p_1(x_0, dx_1) \cdot \dots \cdot p_n(x_{n-1}, dx_n)$$

hat. Zu zeigen ist, dass zu der Familie $\mu_{0:n}, n \geq 0$, von Wahrscheinlichkeitsverteilungen auf dem endlichdimensionalen Produkträumen $S^{\{0,1,\dots,n\}}$ eine eindeutige Wahrscheinlichkeitsverteilung P_μ auf den unendlichen Produktraum S^∞ existiert, bzgl. der die ersten $n+1$ Koordinaten x_0, \dots, x_n für jedes n die Verteilung $\mu_{0:n}$ haben. Die Folge $\pi_n(x) = x_n$ der Koordinatenabbildungen ist dann unter P_μ eine Markovkette mit den vorgegebenen Übergangswahrscheinlichkeiten.

Existenz: Die Wahrscheinlichkeitsverteilungen $\mu_{0:n}$ auf den endlichdimensionalen Produkträumen $S^{\{0,1,\dots,n\}}$ sind *konsistent*, d.h. für $m \leq n$ stimmt die Verteilung der ersten $m+1$ Koordinaten unter $\mu_{0:n}$ mit $\mu_{0:m}$ überein. Aus dem Fortsetzungssatz von Carathéodory folgt nun allgemein, dass zu einer Familie von konsistenten endlichdimensionalen Verteilungen eine Wahrscheinlichkeitsverteilung auf dem unendlichen Produktraum mit den entsprechenden Randverteilungen existiert (*Fortsetzungssatz von Kolmogorov*). Wir verzichten hier auf den Beweis dieser maßtheoretischen Aussage, der sich in vielen Lehrbüchern zur Wahrscheinlichkeitstheorie findet, siehe z.B. [Bauer], [Klenke], oder den Anhang in [Durrett].

Eindeutigkeit: Ein stochastischer Prozess (X_n) auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) ist genau dann eine Markovkette mit Parametern μ und p_n , wenn

$$\int_{B_0} \mu(dx_0) \int_{B_1} p_1(x_0, dx_1) \cdots \int_{B_n} p_n(x_{n-1}, dx_n) = P[X_0 \in B_0, \dots, X_n \in B_n] = P[X \in A] \quad (2.3.1)$$

für jede Zylindermenge der Form

$$A = B_0 \times B_1 \times \dots \times B_n \times S \times S \times \dots = \{\pi_0 \in B_0, \dots, \pi_n \in B_n\},$$

mit $n \in \mathbb{N}$ und $B_0, \dots, B_n \in \mathcal{S}$ gilt. Da die Zylindermengen ein durchschnittsstabiles Erzeugendensystem der Produkt- σ -Algebra \mathcal{F}_∞ bilden, ist die Verteilung von X auf $(S^\infty, \mathcal{F}_\infty)$ durch (2.3.1) eindeutig festgelegt.

Ist $X_n = \pi_n$ der Koordinatenprozess auf dem Produktraum $(\Omega, \mathcal{A}) = (S^\infty, \mathcal{F}_\infty)$, dann stimmt die Verteilung von X mit dem zugrundeliegenden Wahrscheinlichkeitsmaß P überein, d.h. P ist durch (2.3.1) eindeutig festgelegt. □

Bemerkung (Konstruktive Existenzbeweise). Im Fall $S = \mathbb{R}$ erhalten wir die Wahrscheinlichkeitsverteilung P_μ auch direkt als Verteilung der im letzten Abschnitt explizit konstruierten Markovkette (X_n) mit Startverteilung μ und Übergangswahrscheinlichkeiten p_n . Auch auf allgemeineren Zustandsräumen kann man die Existenz von P_μ auf ähnliche Weise aus der Existenz einer Folge von auf $(0, 1)$ gleichverteilten, unabhängigen Zufallsvariablen herleiten (z.B. durch eine messbare Transformation des Zustandsraums nach \mathbb{R}).

Nach Satz 2.8 können wir eine Markovkette mit beliebigen Übergangswahrscheinlichkeiten durch die Koordinatenabbildungen auf dem unendlichen Produktraum $S^\infty = S^{\{0,1,2,\dots\}}$ realisieren.

Definition (Kanonisches Modell). *Der durch die Koordinatenabbildungen $\pi_n(x) = x_n$ gegebene stochastische Prozess auf dem Wahrscheinlichkeitsraum $(S^\infty, \mathcal{F}_\infty, P_\mu)$ heißt **kanonisches Modell** der Markovkette mit Startverteilung μ und Übergangswahrscheinlichkeiten p_n .*

Allgemein kann man jeden stochastischen Prozess im kanonischen Modell realisieren, indem man zur Verteilung des Prozesses auf dem Pfadraum übergeht.

2.3.2 Markov-Eigenschaft

Im folgenden Abschnitt werden wir die wichtige Verbindung von Markovketten und Differenzengleichungen betrachten. Dazu beweisen wir zunächst eine weitergehende Form der definierenden Eigenschaft einer Markovkette.

Sei $(X_n)_{n \geq 0}$ auf (Ω, \mathcal{A}, P) eine Markovkette mit Startverteilung μ und Übergangskernen p_n . Ist (S, \mathcal{S}) der Zustandsraum, dann hat (X_n) nach Satz 2.8 die Verteilung

$$P_\mu(dx) = \mu(dx_0)p_1(x_0, dx_1)p_2(x_1, dx_2) \cdot \dots$$

auf dem unendlichen Produktraum $S^\infty = S^{\{0,1,2,\dots\}}$. Wir bezeichnen im Folgenden die Verteilung P_{δ_x} der Markovkette bei Startwert x kurz mit P_x . Entsprechend sei $P_x^{(n)}$ die Verteilung der Markovkette mit Start in x und Übergangskernen p_{n+1}, p_{n+2}, \dots . In Erweiterung der definierenden Eigenschaft einer Markovkette können wir die bedingte Verteilung der um n Schritte verschobenen Kette gegeben den Verlauf bis zur Zeit n identifizieren:

Satz 2.9 (Markov-Eigenschaft). Für alle $n \geq 0$ und alle \mathcal{F}_∞ -messbaren Funktionen $F : S^\infty \rightarrow [0, \infty)$ gilt:

$$E[F(X_n, X_{n+1}, \dots) \mid X_{0:n}] = E_{X_n}^{(n)}[F] \quad P\text{-fast sicher.} \quad (2.3.2)$$

Bemerkung. (1). Für zeitlich homogene Markovketten gilt $P_x^{(n)} = P_x$ für alle n .

(2). Für diskrete Zustandsräume ergibt sich, dass (X_n, X_{n+1}, \dots) unter der bedingten Verteilung gegeben $X_{0:n} = x_{0:n}$ für jedes $n \geq 0$ und $x_{0:n} \in S^{n+1}$ mit $P[X_{0:n} = x_{0:n}] \neq 0$ eine Markovkette mit Start in x_n und Übergangskernen p_{n+1}, p_{n+2}, \dots ist.

Beweis. Der Beweis erfolgt in mehreren Schritten:

(1). Wir nehmen zunächst an, dass die Funktion F nur von endlich vielen Variablen abhängt, d.h.

$$F(x_0, x_1, \dots) = f(x_{0:k}) \quad \text{für ein } k \geq 0 \text{ und eine messbare Funktion } f : S^{k+1} \rightarrow \mathbb{R}_+. \quad (2.3.3)$$

In diesem Fall können wir direkt verifizieren, dass $E_{X_n}^{(n)}[F]$ eine Version der bedingten Erwartung in (2.3.2) ist:

(a) Es gilt $E_{X_n}^{(n)}[F] = g(X_n)$ mit

$$g(z) = E_z^{(n)}[F] = \int p_1(z, dx_1) \int p_2(x_1, dx_2) \cdots \int p_k(x_{k-1}, dx_k) f(z, x_1, \dots, x_k).$$

Da $f : S^{k+1} \rightarrow \mathbb{R}_+$ produktmessbar ist, ist $g : S \rightarrow \mathbb{R}_+$ messbar.

(b) Für $n \geq 0$ und eine messbare Funktion $h : S^{n+1} \rightarrow \mathbb{R}_+$ gilt

$$\begin{aligned} E[F(X_n, X_{n+1}, \dots)h(X_{0:n})] &= E[f(X_{n:n+k})h(X_{0:n})] \\ &= \int \mu(dx_0) \int p_1(x_0, dx_1) \cdots \int p_n(x_{n-1}, dx_n) h(x_{0:n}) \times \\ &\quad \times \underbrace{\int p_{n+1}(x_n, dx_{n+1}) \cdots \int p_{n+k}(x_{n+k-1}, dx_{n+k}) f(x_{n:n+k})}_{= E_{X_n}^{(n)}[F]} \\ &= E \left[E_{X_n}^{(n)}[F] \cdot h(X_{0:n}) \right]. \end{aligned}$$

- (2). Nach (1) gilt (2.3.2) für Indikatorfunktionen $F = I_A$ von Zylindermengen der Form $A = \{x \in S^\infty : x_0 \in B_0, \dots, x_n \in B_n\}$ mit $n \in \mathbb{N}$ und $B_0, \dots, B_n \in \mathcal{S}$. Wir zeigen nun, dass die Aussage dann auch für Indikatorfunktionen von beliebigen Mengen A aus der Produkt- σ -Algebra \mathcal{F}_∞ gilt. Dazu bemerken wir, dass das System \mathcal{D} aller Mengen $A \in \mathcal{F}_\infty$, für die (2.3.2) mit $F = I_A$ gilt, ein Dynkinsystem ist. Sind beispielsweise $A_1, A_2, \dots \in \mathcal{D}$ disjunkt, dann ist auch $\bigcup_k A_k$ in \mathcal{D} enthalten, denn

$$\begin{aligned} E[I_{\bigcup_k A_k}(X_n, X_{n+1}, \dots) \mid X_{0:n}] &= \sum_k E[I_{A_k}(X_n, X_{n+1}, \dots) \mid X_{0:n}] \\ &= \sum_k E_{X_n}^{(n)}[I_{A_k}] = E_{X_n}^{(n)}[I_{\bigcup_k A_k}] \quad P\text{-fast sicher.} \end{aligned}$$

Da die Zylindermengen ein durchschnittsstabiles Erzeugendensystem der Produkt- σ -Algebra bilden, folgt $\mathcal{D} = \mathcal{F}_\infty$, d.h. (2.3.2) gilt für alle $F = I_A$ mit $A \in \mathcal{F}_\infty$.

- (3). Die Aussage (2.3.2) für beliebige nicht-negative \mathcal{F}_∞ -messbare Funktionen F folgt nun wie üblich durch maßtheoretische Induktion. □

Bemerkung (Markov-Eigenschaft im kanonischen Modell). Im kanonischen Modell können wir die Markoveigenschaft noch etwas kompakter formulieren. Sei $\theta : S^\infty \rightarrow S^\infty$ die durch

$$\theta(x_0, x_1, \dots) = (x_1, x_2, \dots)$$

definierte Shift-Abbildung auf dem Pfadraum S^∞ , und seien $X_n : S^\infty \rightarrow S$,

$$X_n(x_0, x_1, \dots) = x_n,$$

die Koordinatenabbildungen. Dann gilt für alle Wahrscheinlichkeitsverteilungen μ auf (S, \mathcal{S}) und alle messbaren Funktionen $F : S^\infty \rightarrow \mathbb{R}_+$:

$$E_\mu[F \circ \theta^n \mid X_{0:n}] = E_{X_n}^{(n)}[F] \quad P\text{-fast sicher.} \quad (2.3.4)$$

Das folgende Korollar liefert eine weitere äquivalente Formulierung der Markov-Eigenschaft.

Korollar 2.10 (Markov-Eigenschaft, 2. Version). *Ist (X_n) unter P eine Markovkette mit Parametern μ und p_n , dann ist (X_n, X_{n+1}, \dots) **bedingt unabhängig von** (X_0, \dots, X_n) **gegeben** X_n mit bedingter Verteilung $P_{X_n}^{(n)}$, d.h.*

$$\begin{aligned} E[F(X_n, X_{n+1}, \dots)g(X_0, \dots, X_n) \mid X_n] \\ &= E_{X_n}^{(n)}[F] \cdot E[g(X_0, \dots, X_n) \mid X_n] \\ &= E[F(X_n, X_{n+1}, \dots) \mid X_n] \cdot E[g(X_0, \dots, X_n) \mid X_n] \quad P\text{-fast sicher} \end{aligned}$$

für alle messbaren Funktionen $F : S^\infty \rightarrow [0, \infty)$ und $g : S^{n+1} \rightarrow [0, \infty)$.

Beweis. Wegen der Projektivität der bedingten Erwartung gilt nach Satz 2.9:

$$\begin{aligned} E[F(X_{n:\infty})g(X_{0:n}) | X_n] &= E[E[F(X_{n:\infty})g(X_{0:n}) | X_{0:n}] | X_n] \\ &= E \left[E_{X_n}^{(n)}[F] \cdot g(X_{0:n}) \mid X_n \right] = E_{X_n}^{(n)}[F] \cdot E[g(X_{0:n}) | X_n]. \end{aligned}$$

□

Das Korollar besagt anschaulich, dass, gegeben den gegenwärtigen Zustand X_n , die zukünftige Entwicklung einer Markovkette bedingt unabhängig von der vorherigen Entwicklung ist:

„Die Zukunft ist bedingt unabhängig von der Vergangenheit gegeben die Gegenwart.“

2.3.3 Starke Markov-Eigenschaft

Wir beweisen nun eine Erweiterung der Markoveigenschaft auf zufällige Zeiten (Stoppzeiten). Die Information, die über einen stochastischen Prozess $(X_n)_{n \geq 0}$ bis zur Zeit n vorliegt, wird beschrieben durch die σ -Algebra

$$\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n).$$

Sei $T : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$ eine nichtnegative ganzzahlige Zufallsvariable. T heißt eine **Stoppzeit** (bzgl. der σ -Algebren \mathcal{F}_n), falls

$$\{T = n\} \in \mathcal{F}_n \quad \text{für alle } n \geq 0 \text{ gilt.}$$

Nach dem Faktorisierungslemma ist T genau dann eine Stoppzeit bzgl. (\mathcal{F}_n) , wenn $I_{\{T=n\}}$ für jedes n eine Funktion von X_0, \dots, X_n ist. Anschaulich bedeutet dies, dass aufgrund der Information, die bis zur Zeit n vorliegt, entscheidbar ist, ob T den Wert n annimmt.

Beispiel (Trefferzeiten). (1). Die **erste Trefferzeit**

$$T_B = \min\{n \geq 0 \mid X_n \in B\} \quad (\min \emptyset := \infty)$$

und die **erste Passier- bzw. Rückkehrzeit**

$$S_B = \min\{n \geq 1 \mid X_n \in B\} \quad (\min \emptyset := \infty)$$

einer messbaren Teilmenge B des Zustandsraumes S sind Stoppzeiten. Zum Beispiel gilt

$$\{T_B = n\} = \{X_1 \in B^C, \dots, X_{n-1} \in B^C, X_n \in B\} \in \mathcal{F}_n \quad \text{für alle } n \geq 0.$$

Hat man beispielsweise beschlossen, eine Aktie zu verkaufen, sobald ihr Kurs X_n den Wert λ überschreitet, dann ist der Verkaufszeitpunkt gleich $T_{(\lambda, \infty)}$, also eine Stoppzeit.

(2). Die **letzte Besuchszeit**

$$L_B := \sup\{n \geq 0 : X_n \in B\} \quad (\sup \emptyset := 0)$$

ist dagegen in der Regel keine Stopzeit (Übung). Um zu entscheiden, ob $L_B = n$ gilt, benötigt man nämlich Informationen über die zukünftige Entwicklung des Prozesses.

Die Information, die bis zu einer Stopzeit vorliegt, wird beschrieben durch die σ -Algebra

$$\mathcal{F}_T = \{A \in \mathcal{A} : A \cap \{T = n\} \in \mathcal{F}_n \text{ für alle } n \geq 0\},$$

der „bis zur Zeit T beobachtbaren“ Ereignisse. Durch maßtheoretische Induktion zeigt man, dass eine Abbildung $Y : \Omega \rightarrow \mathbb{R}$ genau dann bzgl. \mathcal{F}_T messbar ist, wenn $Y \cdot I_{\{T=n\}}$ für jedes $n \geq 0$ \mathcal{F}_n -messbar, also eine Funktion von X_0, \dots, X_n ist. Insbesondere ist die Position X_T des Prozesses (X_n) zur Stopzeit T eine \mathcal{F}_T -messbare Zufallsvariable, denn für $n \geq 0$ ist

$$X_T \cdot I_{\{T=n\}} = X_n \cdot I_{\{T=n\}} \quad \mathcal{F}_n\text{-messbar.}$$

Wir setzen nun wieder voraus, dass (X_n, P_x) eine zeithomogene Markovkette im kanonischen Modell ist.

Satz 2.11 (Starke Markov-Eigenschaft). *Ist $T : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$ eine Stopzeit bzgl. der σ -Algebren $\mathcal{F}_n = \sigma(X_0, X_1, \dots, X_n)$, dann gilt*

$$E_\nu[F(X_T, X_{T+1}, \dots) \mid \mathcal{F}_T] = E_{X_T}[F(X_0, X_1, \dots)] \quad P_\nu\text{-fast sicher auf } \{T < \infty\}$$

für alle Wahrscheinlichkeitsverteilungen ν auf (S, \mathcal{S}) und alle messbaren Funktionen $F : S^{\{0,1,2,\dots\}} \rightarrow \mathbb{R}_+$, wobei $F(X_T, X_{T+1}, \dots)$ auf $\{T = \infty\}$ willkürlich definiert ist.

Beweis. Sei $\theta(x_0, x_1, \dots) = (x_1, x_2, \dots)$ der Shiftoperator auf $S^{\{0,1,2,\dots\}}$. Wir müssen zeigen, dass

$$E_\nu[F \circ \theta^T \mid \mathcal{F}_T] \cdot I_{\{T < \infty\}} = E_{X_T}[F] \cdot I_{\{T < \infty\}} \quad P_\nu\text{-fast sicher} \quad (2.3.5)$$

gilt, wobei wir die rechte Seite für $T = \infty$ gleich 0 setzen. Für $A \in \mathcal{F}_T$ und $n \geq 0$ gilt $A \cap \{T = n\} \in \mathcal{F}_n$, also nach der Markoveigenschaft:

$$\begin{aligned} E_\nu[F \circ \theta^T ; A \cap \{T = n\}] &= E_\nu[F \circ \theta^n ; A \cap \{T = n\}] \\ &= E_\nu[E_{X_n}[F] ; A \cap \{T = n\}] \\ &= E_\nu[E_{X_T}[F] ; A \cap \{T = n\}] \end{aligned}$$

Durch Summieren über n erhalten wir:

$$E_\nu[F \circ \theta^T; A \cap \{T < \infty\}] = E_\nu[E_{X_T}[F]; A \cap \{T < \infty\}].$$

Also stimmen die Integrale beider Seiten von (2.3.5) über eine beliebige Menge $A \in \mathcal{F}_T$ überein. Da beide Seiten in (2.3.5) \mathcal{F}_T -messbar sind, folgt, dass diese P_ν -fast sicher übereinstimmen. \square

Anschaulich startet eine zeithomogene Markovkette also auch zu einer Stoppzeit T neu im Zustand X_T , d.h. der weitere Verlauf ist unabhängig vom vorherigen Verlauf gegeben den gegenwärtigen Zustand X_T .

2.4 Zusammenhang mit Differenzgleichungen

Wir betrachten den Zusammenhang von Markovketten und Differenzgleichungen zunächst in einem einfachen Beispiel:

Beispiel (Das klassische Ruinproblem). In jeder Runde eines Glücksspiels trete einer der folgenden Fälle ein:

- Mit Wahrscheinlichkeit $p \in (0, 1)$ gewinnt der Spieler 1 Euro dazu.
- Mit Wahrscheinlichkeit $q = 1 - p$ verliert der Spieler 1 Euro.

Die Entwicklung des Kapitals X_n des Spielers kann dann durch einen Random Walk auf \mathbb{Z} mit Übergangswahrscheinlichkeiten $p(x, x+1) = p, p(x, x-1) = q$ beschrieben werden. Sei $x \in \mathbb{Z}$ das Startkapital, und seien $a, b \in \mathbb{Z}$ mit $a \leq x \leq b$. Wir können den Random Walk ohne Beschränkung der Allgemeinheit im kanonischen Modell betrachten, d.h. P_x ist die Verteilung bei Startwert x auf dem Produktraum $\Omega = \mathbb{Z}^{\{0,1,2,\dots\}}$ und $X_n(\omega) = \omega_n$ ist die n -te Koordinatenabbildung. Das Glücksspiel soll folgende mögliche Ausgänge haben:

- Im Fall $X_n \leq a$ ist der Spieler bankrott.
- Im Fall $X_n \geq b$ ist der Gegenspieler (bzw. die Spielbank) bankrott.

Die Zeit, zu der eines dieser beiden Ereignisse zum ersten Mal eintritt, wird durch die Zufallsvariable

$$T(\omega) := \min\{n \geq 0 : X_n(\omega) \leq a \text{ oder } X_n(\omega) \geq b\}$$

beschrieben, wobei wir $\min \emptyset = \infty$ setzen. Wegen $\limsup |X_n| = +\infty$ gilt $T < \infty$ P_x -fast sicher für alle x . Also ist der Austrittspunkt

$$X_T(\omega) := X_{T(\omega)}(\omega)$$

des Random Walks (X_n) aus dem Intervall (a, b) P_x -fast sicher definiert, und mit Wahrscheinlichkeit 1 gilt $X_T = a$ (Spieler bankrott) oder $X_T = b$ (Spielbank bankrott). Wegen

$$X_T = \sum_{n=0}^{\infty} X_n \cdot I_{\{T=n\}}$$

ist auch X_T eine Zufallsvariable. Uns interessiert die *Ruinwahrscheinlichkeit*

$$h(x) := P_x[X_T = a]$$

des Spielers bei Startkapital x . Um diese zu berechnen, bedingen wir auf den ersten Schritt des Random Walks („first step analysis“). Sei dazu

$$\tilde{X}_n(\omega) := X_{n+1}(\omega) = X_n(\theta(\omega))$$

der um einen Schritt verschobene Prozess, und sei

$$\tilde{T} = \min\{n \geq 0 : \tilde{X}_n \leq a \text{ oder } \tilde{X}_n \geq b\}.$$

Für $a < x < b$ gilt $T \geq 1$, also

$$X_T(\omega) = \tilde{X}_{\tilde{T}}(\omega) = X_T(\theta(\omega)) \quad \text{für alle } \omega \in \Omega.$$

Daher folgt mit der Markoveigenschaft:

$$\begin{aligned} h(x) &= P_x[X_T = a] = P_x[X_T \circ \theta = a] \\ &= P_x[X_T \circ \theta = a \mid X_1 = x+1] \cdot P_x[X_1 = x+1] + \\ &\quad + P_x[X_T \circ \theta = a \mid X_1 = x-1] \cdot P_x[X_1 = x-1] \\ &\stackrel{(2.3.2)}{=} P_{x+1}[X_T = a] \cdot p + P_{x-1}[X_T = a] \cdot q \\ &= p \cdot h(x+1) + q \cdot h(x-1). \end{aligned}$$

Die Funktion h hat also die *gewichtete Mittelwerteigenschaft*

$$h(x) = p \cdot h(x+1) + q \cdot h(x-1), \quad \text{für alle } a < x < b.$$

Diese Eigenschaft ist äquivalent zu den Differenzgleichungen

$$0 = p \cdot (h(x+1) - h(x)) - q \cdot (h(x) - h(x-1)) \quad \text{bzw.} \quad (2.4.1)$$

$$0 = q \underbrace{((h(x+1) - h(x)) - (h(x) - h(x-1)))}_{\text{diskrete 2. Ableitung}} + (p - q) \underbrace{(h(x+1) - h(x))}_{\text{diskrete 1. Ableitung}}. \quad (2.4.2)$$

Die gesuchte Ruinwahrscheinlichkeit $h(x)$ löst (2.3.4) bzw. (2.4.1) bzw. (2.4.2) mit den Randbedingungen

$$h(a) = P_a[X_T = a] = 1, \quad h(b) = P_b[X_T = a] = 0.$$

Die Lösung der Differenzgleichung können wir leicht berechnen. Dazu verfahren wir ähnlich wie bei linearen gewöhnlichen Differentialgleichungen. Nach (2.4.1) gilt für die erste Differenz $v(x) := h(x+1) - h(x)$:

$$v(x) = \frac{q}{p} \cdot v(x-1) \quad \text{für alle } a < x < b,$$

d.h. $v(x) = c \cdot (q/p)^x$ für ein $c \in \mathbb{R}$. Wir unterscheiden folgende Fälle:

(1). *Faire Münzwürfe* ($p = q = \frac{1}{2}$): In diesem Fall ist

$$h(x) = cx + d \quad \text{mit } c, d \in \mathbb{R}$$

die allgemeine Lösung von (2.4.1) bzw. (2.4.2). Aus den Randbedingungen folgt:

$$h(x) = \frac{b-x}{b-a} \quad (a \leq x \leq b).$$

(2). $p \neq \frac{1}{2}$: In diesem Fall erhalten wir

$$h(x) = c \cdot \left(\frac{q}{p}\right)^x + d \quad \text{mit } c, d \in \mathbb{R}$$

als allgemeine Lösung. Aus den Randbedingungen folgt:

$$h(x) = \frac{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^x}{\left(\frac{q}{p}\right)^b - \left(\frac{q}{p}\right)^a} = \frac{1 - \left(\frac{p}{q}\right)^{b-x}}{1 - \left(\frac{p}{q}\right)^{b-a}}.$$

Wir haben damit die Ruinwahrscheinlichkeit in allen Fällen berechnet. Ist die Erfolgswahrscheinlichkeit p kleiner als $1/2$, dann gilt $\frac{p}{q} < 1$ und somit $h(x) \geq 1 - (p/q)^{b-x}$. Der letzte Ausdruck hängt nicht von dem Betrag a ab, bei dem der Spieler ruiniert ist. Beispielsweise gilt bei Roulette mit Höchstesatz 1 stets:

$$h(x) \geq 1 - \left(\frac{18}{19}\right)^{b-x}.$$

Bei genügend kleinem Höchstesatz geht also mit an Sicherheit grenzender Wahrscheinlichkeit der Spieler zuerst bankrott - selbst wenn das Kapital, das er mobilisieren kann, über dem der Bank liegt!

2.4.1 Differenzgleichungen für Markovketten

Die beim Ruinproblem verwendete Methode, die Berechnung von Wahrscheinlichkeiten und Erwartungswerten von Markovketten durch Konditionieren auf den ersten Schritt auf eine Differenzgleichung zurückzuführen, ist viel allgemeiner anwendbar. Wir betrachten im Folgenden eine beliebige zeithomogene Markovkette (X_n) mit Zustandsraum (S, \mathcal{S}) und Übergangskern $p(x, dy)$ im kanonischen Modell. Sei $D \in \mathcal{S}$ eine messbare Teilmenge des Zustandsraums, und sei

$$T(\omega) := \min\{n \geq 0 : X_n(\omega) \in D^C\}$$

die **erste Trefferzeit** von $D^C = S \setminus D$, d.h. die **erste Austrittszeit** der Markovkette aus dem Gebiet D . Hierbei setzen wir wieder $\min \emptyset = \infty$. Wir wollen Erwartungswerte von Typ

$$u(x) = E_x \left[\sum_{n=0}^{T-1} c(X_n) \right] + E_x [f(X_T); T < \infty] \quad (2.4.3)$$

berechnen, wobei $c : D \rightarrow \mathbb{R}$ und $f : D^C \rightarrow \mathbb{R}$ gegebene nichtnegative, messbare Funktionen sind. Interpretieren wir beispielsweise $c(x)$ als Kosten, wenn die Markovkette den Punkt x durchläuft, und $f(x)$ als Zusatzkosten, wenn die Markovkette im Punkt x aus der Menge D austritt, dann gibt $u(x)$ die mittleren Gesamtkosten an, die beim Start in x bis zum Austritt aus der Menge D anfallen. Man beachte, dass sich eine Reihe wichtiger Wahrscheinlichkeiten und Erwartungswerte von Markovketten in der Form (2.4.3) darstellen lassen.

Beispiel. (1). $c \equiv 0, f \equiv 1$: *Austrittswahrscheinlichkeit aus D bzw. Trefferwahrscheinlichkeit von D^C :*

$$u(x) = P_x[T < \infty].$$

(2). $c \equiv 0, f = I_B$: *Verteilung des Austrittspunktes X_T :*

$$u(x) = P_x[X_T \in B; T < \infty].$$

(3). $c \equiv 1, f \equiv 0$: *Mittlere Austrittszeit aus D :*

$$u(x) = E_x[T].$$

(4). $c = I_B, f \equiv 0$: *Mittlere Anzahl der Besuche in B vor Austritt aus D :*

$$u(x) = E_x \left[\sum_{n=0}^{T-1} I_B(X_n) \right] = \sum_{n=0}^{\infty} P_x[X_n \in B, n < T].$$

Satz 2.12 (Poissongleichung). u ist die minimale nichtnegative Lösung des Randwertproblems

$$\begin{aligned} u(x) - (pu)(x) &= c(x) && \text{für } x \in D, \\ u(x) &= f(x) && \text{für } x \in D^C. \end{aligned} \quad (2.4.4)$$

Beweis. (1). Wir zeigen zunächst durch Bedingen auf den ersten Schritt, dass u das Randwertproblem löst. Dazu betrachten wir – wie oben – die verschobene Markovkette $\tilde{X}_n = X_{n+1}$ und die entsprechende Austrittszeit $\tilde{T} = \min\{n \geq 0 : \tilde{X}_n \in D^C\}$. Für $x \in D$ gilt P_x -fast sicher $T \geq 1$, also

$$X_T = \tilde{X}_{\tilde{T}} \quad \text{und} \quad \sum_{n=0}^{T-1} c(X_n) = c(X_0) + \sum_{n=0}^{\tilde{T}-1} c(\tilde{X}_n).$$

Damit erhalten wir unter Verwendung der Markoveigenschaft:

$$\begin{aligned} & E_x \left[\sum_{n=0}^{T-1} c(X_n) + f(X_T) \cdot I_{\{T < \infty\}} \mid X_1 \right] \\ &= E_x \left[c(x) + \sum_{n=0}^{\tilde{T}-1} c(\tilde{X}_n) + f(\tilde{X}_{\tilde{T}}) \cdot I_{\{\tilde{T} < \infty\}} \mid X_1 \right] \\ &= c(x) + E_{X_1} \left[\sum_{n=0}^{T-1} c(X_n) + f(X_T) \cdot I_{\{T < \infty\}} \right] \\ &= c(x) + u(X_1) \quad P_x\text{-fast sicher,} \end{aligned}$$

wobei wir $f(X_T) \cdot I_{\{T < \infty\}} := 0$ auf $\{T = \infty\}$ setzen. Durch Bilden des Erwartungswertes bzgl. P_x ergibt sich:

$$u(x) = c(x) + E_x[u(X_1)] = c(x) + (pu)(x) \quad \text{für alle } x \in D.$$

Für $x \in D^C$ gilt $T = 0$ P_x -fast sicher, und damit

$$u(x) = E_x[f(X_0)] = f(x) \quad \text{für alle } x \in D^C.$$

Also löst u das Randwertproblem (2.4.4).

(2). Sei nun $v \geq 0$ eine beliebige Lösung des Randwertproblems. Wir wollen zeigen, dass $v \geq u$ gilt. Dazu betrachten wir für $m \in \mathbb{N}$ die Funktion

$$u_m(x) := E_x \left[\sum_{n=0}^{(T \wedge m)-1} c(X_n) + f(X_T) \cdot I_{\{T \leq m\}} \right], \quad x \in S.$$

Nach dem Satz über monotone Konvergenz gilt $u(x) = \sup_{m \geq 1} u_m(x)$. Durch Konditionieren auf den ersten Schritt erhalten wir ähnlich wie oben:

$$\begin{aligned} u_{m+1}(x) &= c(x) + (p u_m)(x) && \text{für } x \in D, && \text{und} && (2.4.5) \\ u_{m+1}(x) &= f(x) && \text{für } x \in D^C. \end{aligned}$$

Wir zeigen nun durch Induktion nach m :

$$v \geq u_m \quad \text{für alle } m \geq 0. \quad (2.4.6)$$

Für $m = 0$ ist (2.4.6) erfüllt, denn nach Voraussetzung gilt

$$v(x) \geq 0 = u_0(x) \quad \text{für alle } x \in D, \text{ und } v(x) = f(x) = u_0(x) \quad \text{für alle } x \in D^C.$$

Gilt (2.4.6) für ein $m \geq 0$, dann folgt zudem

$$\begin{aligned} v &= pv + c \geq pu_m + c \stackrel{(2.4.5)}{=} u_{m+1} && \text{auf } D, \text{ und} \\ v &= f = u_{m+1} && \text{auf } D^C, \end{aligned}$$

d.h. (2.4.6) gilt auch für $m + 1$. Also ist (2.4.6) für alle $m \geq 0$ erfüllt. Damit folgt aber auch

$$v \geq \sup u_m = u,$$

d.h. u ist tatsächlich die *minimale* nichtnegative Lösung von (2.4.4). □

Wir wollen uns nun das erhaltene Randwertproblem genauer ansehen. In kompakter Notation können wir (2.4.4) schreiben als

$$\begin{aligned} -\mathcal{L}u &= c && \text{auf } D, \\ u &= f && \text{auf } D^C \end{aligned} \quad (2.4.7)$$

mit

$$(\mathcal{L}u)(x) := (pu)(x) - u(x) = \int p(x, dy)(u(y) - u(x)).$$

Der lineare Operator $\mathcal{L} = p - I$ heißt **Generator** der Markovkette. Auf diskreten Zustandsräumen ist \mathcal{L} ein Differenzenoperator:

$$(\mathcal{L}u)(x) = \sum_{y \in S} p(x, y)(u(y) - u(x)).$$

Beispiel (Random Walk auf \mathbb{Z}^d , Poissongleichung und Dirichletproblem). Für den klassischen d -dimensionalen Random Walk gilt

$$p(x, y) = \begin{cases} \frac{1}{2d} & \text{falls } |y - x| = 1, \\ 0 & \text{sonst.} \end{cases}$$

Damit ergibt sich

$$\begin{aligned} (\mathcal{L}u)(x) &= \frac{1}{2d} \sum_{i=1}^d (u(x + e_i) - u(x) + u(x - e_i) - u(x)) \\ &= \frac{1}{2d} \sum_{i=1}^d ((u(x + e_i) - u(x)) - (u(x) - u(x - e_i))). \end{aligned}$$

Also ist

$$\mathcal{L} = \frac{1}{2d} \Delta_{\mathbb{Z}^d}$$

der diskrete Laplace-Operator multipliziert mit der Übergangswahrscheinlichkeit. (2.4.7) ist also ein Randwertproblem für die *diskrete Poissongleichung*

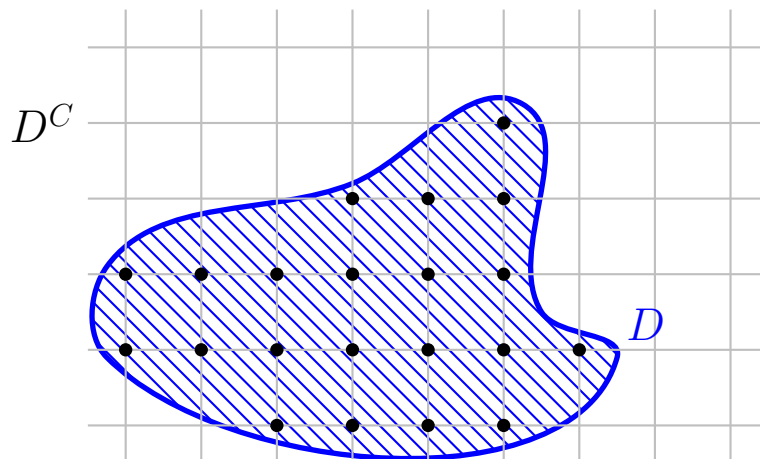
$$(\Delta_{\mathbb{Z}^d} u)(x) = -2dc(x).$$

Beispielsweise ist die mittlere Austrittszeit $u(x)$ des Random Walks mit Start in x aus einer Menge D durch die minimale nichtnegative Lösung des Randwertproblems

$$\begin{aligned} \Delta_{\mathbb{Z}^d} u &= -2d && \text{auf } D, \\ u &= 0 && \text{auf } D^C, \end{aligned}$$

gegeben. Wollen wir die Verteilung des Austrittspunktes X_T berechnen (wie z.B. beim Ruinproblem), dann müssen wir $c \equiv 0$ setzen. In diesem Fall ist (2.4.7) ein *diskretes Dirichletproblem*: Gesucht ist eine Funktion $u : \mathbb{Z}^d \rightarrow \mathbb{R}$ mit

$$\begin{aligned} \Delta_{\mathbb{Z}^d} u &= 0 && \text{auf } D, \\ u &= f && \text{auf } D^C. \end{aligned}$$

Abbildung 2.4: Diskretes Dirichletproblem auf einer Menge $D \subset \mathbb{Z}^2$.

2.4.2 Dirichletproblem und Austrittsverteilung

Allgemein nennen wir Funktionen $h : S \rightarrow \mathbb{R}$ mit $\mathcal{L}h = 0$ harmonisch.

Definition (Harmonische Funktion einer Markovkette). Eine nach unten beschränkte, messbare Funktion $h : S \rightarrow \mathbb{R}$ heißt **harmonisch auf der Menge D bzgl. des stochastischen Kerns p** , falls

$$(\mathcal{L}h)(x) = (ph)(x) - h(x) = 0 \quad \text{für alle } x \in D$$

gilt, d.h. falls h die **verallgemeinerte Mittelwerteigenschaft**

$$\int p(x, dy)h(y) = h(x) \quad \text{für alle } x \in D \quad (2.4.8)$$

besitzt.

Als Spezialfall von Satz 2.12 erhalten wir:

Korollar 2.13 (Stochastische Lösung des Dirichletproblems). Die Funktion

$$u(x) = E_x[f(X_T) ; T < \infty]$$

ist die **minimale nichtnegative Lösung des Dirichletproblems**

$$u \text{ harmonisch auf } D, \quad u = f \quad \text{auf } D^C. \quad (2.4.9)$$

Bemerkung (Lokalität). Ist S abzählbar, dann sind für die Lösung des Dirichletproblems nur die Werte von f auf dem äußeren Rand

$$\partial_{\text{ext}}D = \{y \in D^C : p(x, y) > 0 \text{ für ein } x \in D\}$$

relevant. In der Tat gilt für $u : S \rightarrow \mathbb{R}$ und $x \in D$:

$$(pu)(x) = \sum_{y \in S} p(x, y)u(y) = \sum_{y \in D} p(x, y)u(y) + \sum_{y \in \partial D} p(x, y)u(y),$$

d.h. $(\mathcal{L}u)(x)$ hängt nicht von den Werten von u auf $D^C \setminus \partial D$ ab.

- Bemerkung (Eindeutigkeit des Dirichletproblems).** (1). Im Allgemeinen können mehrere Lösungen des Dirichletproblems (2.4.9) existieren. Ist beispielsweise p der Übergangskern eines klassischen Random Walks auf $\{0, 1, 2, \dots\}$ und $D = \{1, 2, \dots\}$, dann sind die Funktionen $h_a(x) = ax, a \in \mathbb{R}$, alle harmonisch mit Randwerten $h_a(0) = 0$. Ebenso ist die Lösung nicht eindeutig, falls ein $z \in S$ mit $P_z[T = \infty] \neq 0$ existiert, denn in diesem Fall ist $h(x) = P_x[T = \infty]$ eine nichttriviale harmonische Funktion mit Nullrandwerten.
- (2). Ist die Funktion f beschränkt, und ist die Austrittszeit T für alle $x \in S$ P_x -fast sicher endlich, dann ist u die eindeutige beschränkte Lösung von (2.4.9). Dies kann man z.B. mit dem Stoppsatz für Martingale beweisen, siehe Abschnitt 3.3.

Satz 2.12 und Korollar 2.13 sind erste Aspekte weitreichender Beziehungen zwischen Wahrscheinlichkeitstheorie und Analysis (Potentialtheorie) mit fundamentalen Konsequenzen auch für andere Gebiete der Mathematik wie z.B. Diskrete Mathematik, Differentialgeometrie, Numerik und mathematische Physik. Wir erwähnen hier einige wichtige Gesichtspunkte und Konsequenzen des gefundenen Zusammenhangs. Dazu setzen wir $T < \infty$ P_x -fast sicher für alle $x \in S$ voraus. Unter dieser Annahme ist

$$u(x) = E_x[f(X_T)] \tag{2.4.10}$$

für eine nichtnegative bzw. beschränkte Funktion f auf D^C die minimale nichtnegative, bzw. die eindeutige beschränkte Lösung des Dirichletproblems. Damit ergeben sich die folgenden Anwendungen bzw. Interpretationen von Korollar 2.13:

Monte-Carlo-Verfahren zur Berechnung harmonischer Funktionen: Nach dem Gesetz der großen Zahlen und (2.4.10) gilt

$$u(x) \approx \frac{1}{k} \sum_{i=1}^k f(X_{T^{(i)}}^{(i)}) \quad \text{für große } k,$$

wobei $X^{(1)}, X^{(2)}, \dots$ unabhängige Markovketten mit Start in x und Übergangskern p sind, und $T^{(i)}$ die Austrittszeit von $X^{(i)}$ aus der Menge D bezeichnet. Die Simulation von Markovketten kann daher in sehr allgemeinem Rahmen zur näherungsweisen Berechnung harmonischer Funktionen verwendet werden.

Stochastische Darstellung der Lösung des Dirichletproblems als Pfadintegral: Nach (2.4.10)

können wir die harmonische Funktion u schreiben als Integral

$$u(x) = \int_{S^{\{0,1,2,\dots\}}} f(X_{T(\omega)}(\omega)) P(d\omega)$$

über den Raum aller diskreten Pfade auf S . Ähnliche Pfadintegraldarstellungen spielen in der Quantenphysik eine wichtige Rolle, siehe z.B. die Lecture Notes von R. Feynman.

Integralformel für harmonische Funktionen: Sei $\mu_x := P_x \circ X_T^{-1}$ die Austrittsverteilung der Markovkette mit Start in x . Dann gilt:

$$u(x) = \int_{D^C} f(y) \mu_x(dy).$$

Die Austrittsverteilung μ_x ist also das **harmonische Maß** der Potentialtheorie, das eine Berechnung harmonischer Funktionen aus den Randwerten ermöglicht.

2.4.3 Beispiele harmonischer Funktionen

In einfachen Fällen kann man die harmonischen Funktionen explizit berechnen. Dies gilt insbesondere für nicht zu große endliche Zustandsräume, sowie in eindimensionalen und rotations-symmetrischen Situationen:

Diskrete Zustandsräume

Ist S abzählbar, dann ist

$$h_y(x) := P_x[T < \infty \text{ und } X_T = y]$$

für jedes $y \in D^C$ eine nichtnegative, beschränkte harmonische Funktion auf D mit Randwerten

$$h_y(x) = I_{\{y\}}(x) \quad \text{für alle } x \in D^C.$$

Eine Lösung u des Dirichletproblems zu beliebigen Randwerten $f : D^C \rightarrow \mathbb{R}_+$ erhält man als Linearkombination der Funktionen h_y : Gilt $P_x[T = \infty] = 0$ für alle $x \in S$, dann gibt es genau eine beschränkte Lösung des Dirichlet-Problems. Damit folgt, dass die Funktionen $h_y, y \in D^C$, eine Basis des Vektorraums aller beschränkten, harmonischen Funktionen bilden. Wir erhalten also einen Zusammenhang zwischen beschränkten harmonischen Funktionen und den möglichen Austrittspunkten $y \in D^C$ der Markovkette.

Beispiel. (1). *Ruinproblem:* Für den Random Walk auf $\{a, a + 1, \dots, b\} \subset \mathbb{Z}$ mit Übergangskern $p(x, x + 1) = p, p(x, x - 1) = q = 1 - p$, gilt

$$h_a(x) = P_x[X_T = a] = \frac{h(b) - h(x)}{h(b) - h(a)}$$

und

$$h_b(x) = P_x[X_T = b] = \frac{h(x) - h(a)}{h(b) - h(a)}$$

mit

$$h(x) := \begin{cases} x & \text{für } p = q, \\ (q/p)^x & \text{für } p \neq q. \end{cases}$$

Die Funktionen h_a und h_b bilden eine Basis des Vektorraums $\{c \cdot h + d : c, d \in \mathbb{R}\}$ aller harmonischer Funktionen.

(2). *Eine transiente Markovkette auf \mathbb{Z} :* Sei $p \in (\frac{1}{2}, 1)$ und $q = 1 - p$. Wir betrachten die Markovkette (X_n, P_x) auf \mathbb{Z} mit den folgenden Übergangswahrscheinlichkeiten:

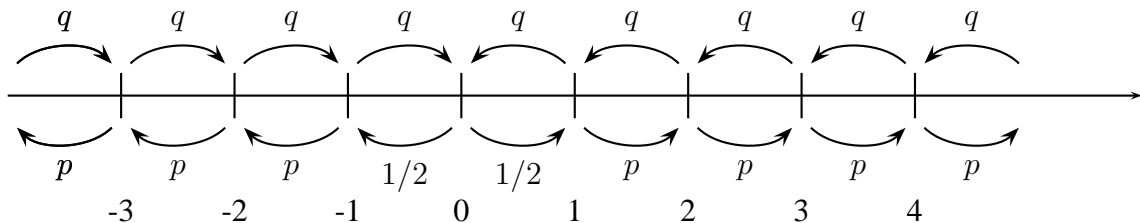


Abbildung 2.5: Übergangswahrscheinlichkeiten der Markovkette (X_n, P_x)

Für $x > 0$ gilt

$$p(x, x + 1) = p > q = p(x, x - 1),$$

für $x < 0$ dagegen umgekehrt

$$p(x, x + 1) = q < p = p(x, x - 1).$$

Daher haben die Ereignisse $\{\lim X_n = \infty\}$ und $\{\lim X_n = -\infty\}$ beide positive Wahrscheinlichkeit. Die Funktion

$$h_+(x) := \begin{cases} 1 - \frac{1}{2} \left(\frac{q}{p}\right)^x & \text{für } x > 0, \\ \frac{1}{2} \left(\frac{q}{p}\right)^{-x} & \text{für } x \leq 0, \end{cases}$$

ist harmonisch mit Randbedingungen

$$\lim_{x \rightarrow \infty} h_+(x) = 1 \quad \text{und} \quad \lim_{x \rightarrow -\infty} h_+(x) = 0.$$

Entsprechend ist $h_-(x) = h_+(-x)$ harmonisch mit

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

und jede harmonische Funktion ist eine Linearkombination von h_+ und h_- . Durch Bedingen auf den ersten Schritt der Markovkette zeigt man

$$h_+(x) = P_x[\lim X_n = \infty] \quad \text{und} \quad h_-(x) = P_x[\lim X_n = -\infty].$$

Die harmonischen Funktionen h_+ und h_- beschreiben in diesem Fall die möglichen Asymptotiken der Markovkette.

Rotationssymmetrischer Fall

Wir betrachten eine Markovkette auf $S = \mathbb{R}^d$, deren Übergangsverteilungen $p(x, dy)$ für jedes x rotationssymmetrisch mit Zentrum x sind. Beispielsweise sei $X_n = x + \sum_{i=1}^n Y_i$ ein Random Walk, dessen Inkremente Y_i unabhängig mit identischer rotationssymmetrischer Verteilung sind. Dann ist jede Funktion $u \in C^2(\mathbb{R}^d)$ mit

$$\Delta u = \sum_{i=1}^d \frac{\partial^2 u}{\partial x_i^2} = 0$$

(also jede harmonische Funktion des Laplaceoperators) auch eine harmonische Funktion des Übergangskerns p , falls u für alle $x \in \mathbb{R}^d$ bzgl. $p(x, dy)$ integrierbar ist. Aus der Greenschen Formel folgt nämlich die Mittelwerteigenschaft

$$u(x) = \text{Mittelwert von } u \text{ auf } \partial B_r(x)$$

für alle Sphären $\partial B_r(x) = \{y \in \mathbb{R}^d : |y-x| = r\}$ mit Radius $r > 0$, siehe z.B. [Forster, Analysis III]. Da $p(x, dy)$ rotationssymmetrisch ist, erhalten wir durch Integration über den Radius:

$$u(x) = \int p(x, dy) u(y),$$

d.h. u ist in der Tat harmonisch bzgl. p .

2.4.4 Mittlere Aufenthaltszeiten und Greenfunktion

Die mittlere Aufenthaltszeit

$$u(x) = E_x \left[\sum_{n=0}^{T-1} I_B(X_n) \right] = \sum_{n=0}^{\infty} P_x[X_n \in B, n < T]$$

einer Markovkette mit Übergangskern p in einer Menge $B \in \mathcal{S}$ vor Austritt aus D löst das Randwertproblem

$$\begin{aligned} u - pu &= I_A && \text{auf } D, \\ u &= 0 && \text{auf } D^C. \end{aligned}$$

Wir betrachten nun den diskreten Fall: Sei S abzählbar, $D \subset S$, und sei

$$B_y^D := \sum_{n=0}^{T-1} I_{\{y\}}(X_n), \quad y \in S,$$

die Anzahl der Besuche der Markovkette in y vor Austritt aus D . Für die mittlere Anzahl der Besuche in y bei Start in x gilt

$$E_x[B_y^D] = E_x \left[\sum_{n=0}^{\infty} I_{\{X_n \in B, n < T\}} \right] = \sum_{n=0}^{\infty} p_n^D(x, y),$$

wobei

$$p_n^D(x, y) = P_x[X_n = y, n < T]$$

die n -Schritt-Übergangswahrscheinlichkeit der Markovkette mit Absorption bei Austritt aus D bezeichnet.

Definition (Greensche Funktion einer Markovkette). Die durch

$$G^D(x, y) := \sum_{n=0}^{\infty} p_n^D(x, y)$$

definierte Funktion $G^D : S \times S \rightarrow [0, \infty]$ heißt **Greensche Funktion** der Markovkette im Gebiet D .

Korollar 2.14 (Greensche Funktion als Fundamentallösung). (1). Für $y \in S$ ist $G^D(\bullet, y)$ ist die minimale Lösung des Randwertproblems

$$\begin{aligned} (I - p)G^D(\bullet, y) &= I_{\{y\}} && \text{auf } D, \\ G^D(\bullet, y) &= 0 && \text{auf } D^C. \end{aligned}$$

(2). Für alle Funktionen $f : S \rightarrow [0, \infty]$ gilt

$$E_x \left[\sum_{n=0}^{T-1} f(X_n) \right] = (G^D f)(x).$$

Beweis. Die erste Aussage folgt unmittelbar aus Satz 2.12. Für eine Funktion $f \geq 0$ gilt:

$$E_x \left[\sum_{n=0}^{T-1} f(X_n) \right] = E_x \left[\sum_{n=0}^{T-1} \sum_{y \in S} f(y) \cdot I_{\{y\}}(X_n) \right] = \sum_{y \in S} G^D(x, y) f(y) = (G^D f)(x).$$

□

Beispiel (Random Walk auf \mathbb{Z}^d). Die Greensche Funktion des klassischen Random Walks auf \mathbb{Z}^d ist die minimale nichtnegative Lösung des Randwertproblems

$$\begin{aligned} \Delta_{\mathbb{Z}^d} G^D(\bullet, y) &= -2d I_{\{y\}} \quad \text{auf } D, \\ G^D(\bullet, y) &= 0 \quad \text{auf } D^C. \end{aligned}$$

Sie ist damit ein diskretes Analogon zur Greenschen Funktion der Analysis, die als Fundamentallösung der Poissongleichung definiert ist. Beispielsweise erhält man für den klassischen eindimensionalen Random Walk als Greensche Funktion eines Intervalls $D = \{a + 1, a + 2, \dots, b - 1\} \subset \mathbb{Z}$:

$$G^D(x, y) = \begin{cases} 2 \frac{(b-y)(x-a)}{b-a} & \text{für } a \leq x < y, \\ 2 \frac{(y-a)(b-x)}{b-a} & \text{für } y \leq x \leq b. \end{cases}$$

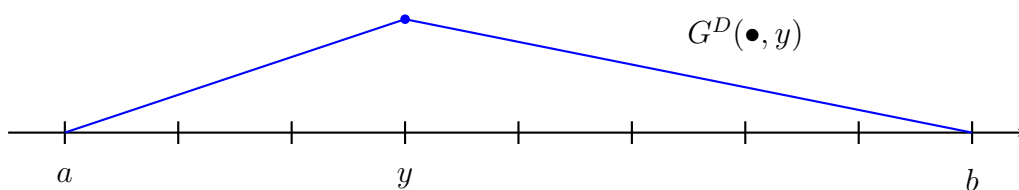


Abbildung 2.6: Greensche Funktion $G^{\{a+1, \dots, b-1\}}$ des klassischen Random Walks.

Kapitel 3

Martingale

Classical analysis starts with studying convergence of sequences of real numbers. Similarly, stochastic analysis relies on basic statements about sequences of real-valued random variables. Any such sequence can be decomposed uniquely into a martingale, i.e., a real-valued stochastic process that is “constant on average”, and a predictable part. Therefore, estimates and convergence theorems for martingales are crucial in stochastic analysis.

3.1 Definitions and examples

We fix a probability space (Ω, \mathcal{A}, P) . Moreover, we assume that we are given an increasing sequence $\mathcal{F}_n (n = 0, 1, 2, \dots)$ of sub- σ -algebras of \mathcal{A} . Intuitively, we often think of \mathcal{F}_n as describing the information available to us at time n . Formally, we define:

Definition (Filtration, adapted process). (1). A *filtration* on (Ω, \mathcal{A}) is an increasing sequence

$$\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \dots$$

of σ -algebras $\mathcal{F}_n \subseteq \mathcal{A}$.

(2). A stochastic process $(X_n)_{n \geq 0}$ is **adapted** to a filtration $(\mathcal{F}_n)_{n \geq 0}$ iff each X_n is \mathcal{F}_n -measurable.

Example. (1). The **canonical filtration** (\mathcal{F}_n^X) generated by a stochastic process (X_n) is given by

$$\mathcal{F}_n^X = \sigma(X_0, X_1, \dots, X_n).$$

If the filtration is not specified explicitly, we will usually consider the canonical filtration.

- (2). Alternatively, filtrations containing additional information are of interest, for example the filtration

$$\mathcal{F}_n = \sigma(Z, X_0, X_1, \dots, X_n)$$

generated by the process (X_n) and an additional random variable Z , or the filtration

$$\mathcal{F}_n = \sigma(X_0, Y_0, X_1, Y_1, \dots, X_n, Y_n)$$

generated by the process (X_n) and a further process (Y_n) . Clearly, the process (X_n) is adapted to any of these filtrations. In general, (X_n) is adapted to a filtration (\mathcal{F}_n) if and only if $\mathcal{F}_n^X \subseteq \mathcal{F}_n$ for any $n \geq 0$.

3.1.1 Martingales and supermartingales

We can now formalize the notion of a real-valued stochastic process that is constant (respectively decreasing or increasing) on average:

Definition (Martingale, supermartingale, submartingale). (1). A sequence of real-valued random variables $M_n : \Omega \rightarrow \mathbb{R}$ ($n = 0, 1, \dots$) on the probability space (Ω, \mathcal{A}, P) is called a **martingale w.r.t. the filtration** (\mathcal{F}_n) if and only if

- (a) (M_n) is adapted w.r.t. (\mathcal{F}_n) ,
- (b) M_n is integrable for any $n \geq 0$, and
- (c) $E[M_n | \mathcal{F}_{n-1}] = M_{n-1}$ for any $n \in \mathbb{N}$.

- (2). Similarly, (M_n) is called a **supermartingale** (resp. a **submartingale**) w.r.t. (\mathcal{F}_n) , if and only if (a) holds, the positive part M_n^+ (resp. the negative part M_n^-) is integrable for any $n \geq 0$, and (c) holds with “=” replaced by “ \leq ”, “ \geq ” respectively.

Condition (c) in the martingale definition can equivalently be written as

$$(c') \quad E[M_{n+1} - M_n | \mathcal{F}_n] = 0 \quad \text{for any } n \in \mathbb{Z}_+,$$

and correspondingly with “=” replaced by “ \leq ” or “ \geq ” for super- or submartingales.

Intuitively, a martingale is a “fair game”, i.e., M_{n-1} is the best prediction (w.r.t. the mean square error) for the next value M_n given the information up to time $n - 1$. A supermartingale is “decreasing on average”, a submartingale is “increasing on average”, and a martingale is both

“decreasing” and “increasing”, i.e., “*constant on average*”. In particular, by induction on n , a martingale satisfies

$$E[M_n] = E[M_0] \quad \text{for any } n \geq 0.$$

Similarly, for a supermartingale, the expectation values $E[M_n]$ are decreasing. More generally, we have:

Lemma 3.1. *If (M_n) is a martingale (respectively a supermartingale) w.r.t. a filtration (\mathcal{F}_n) then*

$$E[M_{n+k} | \mathcal{F}_n] \stackrel{(\leq)}{=} M_n \quad P\text{-almost surely for any } n, k \geq 0.$$

Beweis. By induction on k : The assertion holds for $k = 0$, since M_n is \mathcal{F}_n -measurable. Moreover, the assertion for $k - 1$ implies

$$\begin{aligned} E[M_{n+k} | \mathcal{F}_n] &= E[E[M_{n+k} | \mathcal{F}_{n+k-1}] | \mathcal{F}_n] \\ &= E[M_{n+k-1} | \mathcal{F}_n] = M_n \quad P\text{-a.s.} \end{aligned}$$

by the tower property for conditional expectations. □

Remark (Supermartingale Convergence Theorem). A key fact in analysis is that any lower bounded decreasing sequence of real numbers converges to its infimum. The counterpart of this result in stochastic analysis is the Supermartingale Convergence Theorem: Any lower bounded supermartingale converges almost surely, c.f. Theorem 3.12 below.

3.1.2 Some fundamental examples

a) Sums of independent random variables

A Random Walk

$$S_n = \sum_{i=1}^n \eta_i, \quad n = 0, 1, 2, \dots,$$

with independent increments $\eta_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ is a martingale w.r.t. to the filtration

$$\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n) = \sigma(S_0, S_1, \dots, S_n)$$

if and only if the increments η_i are centered random variables. In fact, for any $n \in \mathbb{N}$,

$$E[S_n - S_{n-1} | \mathcal{F}_{n-1}] = E[\eta_n | \mathcal{F}_{n-1}] = E[\eta_n]$$

by independence of the increments. Correspondingly, (S_n) is an (\mathcal{F}_n) supermartingale if and only if $E[\eta_i] \leq 0$ for any $i \in \mathbb{N}$.

b) Products of independent non-negative random variables

A stochastic process

$$M_n = \prod_{i=1}^n Y_i, \quad n = 0, 1, 2, \dots,$$

with independent non-negative factors $Y_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ is a martingale respectively a supermartingale w.r.t. the filtration

$$\mathcal{F}_n = \sigma(Y_1, \dots, Y_n)$$

if and only if $E[Y_i] = 1$ for any $i \in \mathbb{N}$, or $E[Y_i] \leq 1$ for any $i \in \mathbb{N}$ respectively. In fact, as M_n is \mathcal{F}_n -measurable and Y_{n+1} is independent of \mathcal{F}_n , we have

$$E[M_{n+1} | \mathcal{F}_n] = E[M_n \cdot Y_{n+1} | \mathcal{F}_n] = M_n \cdot E[Y_{n+1}] \quad \text{for any } n \geq 0.$$

Martingales and supermartingales of this type occur naturally in stochastic growth models. For the supermartingale property, integrability of the factors is not required.

Example (Exponential martingales). Consider a Random Walk $S_n = \sum_{i=1}^n \eta_i$ with i.i.d. increments η_i , and let

$$Z(\lambda) = E[\exp(\lambda \eta_i)], \quad \lambda \in \mathbb{R},$$

denote the moment generating function of the increments. Then for any $\lambda \in \mathbb{R}$ with $Z(\lambda) < \infty$, the process

$$M_n^\lambda := e^{\lambda S_n} / Z(\lambda)^n = \prod_{i=1}^n (e^{\lambda \eta_i} / Z(\lambda))$$

is a martingale. This martingale can be used to prove exponential bounds for Random Walks, cf. e.g. Chernov's theorem ["Einführung in die Wahrscheinlichkeitstheorie", Theorem 8.3].

Example (CRR model of stock market). In the Cox-Ross-Rubinstein binomial model of mathematical finance, the price of an asset is changing during each period either by a factor $1 + a$ or by a factor $1 + b$ with $a, b \in (-1, \infty)$ such that $a < b$. We can model the price evolution in a fixed number N of periods by a stochastic process

$$S_n = S_0 \cdot \prod_{i=1}^n X_i, \quad n = 0, 1, 2, \dots, N,$$

defined on $\Omega = \{1 + a, 1 + b\}^N$, where the initial price S_0 is a given constant, and $X_i(\omega) = \omega_i$. Taking into account a constant interest rate $r > 0$, the discounted stock price after n periods is

$$\tilde{S}_n = S_n / (1 + r)^n = S_0 \cdot \prod_{i=1}^n \frac{X_i}{1 + r}.$$

A probability measure P on Ω is called a **martingale measure** if the discounted stock price is a martingale w.r.t. P and the filtration $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$. Martingale measures are important for option pricing under no arbitrage assumptions, cf. Section 3.3 below. For $1 \leq n \leq N$,

$$E[\tilde{S}_n | \mathcal{F}_{n-1}] = E\left[\tilde{S}_{n-1} \cdot \frac{X_n}{1+r} \middle| \mathcal{F}_{n-1}\right] = \tilde{S}_{n-1} \cdot \frac{E[X_n | \mathcal{F}_{n-1}]}{1+r}.$$

Hence (\tilde{S}_n) is an (\mathcal{F}_n) martingale w.r.t. P if and only if

$$E[X_n | \mathcal{F}_{n-1}] = 1+r \quad \text{for any } 1 \leq n \leq N. \quad (3.1.1)$$

On the other hand, since in the CRR model X_n only takes the values $1+a$ and $1+b$, we have

$$\begin{aligned} E[X_n | \mathcal{F}_{n-1}] &= (1+a) \cdot P[X_n = 1+a | \mathcal{F}_{n-1}] + (1+b) \cdot P[X_n = 1+b | \mathcal{F}_{n-1}] \\ &= 1+a + (b-a) \cdot P[X_n = 1+b | \mathcal{F}_{n-1}]. \end{aligned}$$

Therefore, by (3.1.1), (\tilde{S}_n) is a martingale if and only if

$$P[X_n = 1+b | \mathcal{F}_{n-1}] = \frac{r-a}{b-a} \quad \text{for any } n = 1, \dots, N,$$

i.e., if and only if the growth factors X_1, \dots, X_N are independent with

$$P[X_n = 1+b] = \frac{r-a}{b-a} \quad \text{and} \quad P[X_n = 1+a] = \frac{b-r}{b-a}. \quad (3.1.2)$$

Hence for $r \notin [a, b]$, a martingale measure does not exist, and for $r \in [a, b]$, the product measure P on Ω satisfying (3.1.2) is the unique martingale measure. Intuitively this is plausible: If $r < a$ or $r > b$ respectively, then the stock price is always growing more or less than the discount factor $(1+r)^n$, so the discounted stock price can not be a martingale. If, on the other hand, $a < r < b$, then (\tilde{S}_n) is a martingale provided the growth factors are independent with

$$\frac{P[X_n = 1+b]}{P[X_n = 1+a]} = \frac{(1+r) - (1+a)}{(1+b) - (1+r)}.$$

We remark, however, that uniqueness of the martingale measure only follows from (3.1.1) since we have assumed that each X_n takes only two possible values (binomial model). In a corresponding trinomial model there are infinitely many martingale measures!

c) Successive prediction values

Let F be an integrable random variable and (\mathcal{F}_n) a filtration on a probability space (Ω, \mathcal{A}, P) . Then the process

$$M_n := E[F | \mathcal{F}_n], \quad n = 0, 1, 2, \dots,$$

of successive prediction values for F based on the information up to time n is a martingale. Indeed, by the tower property for conditional expectations, we have

$$E[M_n | \mathcal{F}_{n-1}] = E[E[F | \mathcal{F}_n] | \mathcal{F}_{n-1}] = E[F | \mathcal{F}_{n-1}] = M_{n-1}$$

almost surely for any $n \in \mathbb{N}$.

Remark (Representing martingales as successive prediction values). The class of martingales that have a representation as successive prediction values almost contains general martingales. In fact, for an arbitrary (\mathcal{F}_n) martingale (M_n) and any finite integer $m \geq 0$, the representation

$$M_n = E[M_m | \mathcal{F}_n]$$

holds for any $n = 0, 1, \dots, m$. Moreover, the L^1 Martingale Convergence Theorem implies that under appropriate uniform integrability assumptions, the limit $M_\infty = \lim_{m \rightarrow \infty} M_m$ exists in \mathcal{L}^1 , and the representation

$$M_n = E[M_\infty | \mathcal{F}_n]$$

holds for any $n \geq 0$, see e.g. [Introduction to Stochastic Analysis].

d) Functions of martingales

By Jensen's inequality for conditional expectations, convex functions of martingales are submartingales, and concave functions of martingales are supermartingales:

Satz 3.2 (Convex functions of martingales). *Suppose that $(M_n)_{n \geq 0}$ is an (\mathcal{F}_n) martingale, and $u : \mathbb{R} \rightarrow \mathbb{R}$ is a convex function that is bounded from below. Then $(u(M_n))$ is an (\mathcal{F}_n) submartingale.*

Beweis. Since u is lower bounded, $u(M_n)^-$ is integrable for any n . Jensen's inequality for conditional expectations now implies

$$E[u(M_{n+1}) | \mathcal{F}_n] \geq u(E[M_{n+1} | \mathcal{F}_n]) = u(M_n)$$

almost surely for any $n \geq 0$. □

Example. If (M_n) is a martingale then $(|M_n|^p)$ is a submartingale for any $p \geq 1$.

e) Functions of Markov chains

Let $p(x, dy)$ be a stochastic kernel on a measurable space (S, \mathcal{S}) .

Definition (Markov chain, superharmonic function). (1). A stochastic process $(X_n)_{n \geq 0}$ with state space (S, \mathcal{S}) defined on the probability space (Ω, \mathcal{A}, P) is called a **(time-homogeneous) Markov chain w.r.t. the filtration** (\mathcal{F}_n) and with transition kernel p , if and only if

(a) (X_n) is (\mathcal{F}_n) adapted, and

(b) $P[X_{n+1} \in B \mid \mathcal{F}_n] = p(X_n, B)$ holds P -almost surely for any $B \in \mathcal{S}$ and $n \geq 0$.

(2). A measurable function $h : S \rightarrow \mathbb{R}$ is called **superharmonic** (resp. **subharmonic**) w.r.t. p if and only if the integrals

$$(ph)(x) := \int p(x, dy)h(y), \quad x \in S,$$

exist, and

$$(ph)(x) \leq h(x) \quad (\text{respectively } (ph)(x) \geq h(x))$$

holds for any $x \in S$.

The function h is called **harmonic** iff it is both super- and subharmonic, i.e., iff

$$(ph)(x) = h(x) \quad \text{for any } x \in S.$$

By the tower property for conditional expectations, any (\mathcal{F}_n) Markov chain is also a Markov chain w.r.t. the canonical filtration generated by the process.

Example (Classical Random Walk on \mathbb{Z}^d). The standard Random Walk $(X_n)_{n \geq 0}$ on \mathbb{Z}^d is a Markov chain w.r.t. the filtration $\mathcal{F}_n^X = \sigma(X_0, \dots, X_n)$ with transition probabilities $p(x, x+e) = 1/2d$ for any unit vector $e \in \mathbb{Z}^d$. The coordinate processes $(X_n^i)_{n \geq 0}$, $i = 1, \dots, d$, are Markov chains w.r.t. the same filtration with transition probabilities

$$\bar{p}(x, x+1) = \bar{p}(x, x-1) = \frac{1}{2d}, \quad \bar{p}(x, x) = \frac{2d-2}{2d}.$$

A function $h : \mathbb{Z}^d \rightarrow \mathbb{R}$ is superharmonic w.r.t. p if and only if

$$\Delta_{\mathbb{Z}^d} h(x) = \sum_{i=1}^d (h(x+e_i) - 2h(x) + h(x-e_i)) = 2d((ph)(x) - h(x)) \leq 0$$

for any $x \in \mathbb{Z}^d$.

A function $h : \mathbb{Z} \rightarrow \mathbb{R}$ is harmonic w.r.t. \bar{p} if and only if $h(x) = ax + b$ with $a, b \in \mathbb{R}$, and h is superharmonic if and only if it is concave.

It is easy to verify that (super-)harmonic functions of Markov chains are (super-)martingales:

Satz 3.3 (Superharmonic functions of Markov chains are supermartingales). *Suppose that (X_n) is an (\mathcal{F}_n) Markov chain. Then the real-valued process*

$$M_n := h(X_n), \quad n = 0, 1, 2, \dots,$$

is a martingale (resp. a supermartingale) w.r.t. (\mathcal{F}_n) for every harmonic (resp. superharmonic) function $h : S \rightarrow \mathbb{R}$ such that $h(X_n)$ (resp. $h(X_n)^+$) is integrable for each n .

Beweis. Clearly, (M_n) is again (\mathcal{F}_n) adapted. Moreover,

$$E[M_{n+1} | \mathcal{F}_n] = E[h(X_{n+1}) | \mathcal{F}_n] = (ph)(X_n) \quad P\text{-a.s.}$$

The assertion now follows immediately from the definitions. \square

Below, we will show how to construct more general martingales from Markov chains, cf. Theorem 3.5. At first, however, we consider a simple example that demonstrates the usefulness of martingale methods in analyzing Markov chains:

Example (Multinomial resampling). Consider the Wright model for a population of N individuals (replicas) with a finite number of possible types, where each individual in generation $n + 1$ inherits a type from a randomly chosen predecessor in the n th generation. The number X_n of individuals of a given type in generation n is a Markov chain with state space $S = \{0, 1, \dots, N\}$ and transition kernel

$$p(k, \bullet) = \text{Bin}(N, k/N).$$

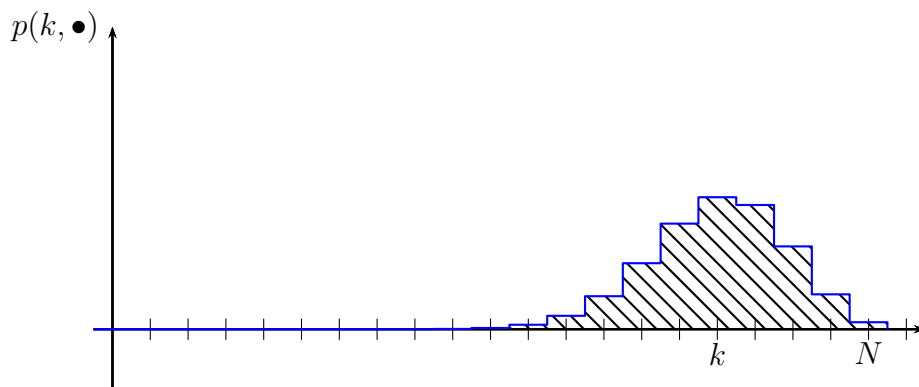


Abbildung 3.1: Transition function of (X_n) .

Moreover, as the average of this binomial distribution is k , the function $h(x) = x$ is harmonic, and the expected number of individuals in generation $n + 1$ given X_0, \dots, X_n is

$$E[X_{n+1} | X_0, \dots, X_n] = X_n.$$

Hence, the process (X_n) is a bounded martingale. The Martingale Convergence Theorem now implies that the limit $X_\infty = \lim X_n$ exists almost surely, cf. Section 3.4 below. Since X_n takes discrete values, we can conclude that $X_n = X_\infty$ eventually with probability one. In particular, X_∞ is almost surely an absorbing state. Hence

$$P[X_n = 0 \text{ or } X_n = N \text{ eventually}] = 1. \quad (3.1.3)$$

In order to compute the probabilities of the events “ $X_n = 0$ eventually” and “ $X_n = N$ eventually” we can apply the Optional Stopping Theorem for martingales, cf. Section 3.3 below. Let

$$T := \min\{n \geq 0 : X_n = 0 \text{ or } X_n = N\}, \quad \min \emptyset := \infty,$$

denote the first hitting time of the absorbing states. If the initial number X_0 of individuals of the given type is k , then by the Optional Stopping Theorem,

$$E[X_T] = E[X_0] = k.$$

Hence by (3.1.3) we obtain

$$\begin{aligned} P[X_n = N \text{ eventually}] &= P[X_T = N] = \frac{1}{N}E[X_T] = \frac{k}{N}, \quad \text{and} \\ P[X_n = 0 \text{ eventually}] &= 1 - \frac{k}{N} = \frac{N - k}{N}. \end{aligned}$$

Hence eventually all individuals have the same type, and a given type occurs eventually with probability determined by its initial relative frequency in the population.

3.2 Doob Decomposition and Martingale Problem

We will show now that any adapted sequence of real-valued random variables can be decomposed into a martingale and a predictable process. In particular, the variance process of a martingale (M_n) is the predictable part in the corresponding Doob decomposition of the process (M_n^2) . The Doob decomposition for functions of Markov chains implies the Martingale Problem characterization of Markov chains.

3.2.1 Doob Decomposition

Let (Ω, \mathcal{A}, P) be a probability space and $(\mathcal{F}_n)_{n \geq 0}$ a filtration on (Ω, \mathcal{A}) .

Definition (Predictable process). A stochastic process $(A_n)_{n \geq 0}$ is called *predictable w.r.t.* (\mathcal{F}_n) if and only if A_0 is constant and A_n is measurable w.r.t. \mathcal{F}_{n-1} for any $n \in \mathbb{N}$.

Intuitively, the value $A_n(\omega)$ of a predictable process can be predicted by the information available at time $n - 1$.

Satz 3.4 (Doob decomposition). *Every (\mathcal{F}_n) adapted sequence of integrable random variables Y_n ($n \geq 0$) has a unique decomposition (up to modification on null sets)*

$$Y_n = M_n + A_n \quad (3.2.1)$$

into an (\mathcal{F}_n) martingale (M_n) and a predictable process (A_n) such that $A_0 = 0$. Explicitly, the decomposition is given by

$$A_n = \sum_{k=1}^n E[Y_k - Y_{k-1} | \mathcal{F}_{k-1}], \quad \text{and} \quad M_n = Y_n - A_n. \quad (3.2.2)$$

Remark. (1). The increments $E[Y_k - Y_{k-1} | \mathcal{F}_{k-1}]$ of the process (A_n) are the predicted increments of (Y_n) given the previous information.

(2). The process (Y_n) is a supermartingale (resp. a submartingale) if and only if the predictable part (A_n) is decreasing (resp. increasing).

Proof of Theorem 3.4. Uniqueness: For any decomposition as in (3.2.1) we have

$$Y_k - Y_{k-1} = M_k - M_{k-1} + A_k - A_{k-1} \quad \text{for any } k \in \mathbb{N}.$$

If (M_n) is a martingale and (A_n) is predictable then

$$E[Y_k - Y_{k-1} | \mathcal{F}_{k-1}] = E[A_k - A_{k-1} | \mathcal{F}_{k-1}] = A_k - A_{k-1} \quad P\text{-a.s.}$$

This implies that (3.2.2) holds almost surely if $A_0 = 0$.

Existence: Conversely, if (A_n) and (M_n) are defined by (3.2.2) then (A_n) is predictable with $A_0 = 0$ and (M_n) is a martingale, since

$$E[M_k - M_{k-1} | \mathcal{F}_{k-1}] = 0 \quad P\text{-a.s. for any } k \in \mathbb{N}.$$

□

3.2.2 Conditional Variance Process

Consider a martingale (M_n) such that M_n is square integrable for any $n \geq 0$. Then, by Jensen's inequality, (M_n^2) is a submartingale and can again be decomposed into a martingale (\widetilde{M}_n) and a predictable process $\langle M \rangle_n$ such that $\langle M \rangle_0 = 0$:

$$M_n^2 = \widetilde{M}_n + \langle M \rangle_n \quad \text{for any } n \geq 0.$$

The increments of the predictable process are given by

$$\begin{aligned}\langle M \rangle_k - \langle M \rangle_{k-1} &= E[M_k^2 - M_{k-1}^2 \mid \mathcal{F}_{k-1}] \\ &= E[(M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1}] + 2 \cdot E[M_{k-1} \cdot (M_k - M_{k-1}) \mid \mathcal{F}_{k-1}] \\ &= \text{Var}[M_k - M_{k-1} \mid \mathcal{F}_{k-1}] \quad \text{for any } k \in \mathbb{N}.\end{aligned}$$

Here we have used in the last step that $E[M_k - M_{k-1} \mid \mathcal{F}_{k-1}]$ vanishes since (M_n) is a martingale.

Definition (Conditional variance process). *The predictable process*

$$\langle M \rangle_n := \sum_{k=1}^n \text{Var}[M_k - M_{k-1} \mid \mathcal{F}_{k-1}], \quad n \geq 0,$$

is called the **conditional variance process** of the square integrable martingale (M_n) .

Example (Random Walks). If $M_n = \sum_{i=1}^n \eta_i$ is a sum of independent centered random variables η_i and $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$ then the conditional variance process is given by $\langle M \rangle_n = \sum_{i=1}^n \text{Var}[\eta_i]$.

The conditional variance process is crucial for generalizations of classical limit theorems such as the Law of Large Numbers or the Central Limit Theorem from sums of independent random variables to martingales. A direct consequence of the fact that $M_n^2 - \langle M \rangle_n$ is a martingale is that

$$E[M_n^2] = E[M_0^2] + E[\langle M \rangle_n] \quad \text{for any } n \geq 0.$$

This can often be used to derive L^2 -estimates for martingales.

Example (Discretizations of stochastic differential equations). Consider an ordinary differential equation

$$\frac{dX_t}{dt} = b(X_t), \quad t \geq 0, \quad (3.2.3)$$

where $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is a given vector field. In order to take into account unpredictable effects on a system, one is frequently interested in studying random perturbations of the dynamics (3.2.3) of type

$$dX_t = b(X_t) dt + \text{“noise”} \quad t \geq 0, \quad (3.2.4)$$

with a random noise term. The solution $(X_t)_{t \geq 0}$ of such a stochastic differential equation (SDE) is a stochastic process in continuous time defined on a probability space (Ω, \mathcal{A}, P) where also the random variables describing the noise effects are defined. The vector field b is called the (deterministic) “drift”. We will need make sense of general SDE in Stochastic Analysis, but we

can already consider time discretizations.

For simplicity let us assume $d = 1$. Let $b, \sigma : \mathbb{R} \rightarrow \mathbb{R}$ be continuous functions, and let $(\eta_i)_{i \in \mathbb{N}}$ be a sequence of i.i.d. random variables $\eta_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ describing the noise effects. We assume

$$E[\eta_i] = 0 \quad \text{and} \quad \text{Var}[\eta_i] = 1 \quad \text{for any } i \in \mathbb{N}.$$

Here, the values 0 and 1 are just a convenient normalization, but it is an important assumption that the random variables are independent with finite variances. Given an initial value $x_0 \in \mathbb{R}$ and a fine discretization step size $h > 0$, we now define a stochastic process $(X_n^{(h)})$ in discrete time by $X_0^{(h)} = x_0$, and

$$X_{k+1}^{(h)} - X_k^{(h)} = b(X_k^{(h)}) \cdot h + \sigma(X_k^{(h)}) \sqrt{h} \eta_{k+1}, \quad \text{for } k = 0, 1, 2, \dots \quad (3.2.5)$$

One should think of $X_k^{(h)}$ as an approximation for the value of the process (X_t) at time $t = k \cdot h$. The equation (3.2.5) can be rewritten as

$$X_n^{(h)} = x_0 + \sum_{k=0}^{n-1} b(X_k^{(h)}) \cdot h + \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \cdot \sqrt{h} \cdot \eta_{k+1}. \quad (3.2.6)$$

To understand the scaling factors h and \sqrt{h} we note first that if $\sigma \equiv 0$ then (3.2.5) respectively (3.2.6) is the Euler discretization of the ordinary differential equation (3.2.3). Furthermore, if $b \equiv 0$ and $\sigma \equiv 1$, then the *diffusive scaling* by a factor \sqrt{h} in the second term ensures that the continuous time process $X_{[t/h]}^{(h)}, t \in [0, \infty)$, converges in distribution as $h \searrow 0$. Indeed, the functional central limit theorem (Donsker's invariance principle) states that the limit process in this case is a Brownian motion $(B_t)_{t \in [0, \infty)}$, see Section 4.1 and [Introduction to Stochastic Analysis]. In general, (3.2.6) is an Euler discretization of a stochastic differential equation of type

$$dX_t = b(X_t) dt + \sigma(X_t) dB_t$$

where $(B_t)_{t \geq 0}$ is a Brownian motion. Let $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$ denote the filtration generated by the random variables η_i . The following exercise summarizes basic properties of the process $X^{(h)}$ in the case of normally distributed increments.

Exercise. Suppose that the random variables η_i are standard normally distributed.

- (1). Prove that the process $X^{(h)}$ is a time-homogeneous \mathcal{F}_n Markov chain with transition kernel

$$p(x, \bullet) = N(x + b(x)h, \sigma(x)^2 h)[\bullet].$$

(2). Show that the Doob decomposition $X^{(h)} = M^{(h)} + A^{(h)}$ is given by

$$\begin{aligned} A_n^{(h)} &= \sum_{k=0}^{n-1} b(X_k^{(h)}) \cdot h \\ M_n^{(h)} &= x_0 + \sum_{k=0}^{n-1} \sigma(X_k^{(h)}) \sqrt{h} \eta_{k+1} \end{aligned} \quad (3.2.7)$$

and the conditional variance process of the martingale part is

$$\langle M^{(h)} \rangle_n = \sum_{k=0}^{n-1} \sigma(X_k^{(h)})^2 \cdot h. \quad (3.2.8)$$

(3). Conclude that

$$E[(M_n^{(h)} - x_0)^2] = \sum_{k=0}^{n-1} E[\sigma(X_k^{(h)})^2] \cdot h. \quad (3.2.9)$$

The last equation can be used in combination with the maximal inequality for martingales to derive bounds for the processes $(X^{(h)})$ in an efficient way, cf. [Introduction to Stochastic Analysis].

3.2.3 Martingale problem

For a Markov chain (X_n) we obtain a Doob decomposition

$$f(X_n) = M_n^{[f]} + A_n^{[f]} \quad (3.2.10)$$

for any function f on the state space such that $f(X_n)$ is integrable for each n . Computation of the predictable part leads to the following general result:

Satz 3.5 (Martingale problem for time-homogeneous Markov chains). *Let p be a stochastic kernel on a measurable space (S, \mathcal{S}) . Then for an (\mathcal{F}_n) adapted stochastic process $(X_n)_{n \geq 0}$ with state space (S, \mathcal{S}) the following statements are equivalent:*

- (1). (X_n) is a time homogeneous (\mathcal{F}_n) Markov chain with transition kernel p .
- (2). (X_n) is a **solution of the martingale problem for the operator $\mathcal{L} = p - I$** , i.e., there is a decomposition

$$f(X_n) = M_n^{[f]} + \sum_{k=0}^{n-1} (\mathcal{L}f)(X_k), \quad n \geq 0,$$

with an (\mathcal{F}_n) martingale $(M_n^{[f]})$ for any function $f : S \rightarrow \mathbb{R}$ such that $f(X_n)$ is integrable for each n , or, equivalently, for any bounded function $f : S \rightarrow \mathbb{R}$.

In particular, we see once more that if $f(X_n)$ is integrable and f is harmonic ($\mathcal{L}f = 0$) then $f(X_n)$ is a martingale, and if f is superharmonic ($\mathcal{L}f \leq 0$), then $f(X_n)$ is a supermartingale. The theorem hence extends Theorem 3.3 above.

Beweis. The implication “(i) \Rightarrow (ii)” is just the Doob decomposition for $f(X_n)$. In fact, by Theorem 3.4, the predictable part is given by

$$\begin{aligned} A_n^{[f]} &= \sum_{k=0}^{n-1} E[f(X_{k+1}) - f(X_k) \mid \mathcal{F}_k] \\ &= \sum_{k=0}^{n-1} E[pf(X_k) - f(X_k) \mid \mathcal{F}_k] = \sum_{k=0}^{n-1} (\mathcal{L}f)(X_k), \end{aligned}$$

and $M_n^{[f]} = f(X_n) - A_n^{[f]}$ is a martingale.

To prove the converse implication “(ii) \Rightarrow (i)” suppose that $M_n^{[f]}$ is a martingale for any bounded $f : S \rightarrow \mathbb{R}$. Then

$$\begin{aligned} 0 &= E[M_{n+1}^{[f]} - M_n^{[f]} \mid \mathcal{F}_n] \\ &= E[f(X_{n+1}) - f(X_n) \mid \mathcal{F}_n] - ((pf)(X_n) - f(X_n)) \\ &= E[f(X_{n+1}) \mid \mathcal{F}_n] - (pf)(X_n) \end{aligned}$$

almost surely for any bounded function f . Hence (X_n) is an (\mathcal{F}_n) Markov chain with transition kernel p . \square

Example (One dimensional Markov chains). Suppose that under P_x , the process (X_n) is a time homogeneous Markov chain with state space $S = \mathbb{R}$ or $S = \mathbb{Z}$, initial state $X_0 = x$, and transition kernel p . Assuming $X_n \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ for each n , we define the “drift” and the “fluctuations” of the process by

$$\begin{aligned} b(x) &:= E_x[X_1 - X_0] \\ a(x) &= \text{Var}_x[X_1 - X_0]. \end{aligned}$$

We now compute the Doob decomposition of X_n . Choosing $f(x) = x$ we have

$$(p - I)f(x) = \int y p(x, dy) - x = E_x[X_1 - X_0] = b(x).$$

Hence by Theorem 3.5,

$$X_n = M_n + \sum_{k=0}^{n-1} b(X_k) \tag{3.2.11}$$

with an (\mathcal{F}_n) martingale (M_n) . To obtain detailed information on M_n , we compute the variance process: By (3.2.11) and the Markov property, we obtain

$$\langle M \rangle_n = \sum_{k=0}^{n-1} \text{Var}[M_{k+1} - M_k \mid \mathcal{F}_k] = \sum_{k=0}^{n-1} \text{Var}[X_{k+1} - X_k \mid \mathcal{F}_k] = \sum_{k=0}^{n-1} a(X_k).$$

Therefore

$$M_n^2 = \widetilde{M}_n + \sum_{k=0}^{n-1} a(X_k) \quad (3.2.12)$$

with another (\mathcal{F}_n) martingale (\widetilde{M}_n) . The functions $a(x)$ and $b(x)$ can now be used in connection with fundamental results for martingales as e.g. the maximal inequality (cf. [Introduction to Stochastic Analysis]) to derive bounds for Markov chains in an efficient way.

3.3 Gambling strategies and stopping times

Throughout this section, we fix a filtration $(\mathcal{F}_n)_{n \geq 0}$ on a probability space (Ω, \mathcal{A}, P) .

3.3.1 Martingale transforms

Suppose that $(M_n)_{n \geq 0}$ is a martingale w.r.t. (\mathcal{F}_n) , and $(C_n)_{n \in \mathbb{N}}$ is a predictable sequence of real-valued random variables. For example, we may think of C_n as the stake in the n -th round of a fair game, and of the martingale increment $M_n - M_{n-1}$ as the net gain (resp. loss) per unit stake. In this case, the capital I_n of a player with gambling strategy (C_n) after n rounds is given recursively by

$$I_n = I_{n-1} + C_n \cdot (M_n - M_{n-1}) \quad \text{for any } n \in \mathbb{N},$$

i.e.,

$$I_n = I_0 + \sum_{k=1}^n C_k \cdot (M_k - M_{k-1}).$$

Definition (Martingale transform). *The stochastic process $C \bullet M$ defined by*

$$(C \bullet M)_n := \sum_{k=1}^n C_k \cdot (M_k - M_{k-1}) \quad \text{for any } n \geq 0,$$

*is called the **martingale transform** of the martingale $(M_n)_{n \geq 0}$ w.r.t. the predictable sequence $(C_k)_{k \geq 1}$, or the discrete stochastic integral of (C_n) w.r.t. (M_n) .*

The process $C \bullet M$ is a time-discrete version of the stochastic integral $\int_0^t C_s dM_s$ for continuous-time processes C and M , cf. [Introduction to Stochastic Analysis].

Example (Martingale strategy). One origin of the word “martingale” is the name of a well-known gambling strategy: In a standard coin-tossing game, the stake is doubled each time a loss occurs, and the player stops the game after the first time he wins. If the net gain in n rounds with unit stake is given by a standard Random Walk

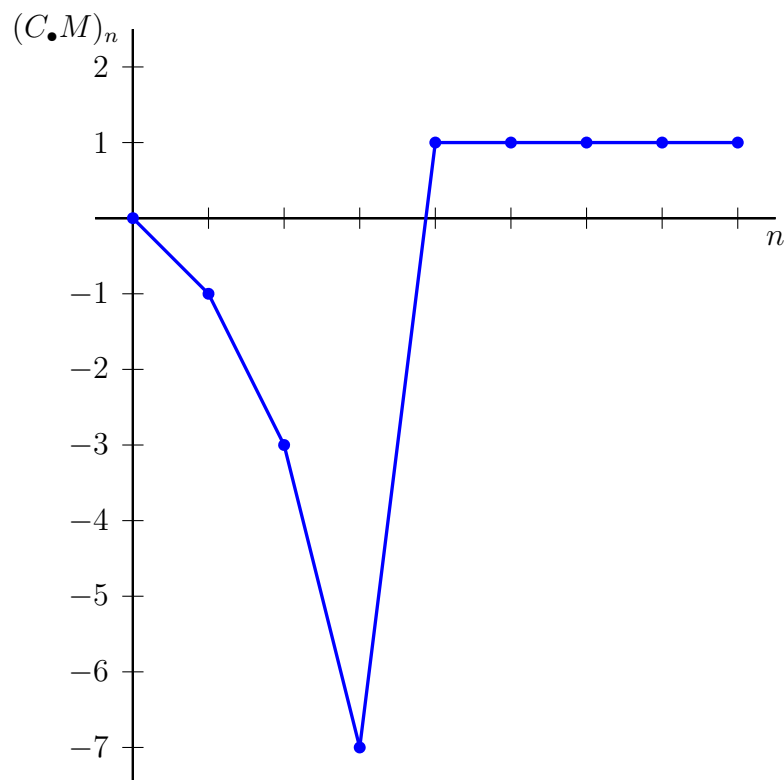
$$M_n = \eta_1 + \dots + \eta_n, \quad \eta_i \text{ i.i.d. with } P[\eta_i = 1] = P[\eta_i = -1] = 1/2,$$

then the stake in the n -th round is

$$C_n = 2^{n-1} \quad \text{if } \eta_1 = \dots = \eta_{n-1} = -1, \text{ and } \quad C_n = 0 \quad \text{otherwise.}$$

Clearly, with probability one, the game terminates in finite time, and at that time the player has always won one unit, i.e.,

$$P[(C \bullet M)_n = 1 \text{ eventually}] = 1.$$



At first glance this looks like a safe winning strategy, but of course this would only be the case, if the player had unlimited capital and time available.

Satz 3.6 (You can't beat the system!). (1). If $(M_n)_{n \geq 0}$ is an (\mathcal{F}_n) martingale, and $(C_n)_{n \geq 1}$ is predictable with $C_n \cdot (M_n - M_{n-1}) \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ for any $n \geq 1$, then $C_\bullet M$ is again an (\mathcal{F}_n) martingale.

(2). If (M_n) is an (\mathcal{F}_n) supermartingale and $(C_n)_{n \geq 1}$ is non-negative and predictable with $C_n \cdot (M_n - M_{n-1}) \in \mathcal{L}^1$ for any n , then $C_\bullet M$ is again a supermartingale.

Beweis. For $n \geq 1$ we have

$$\begin{aligned} E[(C_\bullet M)_n - (C_\bullet M)_{n-1} \mid \mathcal{F}_{n-1}] &= E[C_n \cdot (M_n - M_{n-1}) \mid \mathcal{F}_{n-1}] \\ &= C_n \cdot E[M_n - M_{n-1} \mid \mathcal{F}_{n-1}] = 0 \quad P\text{-a.s.} \end{aligned}$$

This proves the first part of the claim. The proof of the second part is similar. \square

The theorem shows that a fair game (a martingale) can not be transformed by choice of a clever gambling strategy into an unfair (or “superfair”) game. In models of financial markets this fact is crucial to exclude the existence of arbitrage possibilities (riskless profit).

Example (Martingale strategy, cont.). For the classical martingale strategy, we obtain

$$E[(C_\bullet M)_n] = E[(C_\bullet M)_0] = 0 \quad \text{for any } n \geq 0$$

by the martingale property, although

$$\lim_{n \rightarrow \infty} (C_\bullet M)_n = 1 \quad P\text{-a.s.}$$

This is a classical example showing that the assertion of the dominated convergence theorem may not hold if the assumptions are violated.

Remark. The integrability assumption in Theorem 3.6 is always satisfied if the random variables C_n are bounded, or if both C_n and M_n are square-integrable for any n .

Example (Financial market model with one risky asset). Suppose an investor is holding in the time interval $(n-1, n)$ Φ_n units of an asset with price S_n per unit at time n . We assume that (S_n) is an adapted and (Φ_n) is a predictable stochastic process w.r.t. a filtration (\mathcal{F}_n) . If the investor always puts his remaining capital onto a bank account with guaranteed interest rate r (“riskless asset”) then the change of his capital V_n during the time interval $(n-1, n)$ is given by

$$V_n = V_{n-1} + \Phi_n \cdot (S_n - S_{n-1}) + (V_{n-1} - \Phi_n \cdot S_{n-1}) \cdot r. \quad (3.3.1)$$

Considering the discounted quantity $\tilde{V}_n = V_n/(1+r)^n$, we obtain the equivalent recursion

$$\tilde{V}_n = \tilde{V}_{n-1} + \Phi_n \cdot (\tilde{S}_n - \tilde{S}_{n-1}) \quad \text{for any } n \geq 1. \quad (3.3.2)$$

In fact, (3.3.1) holds if and only if

$$V_n - (1+r)V_{n-1} = \Phi_n \cdot (S_n - (1+r)S_{n-1}),$$

which is equivalent to (3.3.2). Therefore, the discounted capital at time n is given by

$$\tilde{V}_n = V_0 + (\Phi_{\bullet} \tilde{S})_n.$$

By Theorem 3.6, we can conclude that, if the discounted price progress (\tilde{S}_n) is an (\mathcal{F}_n) martingale w.r.t. a given probability measure, then (\tilde{V}_n) is a martingale as well. In this case, assuming that V_0 is constant, we obtain in particular

$$E[\tilde{V}_n] = V_0,$$

or, equivalently,

$$E[V_n] = (1+r)^n V_0 \quad \text{for any } n \geq 0. \quad (3.3.3)$$

This fact, together with the existence of a martingale measure, can now be used for option pricing under a **no-arbitrage assumption**. To this end we assume that the payoff of an option at time N is given by an (\mathcal{F}_N) -measurable random variable F . For example, the payoff of a European call option with strike price K based on the asset with price process (S_n) is $S_N - K$ if the price S_N at maturity exceeds K , and 0 otherwise, i.e.,

$$F = (S_N - K)^+.$$

Suppose further that the option can be *replicated by a hedging strategy* (Φ_n) , i.e., there exists a \mathcal{F}_0 -measurable random variable V_0 and a predictable sequence of random variables $(\Phi_n)_{1 \leq n \leq N}$ such that

$$F = V_N$$

is the value at time N of a portfolio with initial value V_0 w.r.t. the trading strategy (Φ_n) .

Then, assuming the non-existence of arbitrage possibilities, the option price at time 0 has to be V_0 , since otherwise one could construct an arbitrage strategy by selling the option and investing money in the stock market with strategy (Φ_n) , or conversely. Therefore, if a martingale measure exists (i.e., an underlying probability measure such that the discounted stock price (\tilde{S}_n) is a martingale), then the no-arbitrage price of the option at time 0 can be computed by (3.3.3) where the expectation is taken w.r.t. the martingale measure.

The following exercise shows how this works out in the Cox-Ross-Rubinstein binomial model:

Exercise (No-Arbitrage Pricing in the CRR model). Consider the CRR binomial model, i.e., $\Omega = \{1 + a, 1 + b\}^N$ with $-1 < a < r < b < \infty$, $X_i(\omega_1, \dots, \omega_N) = \omega_i$, $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, and

$$S_n = S_0 \cdot \prod_{i=1}^n X_i, \quad n = 0, 1, \dots, N,$$

where S_0 is a constant.

- (1). *Completeness of the CRR model:* Prove that for any function $F : \Omega \rightarrow \mathbb{R}$ there exists a constant V_0 and a predictable sequence $(\Phi_n)_{1 \leq n \leq N}$ such that $F = V_N$ where $(V_n)_{1 \leq n \leq N}$ is defined by (3.3.1), or, equivalently,

$$\frac{F}{(1+r)^N} = \tilde{V}_N = V_0 + (\Phi \bullet \tilde{S})_N.$$

Hence in the CRR model, any \mathcal{F}_N -measurable function F can be replicated by a predictable trading strategy. Market models with this property are called *complete*.

Hint: Prove inductively that for $n = N, N-1, \dots, 0$, $\tilde{F} = F/(1+r)^N$ can be represented as

$$\tilde{F} = \tilde{V}_n + \sum_{i=n+1}^N \Phi_i \cdot (\tilde{S}_i - \tilde{S}_{i-1})$$

with an \mathcal{F}_n -measurable function \tilde{V}_n and a predictable sequence $(\Phi_i)_{n+1 \leq i \leq N}$.

- (2). *Option pricing:* Derive a general formula for the no-arbitrage price of an option with payoff function $F : \Omega \rightarrow \mathbb{R}$ in the CRR model. Compute the no-arbitrage price for a European call option with maturity N and strike K explicitly.

3.3.2 Stopped Martingales

One possible strategy for controlling a fair game is to terminate the game at a time depending on the previous development. Recall that a random variable $T : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$ is called a **stopping time** w.r.t. the filtration (\mathcal{F}_n) if and only if the event $\{T = n\}$ is contained in \mathcal{F}_n for any $n \geq 0$, or equivalently, iff $\{T \leq n\} \in \mathcal{F}_n$ for any $n \geq 0$.

We consider an (\mathcal{F}_n) -adapted stochastic process $(M_n)_{n \geq 0}$, and an (\mathcal{F}_n) -stopping time T on the probability space (Ω, \mathcal{A}, P) . The process stopped at time T is defined as $(M_{T \wedge n})_{n \geq 0}$ where

$$M_{T \wedge n}(\omega) = M_{T(\omega) \wedge n}(\omega) = \begin{cases} M_n(\omega) & \text{for } n \leq T(\omega), \\ M_{T(\omega)}(\omega) & \text{for } n \geq T(\omega). \end{cases}$$

For example, the process stopped at a hitting time T_A gets stuck at the first time it enters the set A .

Satz 3.7 (Optional Stopping Theorem, Version 1). *If $(M_n)_{n \geq 0}$ is a martingale (resp. a supermartingale) w.r.t. (\mathcal{F}_n) , and T is an (\mathcal{F}_n) -stopping time, then the stopped process $(M_{T \wedge n})_{n \geq 0}$ is again an (\mathcal{F}_n) -martingale (resp. supermartingale). In particular, we have*

$$E[M_{T \wedge n}] \stackrel{(\leq)}{=} E[M_0] \quad \text{for any } n \geq 0.$$

Beweis. Consider the following strategy:

$$C_n = I_{\{T \geq n\}} = 1 - I_{\{T \leq n-1\}},$$

i.e., we put a unit stake in each round before time T and quit playing at time T . Since T is a stopping time, the sequence (C_n) is predictable. Moreover,

$$M_{T \wedge n} - M_0 = (C \bullet M)_n \quad \text{for any } n \geq 0. \quad (3.3.4)$$

In fact, for the increments of the stopped process we have

$$M_{T \wedge n} - M_{T \wedge (n-1)} = \begin{cases} M_n - M_{n-1} & \text{if } T \geq n \\ 0 & \text{if } T \leq n-1 \end{cases} = C_n \cdot (M_n - M_{n-1}),$$

and (3.3.4) follows by summing over n . Since the sequence (C_n) is predictable, bounded and non-negative, the process $C \bullet M$ is a martingale, supermartingale respectively, provided the same holds for M . \square

Remark (IMPORTANT). (1). In general, it is NOT TRUE under the assumptions in Theorem 3.7 that

$$E[M_T] = E[M_0], \quad E[M_T] \leq E[M_0] \quad \text{respectively.} \quad (3.3.5)$$

Suppose for example that (M_n) is the classical Random Walk starting at 0 and $T = T_{\{1\}}$ is the first hitting time of the point 1. Then, by recurrence of the Random Walk, $T < \infty$ and $M_T = 1$ hold almost surely although $M_0 = 0$.

(2). If, on the other hand, T is a *bounded stopping time*, then there exists $n \in \mathbb{N}$ such that $T(\omega) \leq n$ for any ω . In this case, the optional stopping theorem implies

$$E[M_T] = E[M_{T \wedge n}] \stackrel{(\leq)}{=} E[M_0].$$

Example (Classical Ruin Problem). Let $a, b, x \in \mathbb{Z}$ with $a < x < b$. We consider the classical Random Walk

$$X_n = x + \sum_{i=1}^n \eta_i, \quad \eta_i \text{ i.i.d. with } P[\eta_i = \pm 1] = \frac{1}{2},$$

with initial value $X_0 = x$. We now show how to apply the optional stopping theorem to compute the distributions of the exit time

$$T(\omega) = \min\{n \geq 0 : X_n(\omega) \notin (a, b)\},$$

and the exit point X_T . These distributions can also be computed by more traditional methods (first step analysis, reflection principle), but martingales yield an elegant and general approach.

(1). *Ruin probability* $r(x) = P[X_T = a]$.

The process (X_n) is a martingale w.r.t. the filtration $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$, and $T < \infty$ almost surely holds by elementary arguments. As the stopped process $X_{T \wedge n}$ is bounded ($a \leq X_{T \wedge n} \leq b$), we obtain

$$x = E[X_0] = E[X_{T \wedge n}] \xrightarrow{n \rightarrow \infty} E[X_T] = a \cdot r(x) + b \cdot (1 - r(x))$$

by the Optional Stopping Theorem and the Dominated Convergence Theorem. Hence

$$r(x) = \frac{b - x}{a - x}. \quad (3.3.6)$$

(2). *Mean exit time from (a, b) .*

To compute the expectation value $E[T]$, we apply the Optional Stopping Theorem to the (\mathcal{F}_n) martingale

$$M_n := X_n^2 - n.$$

By monotone and dominated convergence, we obtain

$$\begin{aligned} x^2 &= E[M_0] = E[M_{T \wedge n}] = E[X_{T \wedge n}^2] - E[T \wedge n] \\ &\xrightarrow{n \rightarrow \infty} E[X_T^2] - E[T]. \end{aligned}$$

Therefore, by (3.3.6),

$$\begin{aligned} E[T] &= E[X_T^2] - x^2 = a^2 \cdot r(x) + b^2 \cdot (1 - r(x)) - x^2 \\ &= (b - x) \cdot (x - a). \end{aligned} \quad (3.3.7)$$

(3). *Mean passage time of b is infinite.*

The first passage time $T_b = \min\{n \geq 0 : X_n = b\}$ is greater or equal than the exit time from the interval (a, b) for any $a < x$. Thus by (3.3.7), we have

$$E[T_b] \geq \lim_{a \rightarrow -\infty} (b - x) \cdot (x - a) = \infty,$$

i.e., T_b is **not integrable!** These and some other related passage times are important examples of random variables with a heavy-tailed distribution and infinite first moment.

(4). *Distribution of passage times.*

We now compute the distribution of the first passage time T_b explicitly in the case $x = 0$ and $b = 1$. Hence let $T = T_1$. As shown above, the process

$$M_n^\lambda := e^{\lambda X_n} / (\cosh \lambda)^n, \quad n \geq 0,$$

is a martingale for each $\lambda \in \mathbb{R}$. Now suppose $\lambda > 0$. By the Optional Stopping Theorem,

$$1 = E[M_0^\lambda] = E[M_{T \wedge n}^\lambda] = E[e^{\lambda X_{T \wedge n}} / (\cosh \lambda)^{T \wedge n}] \quad (3.3.8)$$

for any $n \in \mathbb{N}$. As $n \rightarrow \infty$, the integrands on the right hand side converge to $e^\lambda (\cosh \lambda)^{-T} \cdot I_{\{T < \infty\}}$. Moreover, they are uniformly bounded by e^λ , since $X_{T \wedge n} \leq 1$ for any n . Hence by the Dominated Convergence Theorem, the expectation on the right hand side of (3.3.8) converges to $E[e^\lambda / (\cosh \lambda)^T ; T < \infty]$, and we obtain the identity

$$E[(\cosh \lambda)^{-T} ; T < \infty] = e^{-\lambda} \quad \text{for any } \lambda > 0. \quad (3.3.9)$$

Taking the limit as $\lambda \searrow 0$, we see that $P[T < \infty] = 1$. Taking this into account, and substituting $s = 1 / \cosh \lambda$ in (3.3.9), we can now compute the generating function of T explicitly:

$$E[s^T] = e^{-\lambda} = (1 - \sqrt{1 - s^2}) / s \quad \text{for any } s \in (0, 1). \quad (3.3.10)$$

Developing both sides into a power series finally yields

$$\sum_{n=0}^{\infty} s^n \cdot P[T = n] = \sum_{m=1}^{\infty} (-1)^{m+1} \binom{1/2}{m} s^{2m-1}.$$

Therefore, the distribution of the first passage time of 1 is given by $P[T = 2m] = 0$ and

$$P[T = 2m - 1] = (-1)^{m+1} \binom{1/2}{m} = (-1)^{m+1} \cdot \frac{1}{2} \cdot \left(-\frac{1}{2}\right) \cdots \left(\frac{1}{2} - m + 1\right) / m!$$

for any $m \geq 1$.

3.3.3 Optional Stopping Theorems

Stopping times occurring in applications are typically not bounded, see the example above. Therefore, we need more general conditions guaranteeing that (3.3.5) holds nevertheless. A first general criterion is obtained by applying the Dominated Convergence Theorem:

Satz 3.8 (Optional Stopping Theorem, Version 2). *Suppose that (M_n) is a martingale w.r.t. (\mathcal{F}_n) , T is an (\mathcal{F}_n) -stopping time with $P[T < \infty] = 1$, and there exists a random variable $Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ such that*

$$|M_{T \wedge n}| \leq Y \quad P\text{-almost surely for any } n \in \mathbb{N}.$$

Then

$$E[M_T] = E[M_0].$$

Beweis. Since $P[T < \infty] = 1$, we have

$$M_T = \lim_{n \rightarrow \infty} M_{T \wedge n} \quad P\text{-almost surely.}$$

By Theorem 3.7, $E[M_0] = E[M_{T \wedge n}]$, and by the Dominated Convergence Theorem, $E[M_{T \wedge n}] \rightarrow E[M_T]$ as $n \rightarrow \infty$. \square

Remark (Weakening the assumptions). Instead of the existence of an integrable random variable Y dominating the random variables $M_{T \wedge n}$, $n \in \mathbb{N}$, it is enough to assume that these random variables are **uniformly integrable**, i.e.,

$$\sup_{n \in \mathbb{N}} E[|M_{T \wedge n}|; |M_{T \wedge n}| \geq c] \rightarrow 0 \quad \text{as } c \rightarrow \infty.$$

For non-negative supermartingales, we can apply Fatou's Lemma instead of the Dominated Convergence Theorem to pass to the limit as $n \rightarrow \infty$ in the Stopping Theorem. The advantage is that no integrability assumption is required. Of course, the price to pay is that we only obtain an inequality:

Satz 3.9 (Optional Stopping Theorem, Version 3). *If (M_n) is a non-negative supermartingale w.r.t. (\mathcal{F}_n) , then*

$$E[M_0] \geq E[M_T; T < \infty]$$

holds for any (\mathcal{F}_n) stopping time T .

Beweis. Since $M_T = \lim_{n \rightarrow \infty} M_{T \wedge n}$ on $\{T < \infty\}$, and $M_T \geq 0$, Theorem 3.7 combined with Fatou's Lemma implies

$$E[M_0] \geq \liminf_{n \rightarrow \infty} E[M_{T \wedge n}] \geq E\left[\liminf_{n \rightarrow \infty} M_{T \wedge n}\right] \geq E[M_T; T < \infty].$$

□

Example (Markov chains and Dirichlet problem). Suppose that w.r.t. the probability measure P_x , the process (X_n) is a time-homogeneous Markov chain with measurable state space (S, \mathcal{S}) , transition kernel p , and start in x . Let $D \in \mathcal{S}$ be a measurable subset of the state space, and $f : D^C \rightarrow \mathbb{R}$ a measurable function (the given “boundary values”), and let

$$T = \min\{n \geq 0 : X_n \in D^C\}$$

denote the first exit time of the Markov chain from D . By conditioning on the first step of the Markov chain, we have shown above that if f is non-negative or bounded, then the function

$$h(x) = E_x[f(X_T); T < \infty], \quad (x \in S),$$

is a solution of the *Dirichlet problem*

$$\begin{aligned} (ph)(x) &= h(x) && \text{for } x \in D, \\ h(x) &= f(x) && \text{for } x \in D^C. \end{aligned}$$

By considering the martingale $h(X_{T \wedge n})$ for a function h that is harmonic on D , we obtain a converse statement:

Exercise (Uniqueness of the Dirichlet problem). Suppose that $P_x[T < \infty] = 1$ for any $x \in S$.

- (1). Prove that $h(X_{T \wedge n})$ is a martingale w.r.t. P_x for any bounded solution h of the Dirichlet problem and any $x \in S$.
- (2). Conclude that if f is bounded, then

$$h(x) = E_x[f(X_T)] \tag{3.3.11}$$

is the unique bounded solution of the Dirichlet problem.

- (3). Similarly, show that for any non-negative f , the function h defined by (3.3.11) is the minimal non-negative solution of the Dirichlet problem.

We finally state a version of the Optional Stopping Theorem that applies in particular to martingales with bounded increments:

Korollar 3.10 (Optional Stopping for martingales with bounded increments). *Suppose that (M_n) is an (\mathcal{F}_n) martingale, and there exists a finite constant $K \in (0, \infty)$ such that*

$$E[|M_{n+1} - M_n| \mid \mathcal{F}_n] \leq K \quad P\text{-almost surely for any } n \geq 0. \quad (3.3.12)$$

Then for any (\mathcal{F}_n) stopping time T with $E[T] < \infty$, we have

$$E[M_T] = E[M_0].$$

Beweis. For any $n \geq 0$,

$$|M_{T \wedge n}| \leq |M_0| + \sum_{i=0}^{\infty} |M_{i+1} - M_i| \cdot I_{\{T > i\}}.$$

Let Y denote the expression on the right hand side. We will show that Y is an integrable random variable – this implies the assertion by Theorem 3.8. To verify integrability of Y note that the event $\{T > i\}$ is contained in \mathcal{F}_i for any $i \geq 0$ since T is a stopping time. Therefore and by (3.3.12),

$$E[|M_{i+1} - M_i|; T > i] = E[E[|M_{i+1} - M_i| \mid \mathcal{F}_i]; T > i] \leq k \cdot P[T > i].$$

Summing over i , we obtain

$$E[Y] \leq E[|M_0|] + k \cdot \sum_{i=0}^{\infty} P[T > i] = E[|M_0|] + k \cdot E[T] < \infty$$

by the assumptions. □

Exercise (Integrability of stopping times). Prove that the expectation value $E[T]$ of a stopping time T is finite if there exist constants $\varepsilon > 0$ and $k \in \mathbb{N}$ such that

$$P[T \leq n + k \mid \mathcal{F}_n] > \varepsilon \quad P\text{-a.s. for any } n \in \mathbb{N}.$$

3.3.4 Wald's identity for random sums

We finally apply the Optional Stopping Theorem to sums of independent random variables with a random number T of summands. The point is that we do not assume that T is independent of the summands but only that it is a stopping time w.r.t. the filtration generated by the summands.

Let $S_n = \eta_1 + \dots + \eta_n$ with i.i.d. random variables $\eta_i \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. Denoting by m the expectation values of the increments η_i , the process

$$M_n = S_n - n \cdot m$$

is a martingale w.r.t. $\mathcal{F}_n = \sigma(\eta_1, \dots, \eta_n)$. By applying Corollary 3.10 to this martingale, we obtain:

Satz 3.11 (Wald's identity). *Suppose that T is an (\mathcal{F}_n) stopping time with $E[T] < \infty$. Then*

$$E[S_T] = m \cdot E[T].$$

Beweis. For any $n \geq 0$, we have

$$E[|M_{n+1} - M_n| | \mathcal{F}_n] = E[|\eta_{n+1} - m| | \mathcal{F}_n] = E[|\eta_{n+1} - m|]$$

by the independence of the η_i . As the η_i are identically distributed and integrable, the right hand side is a finite constant. Hence Corollary 3.10 applies, and we obtain

$$0 = E[M_0] = E[M_t] = E[S_t] - m \cdot E[T].$$

□

3.4 Almost sure convergence of supermartingales

The strength of martingale theory is partially due to powerful general convergence theorems that hold for martingales, sub- and supermartingales. Let $(Z_n)_{n \geq 0}$ be a discrete-parameter supermartingale w.r.t. a filtration $(\mathcal{F}_n)_{n \geq 0}$ on a probability space (Ω, \mathcal{A}, P) . The following theorem yields a stochastic counterpart to the fact that any lower bounded decreasing sequence of reals converges to a finite limit:

Satz 3.12 (Supermartingale Convergence Theorem, Doob). *If $\sup_{n \geq 0} E[Z_n^-] < \infty$ then (Z_n) converges almost surely to an integrable random variable $Z_\infty \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$. In particular, supermartingales that are uniformly bounded from above converge almost surely to an integrable random variable.*

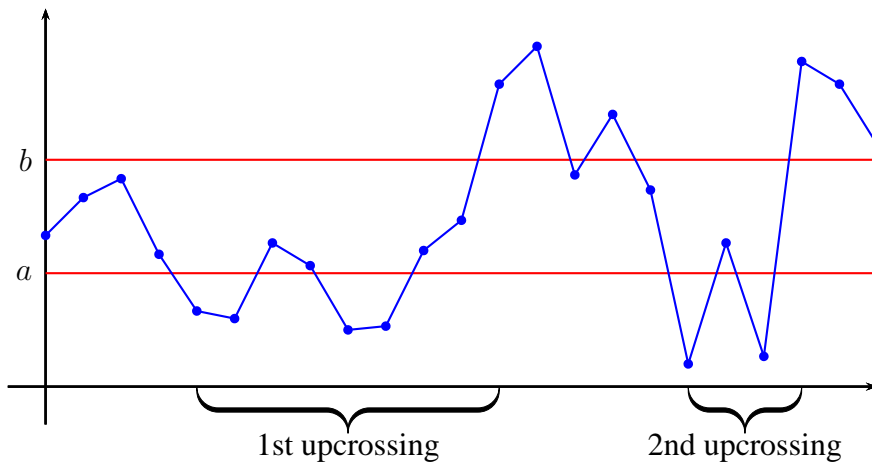
Remark (L^1 boundedness and L^1 convergence).

- (1). Although the limit is integrable, L^1 convergence does **not** hold in general.

(2). The condition $\sup E[Z_n^-] < \infty$ holds if and only if (Z_n) is bounded in L^1 . Indeed, as $E[Z_n^+] < \infty$ by our definition of a supermartingale, we have

$$E[|Z_n|] = E[Z_n] + 2E[Z_n^-] \leq E[Z_0] + 2E[Z_n^-] \quad \text{for any } n \geq 0.$$

For proving the Supermartingale Convergence Theorem, we introduce the number $U^{(a,b)}(\omega)$ of upcrossings over an interval (a, b) by the sequence $Z_n(\omega)$, cf. below for the exact definition.



Note that if $U^{(a,b)}(\omega)$ is finite for any non-empty bounded interval (a, b) then $\limsup Z_n(\omega)$ and $\liminf Z_n(\omega)$ coincide, i.e., the sequence $(Z_n(\omega))$ converges. Therefore, to show almost sure convergence of (Z_n) , we derive an upper bound for $U^{(a,b)}$. We first prove this key estimate and then complete the proof of the theorem.

3.4.1 Doob's upcrossing inequality

For $n \in \mathbb{N}$ and $a, b \in \mathbb{R}$ with $a < b$ we define the number $U_n^{(a,b)}$ of upcrossings over the interval (a, b) before time n by

$$U_n^{(a,b)} = \max\{k \geq 0 : \exists 0 \leq s_1 < t_1 < s_2 < t_2 \dots < s_k < t_k \leq n : Z_{s_i} \leq a, Z_{t_i} \geq b\}.$$

Lemma 3.13 (Doob). *If (Z_n) is a supermartingale then*

$$(b - a) \cdot E[U_n^{(a,b)}] \leq E[(Z_n - a)^-] \quad \text{for any } a < b \text{ and } n \geq 0.$$

Beweis. We may assume $E[Z_n^-] < \infty$ since otherwise there is nothing to prove. The key idea is to set up a predictable gambling strategy that increases our capital by $(b - a)$ for each completed upcrossing. Since the net gain with this strategy should again be a supermartingale this yields an upper bound for the average number of upcrossings. Here is the strategy:

- repeat
- Wait until $Z_k \leq a$.
 - Then play unit stakes until $Z_k \geq b$.
 -

The stake C_k in round k is

$$C_1 = \begin{cases} 1 & \text{if } Z_0 \leq a, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$C_k = \begin{cases} 1 & \text{if } (C_{k-1} = 1 \text{ and } Z_{k-1} \leq b) \text{ or } (C_{k-1} = 0 \text{ and } Z_{k-1} \leq a), \\ 0 & \text{otherwise.} \end{cases}$$

Clearly, (C_k) is a predictable, bounded and non-negative sequence of random variables. Moreover, $C_k \cdot (Z_k - Z_{k-1})$ is integrable for any $k \leq n$, because C_k is bounded and

$$E[|Z_k|] = 2E[Z_k^+] - E[Z_k] \leq 2E[Z_k^+] - E[Z_n] \leq 2E[Z_k^+] - E[Z_n^-]$$

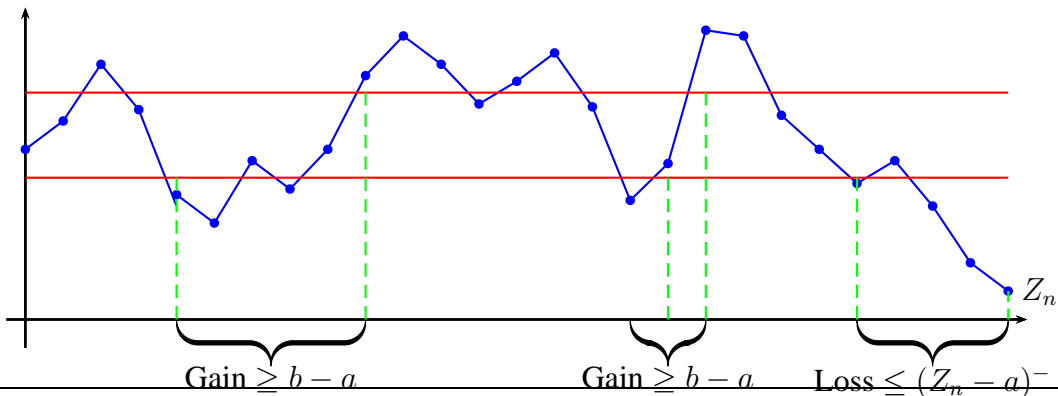
for $k \leq n$. Therefore, by Theorem 3.6 and the remark below, the process

$$(C \bullet Z)_k = \sum_{i=1}^k C_i \cdot (Z_i - Z_{i-1}), \quad 0 \leq k \leq n,$$

is again a supermartingale.

Clearly, the value of the process $C \bullet Z$ increases by at least $(b - a)$ units during each completed upcrossing. Between upcrossing periods, the value of $(C \bullet Z)_k$ is constant. Finally, if the final time n is contained in an upcrossing period, then the process can decrease by at most $(Z_n - a)^-$ units during that last period (since Z_k might decrease before the next upcrossing is completed). Therefore, we have

$$\begin{aligned} (C \bullet Z)_n &\geq (b - a) \cdot U_n^{(a,b)} - (Z_n - a)^-, \quad \text{i.e.,} \\ (b - a) \cdot U_n^{(a,b)} &\leq (C \bullet Z)_n + (Z_n - a)^-. \end{aligned}$$



Since $C \bullet Z$ is a supermartingale with initial value 0, we obtain the upper bound

$$(b - a)E[U_n^{(a,b)}] \leq E[(C \bullet Z)_n] + E[(Z_n - a)^-] \leq E[(Z_n - a)^-].$$

□

3.4.2 Proof of Doob's Convergence Theorem

We can now complete the proof of Theorem 3.12.

Beweis. Let

$$U^{(a,b)} = \sup_{n \in \mathbb{N}} U_n^{(a,b)}$$

denote the total number of upcrossings of the supermartingale (Z_n) over an interval (a, b) with $-\infty < a < b < \infty$. By the upcrossing inequality and monotone convergence,

$$E[U^{(a,b)}] = \lim_{n \rightarrow \infty} E[U_n^{(a,b)}] \leq \frac{1}{b - a} \cdot \sup_{n \in \mathbb{N}} E[(Z_n - a)^-]. \quad (3.4.1)$$

Assuming $\sup E[Z_n^-] < \infty$, the right hand side of (3.4.1) is finite since $(Z_n - a)^- \leq |a| + Z_n^-$. Therefore,

$$U^{(a,b)} < \infty \quad P\text{-almost surely,}$$

and hence the event

$$\{\liminf Z_n \neq \limsup Z_n\} = \bigcup_{\substack{a, b \in \mathbb{Q} \\ a < b}} \{U^{(a,b)} = \infty\}$$

has probability zero. This proves almost sure convergence.

It remains to show that the almost sure limit $Z_\infty = \lim Z_n$ is an integrable random variable (in particular, it is finite almost surely). This holds true as, by the remark below Theorem 3.12, $\sup E[Z_n^-] < \infty$ implies that (Z_n) is bounded in L^1 , and therefore

$$E[|Z_\infty|] = E[\lim |Z_n|] \leq \liminf E[|Z_n|] < \infty$$

by Fatou's lemma. □

3.4.3 Examples and first applications

We now consider a few prototypic applications of the almost sure convergence theorem:

Example (Sums of i.i.d. random variables). Consider a Random Walk

$$S_n = \sum_{i=1}^n \eta_i$$

on \mathbb{R} with centered and bounded increments:

$$\eta_i \text{ i.i.d. with } |\eta_i| \leq c \text{ and } E[\eta_i] = 0, \quad c \in \mathbb{R}.$$

Suppose that $P[\eta_i \neq 0] > 0$. Then there exists $\varepsilon > 0$ such that $P[|\eta_i| \geq \varepsilon] > 0$. As the increments are i.i.d., the event $\{|\eta_i| \geq \varepsilon\}$ occurs infinitely often with probability one. Therefore, almost surely the martingale (S_n) does not converge as $n \rightarrow \infty$.

Now let $a \in \mathbb{R}$. We consider the first hitting time

$$T_a = \min\{n \geq 0 : S_n \geq a\}$$

of the interval $[a, \infty)$. By the Optional Stopping Theorem, the stopped Random Walk $(S_{T_a \wedge n})_{n \geq 0}$ is again a martingale. Moreover, as $S_k < a$ for any $k < T_a$ and the increments are bounded by c , we obtain the upper bound

$$S_{T_a \wedge n} < a + c \quad \text{for any } n \in \mathbb{N}.$$

Therefore, the stopped Random Walk converges almost surely by the Supermartingale Convergence Theorem. As (S_n) does not converge, we can conclude that $P[T_a < \infty] = 1$ for any $a > 0$, i.e.,

$$\limsup S_n = \infty \quad \text{almost surely.}$$

Since (S_n) is also a submartingale, we obtain

$$\liminf S_n = -\infty \quad \text{almost surely}$$

by an analogue argument. A generalization of this result is given in Theorem 3.14 below.

Remark (Almost sure vs. L^p convergence). In the last example, the stopped process does not converge in L^p for any $p \in [1, \infty)$. In fact,

$$\lim_{n \rightarrow \infty} E[S_{T_a \wedge n}] = E[S_{T_a}] \geq a \quad \text{whereas} \quad E[S_0] = 0.$$

Example (Products of non-negative i.i.d. random variables). Consider a growth process

$$Z_n = \prod_{i=1}^n Y_i$$

with i.i.d. factors $Y_i \geq 0$ with finite expectation $\alpha \in (0, \infty)$. Then

$$M_n = Z_n / \alpha^n$$

is a martingale. By the almost sure convergence theorem, a finite limit M_∞ exists almost surely, because $M_n \geq 0$ for all n . For the almost sure asymptotics of (Z_n) , we distinguish three different cases:

(1). $\alpha < 1$ (*subcritical*): In this case,

$$Z_n = M_n \cdot \alpha^n$$

converges to 0 exponentially fast with probability one.

(2). $\alpha = 1$ (*critical*): Here (Z_n) is a martingale and converges almost surely to a finite limit. If $P[Y_i \neq 1] > 0$ then there exists $\varepsilon > 0$ such that $Y_i \geq 1 + \varepsilon$ infinitely often with probability one. This is consistent with convergence of (Z_n) only if the limit is zero. Hence, if (Z_n) is not almost surely constant, then also in the critical case $Z_n \rightarrow 0$ almost surely.

(3). $\alpha > 1$ (*supercritical*): In this case, on the set $\{M_\infty > 0\}$,

$$Z_n = M_n \cdot \alpha^n \sim M_\infty \cdot \alpha^n,$$

i.e., (Z_n) grows exponentially fast. The asymptotics on the set $\{M_\infty = 0\}$ is not evident and requires separate considerations depending on the model.

Although most of the conclusions in the last example could have been obtained without martingale methods (e.g. by taking logarithms), the martingale approach has the advantage of carrying over to far more general model classes. These include for example branching processes or exponentials of continuous time processes.

Example (Boundary behaviour of harmonic functions). Let $D \subseteq \mathbb{R}^d$ be a bounded open domain, and let $h : D \rightarrow \mathbb{R}$ be a harmonic function on D that is bounded from below:

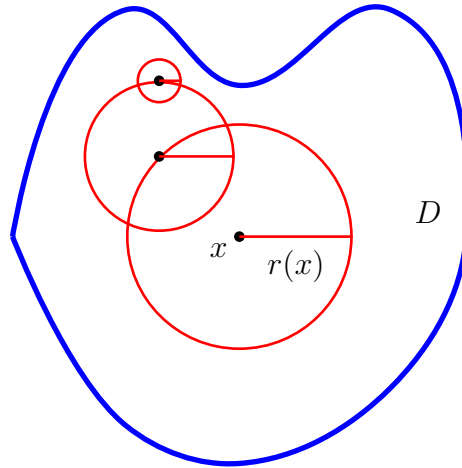
$$\Delta h(x) = 0 \quad \text{for any } x \in D, \quad \inf_{x \in D} h(x) > -\infty. \quad (3.4.2)$$

To study the asymptotic behavior of $h(x)$ as x approaches the boundary ∂D , we construct a Markov chain (X_n) such that $h(X_n)$ is a martingale: Let $r : D \rightarrow (0, \infty)$ be a continuous function such that

$$0 < r(x) < \text{dist}(x, \partial D) \quad \text{for any } x \in D, \quad (3.4.3)$$

and let (X_n) w.r.t P_x denote the canonical time-homogeneous Markov chain with state space D , initial value x , and transition probabilities

$$p(x, dy) = \text{Uniform distribution on the sphere } \{y \in \mathbb{R}^d : |y - x| = r(x)\}.$$



By (3.4.3), the function h is integrable w.r.t. $p(x, dy)$, and, by the mean value property,

$$(ph)(x) = h(x) \quad \text{for any } x \in D.$$

Therefore, the process $h(X_n)$ is a martingale w.r.t. P_x for each $x \in D$. As $h(X_n)$ is lower bounded by (3.4.2), the limit as $n \rightarrow \infty$ exists P_x -almost surely by the Supermartingale Convergence Theorem. In particular, since the coordinate functions $x \mapsto x_i$ are also harmonic and lower bounded on \overline{D} , the limit $X_\infty = \lim X_n$ exists P_x -almost surely. Moreover, X_∞ is in ∂D , because r is bounded from below by a strictly positive constant on any compact subset of D .

Summarizing we have shown:

- (1). *Boundary regularity:* If h is harmonic and bounded from below on D then the limit $\lim_{n \rightarrow \infty} h(X_n)$ exists along almost every trajectory X_n to the boundary ∂D .
- (2). *Representation of h in terms of boundary values:* If h is continuous on \overline{D} , then $h(X_n) \rightarrow h(X_\infty)$ P_x -almost surely and hence

$$h(x) = \lim_{n \rightarrow \infty} E_x[h(X_n)] = E[h(X_\infty)],$$

i.e., the distribution of X_∞ w.r.t. P_x is the harmonic measure on ∂D .

Note that, in contrast to classical results from analysis, the first statement holds without any smoothness condition on the boundary ∂D . Thus, although boundary values of h may not exist in the classical sense, they still do exist along almost every trajectory of the Markov chain!

3.4.4 Generalized Borel-Cantelli Lemma

Another application of the almost sure Convergence Theorem is a generalization of the Borel-Cantelli lemmas. We first prove a dichotomy for the asymptotic behavior of martingales with L^1 -bounded increments:

Satz 3.14 (Asymptotics of martingales with L^1 bounded increments). *Suppose that (M_n) is a martingale, and there exists an integrable random variable Y such that*

$$|M_n - M_{n-1}| \leq Y \quad \text{for any } n \in \mathbb{N}.$$

Then for P -almost every ω , the following dichotomy holds:

Either: *The limit $\lim_{n \rightarrow \infty} M_n(\omega)$ exists in \mathbb{R} ,*

or: $\limsup_{n \rightarrow \infty} M_n(\omega) = +\infty$ *and* $\liminf_{n \rightarrow \infty} M_n(\omega) = -\infty$.

The theorem and its proof are a generalization of Example 1 above.

Beweis. For $a \in (-\infty, 0)$ let $T_a = \min\{n \geq 0 : M_n \geq a\}$. By the Optional Stopping Theorem, $(M_{T_a \wedge n})$ is a martingale. Moreover,

$$M_{T_a \wedge n} \geq \min(M_0, a - Y) \quad \text{for any } n \geq 0,$$

and the right hand side is an integrable random variable. Therefore, (M_n) converges almost surely on $\{T_a = \infty\}$. Since this holds for any $a < 0$, we obtain almost sure convergence on the set

$$\{\liminf M_n > -\infty\} = \bigcup_{\substack{a < 0 \\ a \in \mathbb{Q}}} \{T_a = \infty\}.$$

Similarly, almost sure convergence follows on the set $\{\limsup M_n < \infty\}$. □

Now let $(\mathcal{F}_n)_{n \geq 0}$ be an arbitrary filtration. As a consequence of Theorem 3.14 we obtain:

Korollar 3.15 (Generalized Borel-Cantelli Lemma). *If (A_n) is a sequence of events with $A_n \in \mathcal{F}_n$ for any n , then the equivalence*

$$\omega \in A_n \quad \text{infinitely often} \quad \iff \quad \sum_{n=1}^{\infty} P[A_n | \mathcal{F}_{n-1}](\omega) = \infty$$

holds for almost every $\omega \in \Omega$.

Beweis. Let $S_n = \sum_{k=1}^n I_{A_k}$ and $T_n = \sum_{k=1}^n E[I_{A_k} | \mathcal{F}_{k-1}]$. Then S_n and T_n are almost surely increasing sequences. Let $S_\infty = \sup S_n$ and $T_\infty = \sup T_n$ denote the limits on $[0, \infty]$. The claim is that almost surely,

$$S_\infty = \infty \iff T_\infty = \infty. \quad (3.4.4)$$

To prove (3.4.4) we note that $S_n - T_n$ is a martingale with bounded increments. Therefore, almost surely, $S_n - T_n$ converges to a finite limit, or both $\limsup (S_n - T_n) = \infty$ and $\liminf (S_n - T_n) = -\infty$. In the first case, (3.4.4) holds. In the second case, $S_\infty = \infty$ and $T_\infty = \infty$, so (3.4.4) holds, too. \square

The assertion of Corollary 3.15 generalizes both classical Borel-Cantelli Lemmas: If (A_n) is an arbitrary sequence of events in a probability space (Ω, \mathcal{A}, P) then we can consider the filtration $\mathcal{F}_n = \sigma(A_1, \dots, A_n)$. By Corollary 3.15 we obtain:

1st Borel-Cantelli Lemma: If $\sum P[A_n] < \infty$ then $\sum P[A_n | \mathcal{F}_{n-1}] < \infty$ almost surely, and therefore

$$P[A_n \text{ infinitely often}] = 0.$$

2nd Borel-Cantelli Lemma: If $\sum P[A_n] = \infty$ and the events A_n are independent then we obtain $\sum P[A_n | \mathcal{F}_{n-1}] = \sum P[A_n] = \infty$ almost surely, and therefore

$$P[A_n \text{ infinitely often}] = 1.$$

Kapitel 4

Stochastische Prozesse in stetiger Zeit

Für viele Anwendungsprobleme sind Modelle, die auf Markovprozessen in kontinuierlicher Zeit basieren, natürlicher. Ändert der Prozess nur an abzählbar vielen zufälligen Zeitpunkten seinen Zustand, dann nennt man ihn eine zeitstetige Markovkette. Ein Markovprozess mit stetigen Pfaden heißt dagegen Diffusionsprozess.

Klassische Anwendungsbereiche zeitstetiger Markovketten sind die Modellierung von Warteschlangen und chemischen Reaktionen. Wir zeigen hier, wie man zeitstetige aus zeitdiskreten Markovketten konstruiert und beschreibt. Viele der Aussagen aus den letzten Abschnitten haben Entsprechungen im zeitstetigen Fall – wir verweisen dazu auf das einführende Lehrbuch [J. Norris: Markov Chains].

Der wichtigste Diffusionsprozess ist die Brownsche Bewegung, die sich ausgehend vom zentralen Grenzwertsatz als universeller zeitstetiger Skalierungslimes von Random Walks mit quadratintegrierbaren Inkrementen ergibt. In der stochastischen Analysis konstruiert man andere Diffusionsprozesse über stochastische Differentialgleichungen aus der Brownschen Bewegung – mit zahlreichen Anwendungen z.B. in der Finanzmathematik, Physik und mathematischen Biologie, aber auch mit weitreichenden Konsequenzen für viele Bereiche der Mathematik.

4.1 From Random Walks to Brownian Motion

Brownian Motion is a diffusion process, i.e., a continuous-time Markov process $(B_t)_{t \geq 0}$ with continuous sample paths $t \mapsto B_t(\omega)$. In fact, it is the only nontrivial continuous-time process that is a Lévy process as well as a martingale and a Gaussian process. A rigorous construction of this process has been carried out first by N. Wiener in 1923. Already about 20 years earlier, related models had been introduced independently for financial markets by L. Bachelier [*Théorie de la*

spéculation, Ann. Sci. École Norm. Sup. 17, 1900], and for the velocity of molecular motion by A. Einstein [Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen, Annalen der Physik 17, 1905].

The fundamental rôle played by Brownian motion in stochastic analysis is due to the Central Limit Theorem. Similarly as the normal distribution arises as a universal scaling limit of standardized sums of independent, identically distributed, square integrable random variables, Brownian motion shows up as a universal scaling limit of Random Walks with square integrable increments.

To motivate the definition of Brownian motion below, we first briefly discuss discrete-time stochastic processes and possible continuous-time scaling limits on an informal level.

A standard approach to model stochastic dynamics in discrete time is to start from a sequence of random variables η_1, η_2, \dots defined on a common probability space (Ω, \mathcal{A}, P) . The random variables η_n describe the stochastic influences (*noise*) on the system. Often they are assumed to be *independent and identically distributed (i.i.d.)*. In this case the collection (η_n) is also called a **white noise**, whereas a **colored noise** is given by dependent random variables. A stochastic process X_n ($n = 0, 1, 2, \dots$) taking values in \mathbb{R}^d is then defined recursively on (Ω, \mathcal{A}, P) by

$$X_{n+1} = X_n + \Phi_{n+1}(X_n, \eta_{n+1}), \quad n = 0, 1, 2, \dots \quad (4.1.1)$$

Here the Φ_n are measurable maps describing the *random law of motion*. If X_0 and η_1, η_2, \dots are independent random variables, then the process (X_n) is a Markov chain with respect to P .

Now let us assume that the random variables η_n are independent and identically distributed taking values in \mathbb{R} , or, more generally, \mathbb{R}^d . The easiest type of a nontrivial stochastic dynamics as described above is the Random Walk $S_n = \sum_{i=1}^n \eta_i$ which satisfies

$$S_{n+1} = S_n + \eta_{n+1} \quad \text{for } n = 0, 1, 2, \dots$$

Since the noise random variables η_n are the increments of the Random Walk (S_n) , the law of motion (4.1.1) in the general case can be rewritten as

$$X_{n+1} - X_n = \Phi_{n+1}(X_n, S_{n+1} - S_n), \quad n = 0, 1, 2, \dots \quad (4.1.2)$$

This equation is a difference equation for (X_n) driven by the stochastic process (S_n) .

Our aim is to carry out a similar construction as above for stochastic dynamics in continuous time. The stochastic difference equation (4.1.2) will then eventually be replaced by a *stochastic differential equation (SDE)*. However, before even being able to think about how to write down and make sense of such an equation, we have to identify a continuous-time stochastic process

that takes over the rôle of the Random Walk. For this purpose, we first determine possible scaling limits of Random Walks when the time steps tend to 0. It will turn out that if the increments are square integrable and the size of the increments goes to 0 as the length of the time steps tends to 0, then by the Central Limit Theorem there is essentially only one possible limit process in continuous time: Brownian motion.

4.1.1 Central Limit Theorem

Suppose that $Y_{n,i} : \Omega \rightarrow \mathbb{R}^d$, $1 \leq i \leq n < \infty$, are identically distributed, square-integrable random variables on a probability space (Ω, \mathcal{A}, P) such that $Y_{n,1}, \dots, Y_{n,n}$ are independent for each $n \in \mathbb{N}$. Then the rescaled sums

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_{n,i} - E[Y_{n,i}])$$

converge in distribution to a multivariate normal distribution $N(0, C)$ with covariance matrix

$$C_{kl} = \text{Cov}[Y_{n,i}^{(k)}, Y_{n,i}^{(l)}].$$

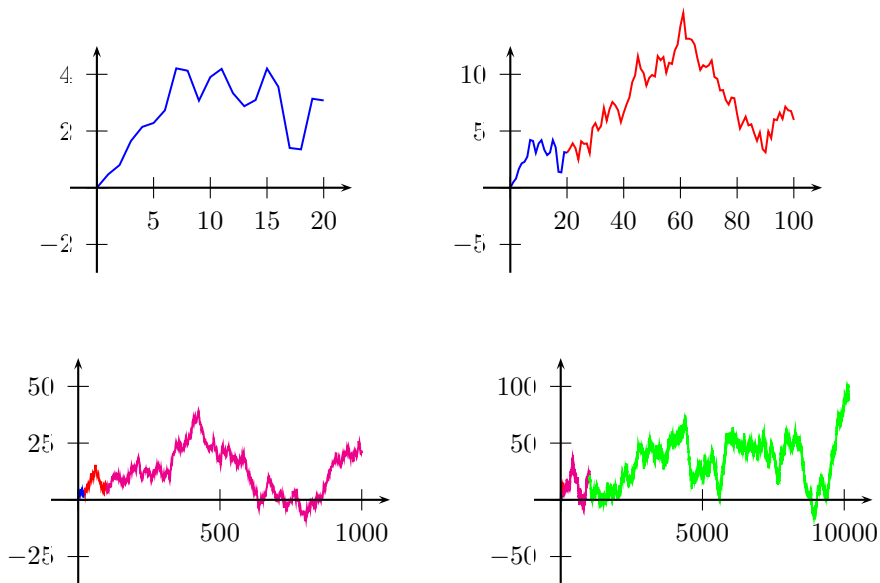
In order to see, how the CLT determines the possible scaling limits of Random Walks, let us consider a one-dimensional Random Walk

$$S_n = \sum_{i=1}^n \eta_i, \quad n = 0, 1, 2, \dots,$$

on a probability space (Ω, \mathcal{A}, P) with independent increments $\eta_i \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ normalized such that

$$E[\eta_i] = 0 \quad \text{and} \quad \text{Var}[\eta_i] = 1. \quad (4.1.3)$$

Plotting many steps of the Random Walk seems to indicate that there is a limit process with continuous sample paths after appropriate rescaling:



To see what appropriate means, we fix a positive integer m , and try to define a rescaled Random Walk $S_t^{(m)}$ ($t = 0, 1/m, 2/m, \dots$) with time steps of size $1/m$ by

$$S_{k/m}^{(m)} = c_m \cdot S_k \quad (k = 0, 1, 2, \dots)$$

for some constants $c_m > 0$. If t is a multiple of $1/m$, then

$$\text{Var}[S_t^{(m)}] = c_m^2 \cdot \text{Var}[S_{mt}] = c_m^2 \cdot m \cdot t.$$

Hence in order to achieve convergence of $S_t^{(m)}$ as $m \rightarrow \infty$, we should choose c_m proportional to $m^{-1/2}$. This leads us to define a continuous time process $(S_t^{(m)})_{t \geq 0}$ by

$$S_t^{(m)}(\omega) := \frac{1}{\sqrt{m}} S_{mt}(\omega) \quad \text{whenever } t = k/m \text{ for some integer } k,$$

and by linear interpolation for $t \in (\frac{k-1}{m}, \frac{k}{m}]$.

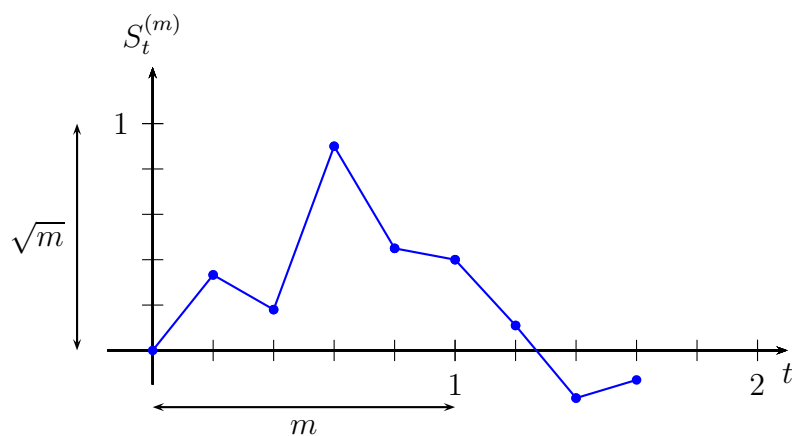


Abbildung 4.1: Rescaling of a Random Walk.

Clearly,

$$E[S_t^{(m)}] = 0 \quad \text{for any } t \geq 0,$$

and

$$\text{Var}[S_t^{(m)}] = \frac{1}{m} \text{Var}[S_{mt}] = t$$

whenever t is a multiple of $1/m$. In particular, the expectation values and variances for a fixed time t do not depend on m . Moreover, if we fix a partition $0 \leq t_0 < t_1 < \dots < t_n$ such that each t_i is a multiple of $1/m$, then the increments

$$S_{t_{i+1}}^{(m)} - S_{t_i}^{(m)} = \frac{1}{\sqrt{m}} (S_{mt_{i+1}} - S_{mt_i}), \quad i = 0, 1, 2, \dots, n-1, \quad (4.1.4)$$

of the rescaled process $(S_t^{(m)})_{t \geq 0}$ are independent centered random variables with variances $t_{i+1} - t_i$. If t_i is not a multiple of $1/m$, then a corresponding statement holds approximately with an error that should be negligible in the limit $m \rightarrow \infty$. Hence, if the rescaled Random Walks $(S_t^{(m)})_{t \geq 0}$ converge in distribution to a limit process $(B_t)_{t \geq 0}$, then $(B_t)_{t \geq 0}$ should have *independent increments* $B_{t_{i+1}} - B_{t_i}$ over disjoint time intervals with mean 0 and variances $t_{i+1} - t_i$.

It remains to determine the precise distributions of the increments. Here the Central Limit Theorem applies. In fact, we can observe that by (4.1.4) each increment

$$S_{t_{i+1}}^{(m)} - S_{t_i}^{(m)} = \frac{1}{\sqrt{m}} \sum_{k=mt_i+1}^{mt_{i+1}} \eta_k$$

of the rescaled process is a rescaled sum of $m \cdot (t_{i+1} - t_i)$ i.i.d. random variables with mean 0 and variance 1. Therefore, the CLT implies that the distributions of the increments converge weakly to a normal distribution:

$$S_{t_{i+1}}^{(m)} - S_{t_i}^{(m)} \xrightarrow{\mathcal{D}} N(0, t_{i+1} - t_i).$$

Hence if a limit process (B_t) exists, then it should have *independent, normally distributed increments*.

Our considerations motivate the following definition:

Definition (Brownian Motion).

(1). Let $a \in \mathbb{R}$. A continuous-time stochastic process $B_t : \Omega \rightarrow \mathbb{R}$, $t \geq 0$, defined on a probability space (Ω, \mathcal{A}, P) , is called a **Brownian motion (starting in a)** if and only if

(a) $B_0(\omega) = a$ for each $\omega \in \Omega$.

(b) For any partition $0 \leq t_0 < t_1 < \dots < t_n$, the increments $B_{t_{i+1}} - B_{t_i}$ are independent random variables with distribution

$$B_{t_{i+1}} - B_{t_i} \sim N(0, t_{i+1} - t_i).$$

(c) P -almost every sample path $t \mapsto B_t(\omega)$ is continuous.

(2). An \mathbb{R}^d -valued stochastic process $B_t(\omega) = (B_t^{(1)}(\omega), \dots, B_t^{(d)}(\omega))$ is called a *multi-dimensional Brownian motion* if and only if the component processes $(B_t^{(1)}), \dots, (B_t^{(d)})$ are independent one-dimensional Brownian motions.

Thus the increments of a d -dimensional Brownian motion are independent over disjoint time intervals and have a multivariate normal distribution:

$$B_t - B_s \sim N(0, (t - s) \cdot I_d) \quad \text{for any } 0 \leq s \leq t.$$

Remark. (1). *Continuity:* Continuity of the sample paths has to be assumed separately: If $(B_t)_{t \geq 0}$ is a one-dimensional Brownian motion, then the modified process $(\tilde{B}_t)_{t \geq 0}$ defined by $\tilde{B}_0 = B_0$ and

$$\tilde{B}_t = B_t \cdot I_{\{B_t \in \mathbb{R} \setminus \mathbb{Q}\}} \quad \text{for } t > 0$$

has almost surely discontinuous paths. On the other hand, it satisfies (a) and (b) since the laws of $(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_n})$ and $(B_{t_1}, \dots, B_{t_n})$ coincide for any $n \in \mathbb{N}$ and $t_1, \dots, t_n \geq 0$.

(2). *Spatial Homogeneity:* If $(B_t)_{t \geq 0}$ is a Brownian motion starting at 0, then the translated process $(a + B_t)_{t \geq 0}$ is a Brownian motion starting at a .

(3). *Existence:* There are several constructions and existence proofs for Brownian motion. In Section 5.1 below we will discuss in detail the Wiener-Lévy construction of Brownian motion as a random superposition of infinitely many deterministic paths. This explicit construction is also very useful for numerical approximations. A more general (but less constructive) existence proof is based on Kolmogorov's extension theorem, cf. e.g. [Klenke].

(4). *Functional Central Limit Theorem:* The construction of Brownian motion as a scaling limit of Random Walks sketched above can also be made rigorous: *Donsker's invariance principle* is a functional version of the Central Limit Theorem which states that the rescaled Random Walks $(S_t^{(m)})$ converge in distribution to a Brownian motion. Similarly as in the classical CLT, the limit is universal, i.e., it does not depend on the distribution of the increments η_i provided (4.1.3) holds.

4.1.2 Brownian motion as a Lévy process.

The definition of Brownian motion shows in particular that Brownian motion is a *Lévy process*, i.e., it has stationary independent increments (over disjoint time intervals). In fact, the analogues of Lévy processes in discrete time are Random Walks, and it is rather obvious, that all scaling limits of Random Walks should be Lévy processes. Brownian motion is the only Lévy process L_t in continuous time with continuous paths such that $E[L_1] = 0$ and $\text{Var}[L_1] = 1$. The normal distribution of the increments follows under these assumptions by an extension of the CLT, cf. e.g. [Breiman: Probability]. A simple example of a Lévy process with non-continuous paths is the Poisson process. Other examples are α -stable processes which arise as scaling limits of Random Walks when the increments are not square-integrable.

Let us now consider a Brownian motion $(B_t)_{t \geq 0}$ starting at a fixed point $a \in \mathbb{R}^d$, defined on a probability space (Ω, \mathcal{A}, P) . The information on the process up to time t is encoded in the σ -algebra

$$\mathcal{F}_t^B = \sigma(B_s : 0 \leq s \leq t)$$

generated by the process. The independence of the increments over disjoint intervals immediately implies:

Lemma 4.1. *For any $0 \leq s \leq t$, the increment $B_t - B_s$ is independent of \mathcal{F}_s^B .*

Beweis. For any partition $0 = t_0 \leq t_1 \leq \dots \leq t_n = s$ of the interval $[0, s]$, the increment $B_t - B_s$ is independent of the σ -algebra

$$\sigma(B_{t_1} - B_{t_0}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}})$$

generated by the increments up to time s . Since

$$B_{t_k} = B_{t_0} + \sum_{i=1}^k (B_{t_i} - B_{t_{i-1}})$$

and B_{t_0} is constant, this σ -algebra coincides with $\sigma(B_{t_0}, B_{t_1}, \dots, B_{t_n})$. Hence $B_t - B_s$ is independent of all finite subcollections of $(B_u : 0 \leq u \leq s)$ and therefore independent of \mathcal{F}_s^B . \square

4.1.3 Brownian motion as a Markov process.

As a process with independent stationary increments, Brownian motion is in particular a time-homogeneous Markov process. In fact, we have:

Satz 4.2 (Markov property). *A Brownian motion $(B_t)_{t \geq 0}$ in \mathbb{R}^d is a time-homogeneous Markov process with transition densities*

$$p_t(x, y) = (2\pi t)^{-d/2} \cdot \exp\left(-\frac{|x - y|^2}{2t}\right), \quad t > 0, \quad x, y \in \mathbb{R}^d,$$

i.e., for any Borel set $A \subseteq \mathbb{R}^d$ and $0 \leq s < t$,

$$P[B_t \in A \mid \mathcal{F}_s^B] = \int_A p_{t-s}(B_s, y) dy \quad P\text{-almost surely.}$$

Beweis. For $0 \leq s < t$ we have $B_t = B_s + (B_t - B_s)$ where B_s is \mathcal{F}_s^B -measurable, and $B_t - B_s$ is independent of \mathcal{F}_s^B by Lemma 4.1. Hence

$$\begin{aligned} P[B_t \in A \mid \mathcal{F}_s^B](\omega) &= P[B_s(\omega) + B_t - B_s \in A] = N(B_s(\omega), (t - s) \cdot I_d)[A] \\ &= \int_A (2\pi(t - s))^{-d/2} \cdot \exp\left(-\frac{|y - B_s(\omega)|^2}{2(t - s)}\right) dy \quad P\text{-almost surely.} \end{aligned}$$

□

We will consider Markov processes in continuous time more systematically in Section 4.3. The Markov property enables us to compute the marginal distributions of Brownian motion:

Korollar 4.3 (Finite dimensional marginals). *Suppose that $(B_t)_{t \geq 0}$ is a Brownian motion starting at $x_0 \in \mathbb{R}^d$ defined on a probability space (Ω, \mathcal{A}, P) . Then for any $n \in \mathbb{N}$ and $0 = t_0 < t_1 < t_2 < \dots < t_n$, the joint distribution of $B_{t_1}, B_{t_2}, \dots, B_{t_n}$ is absolutely continuous with density*

$$\begin{aligned} f_{B_{t_1}, \dots, B_{t_n}}(x_1, \dots, x_n) &= p_{t_1}(x_0, x_1) p_{t_2 - t_1}(x_1, x_2) p_{t_3 - t_2}(x_2, x_3) \cdots p_{t_n - t_{n-1}}(x_{n-1}, x_n) \\ &= \prod_{i=1}^n (2\pi(t_i - t_{i-1}))^{-d/2} \cdot \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right). \end{aligned} \quad (4.1.5)$$

Beweis. By the Markov property and induction on n , we obtain

$$\begin{aligned} &P[B_{t_1} \in A_1, \dots, B_{t_n} \in A_n] \\ &= E[P[B_{t_n} \in A_n \mid \mathcal{F}_{t_{n-1}}^B]; B_{t_1} \in A_1, \dots, B_{t_{n-1}} \in A_{n-1}] \\ &= E[p_{t_n - t_{n-1}}(B_{t_{n-1}}, A_n); B_{t_1} \in A_1, \dots, B_{t_{n-1}} \in A_{n-1}] \\ &= \int_{A_1} \cdots \int_{A_{n-1}} p_{t_1}(x_0, x_1) p_{t_2 - t_1}(x_1, x_2) \cdots \\ &\quad \cdots p_{t_{n-1} - t_{n-2}}(x_{n-2}, x_{n-1}) p_{t_n - t_{n-1}}(x_{n-1}, A_n) dx_{n-1} \cdots dx_1 \\ &= \int_{A_1} \cdots \int_{A_n} \left(\prod_{i=1}^n p_{t_i - t_{i-1}}(x_{i-1}, x_i) \right) dx_n \cdots dx_1 \end{aligned}$$

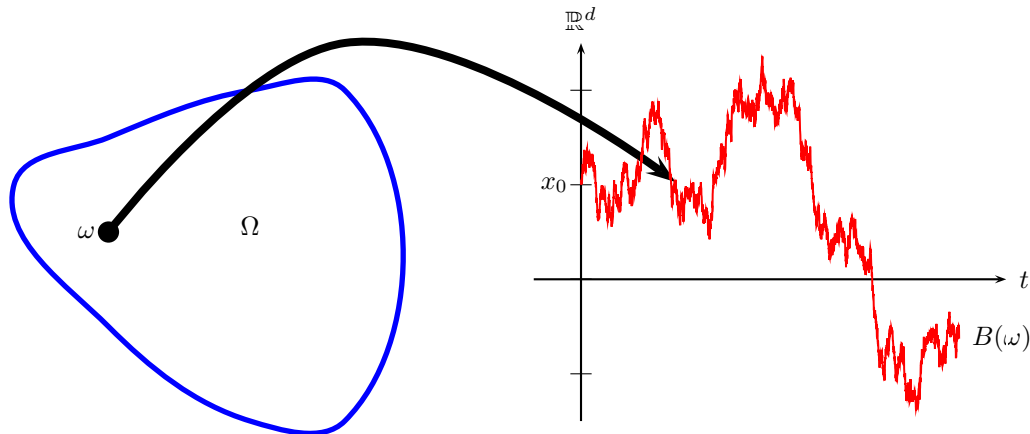
for any $n \geq 0$ and $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^d)$. □

Remark (Brownian motion as a Gaussian process). The corollary shows in particular that Brownian motion is a Gaussian process, i.e., all the marginal distributions in (4.1.5) are multivariate normal distributions. We will come back to this important aspect in the next section.

4.1.4 Wiener Measure

The distribution of Brownian motion could be considered as a probability measure on the product space $(\mathbb{R}^d)^{[0, \infty)}$ consisting of all maps $x : [0, \infty) \rightarrow \mathbb{R}^d$. A disadvantage of this approach is that the product space is far too large for our purposes: It contains extremely irregular paths $x(t)$, although at least almost every path of Brownian motion is continuous by definition. Actually, since $[0, \infty)$ is uncountable, the subset of all continuous paths is not even measurable w.r.t. the product σ -algebra on $(\mathbb{R}^d)^{[0, \infty)}$.

Instead of the product space, we will directly consider the distribution of Brownian motion on the continuous path space $C([0, \infty), \mathbb{R}^d)$. For this purpose, we fix a Brownian motion $(B_t)_{t \geq 0}$ starting at $x_0 \in \mathbb{R}^d$ on a probability space (Ω, \mathcal{A}, P) , and we *assume that every sample path $t \mapsto B_t(\omega)$ is continuous*. This assumption can always be fulfilled by modifying a given Brownian motion on a set of measure zero. The full process $(B_t)_{t \geq 0}$ can then be interpreted as a single path-space valued random variable (or a “random path”).



We endow the space of continuous paths $x : [0, \infty) \rightarrow \mathbb{R}^d$ with the σ -algebra

$$\mathcal{B} = \sigma(X_t : t \geq 0)$$

generated by the coordinate maps

$$X_t : C([0, \infty), \mathbb{R}^d) \rightarrow \mathbb{R}^d, \quad X_t(x) = x_t, \quad t \geq 0.$$

Note that we also have

$$\mathcal{B} = \sigma(X_t : t \in \mathcal{D})$$

for any dense subset \mathcal{D} of $[0, \infty)$, because $X_t = \lim_{s \rightarrow t} X_s$ for each $t \in [0, \infty)$ by continuity. Furthermore, it can be shown that \mathcal{B} is the Borel σ -algebra on $C([0, \infty), \mathbb{R}^d)$ endowed with the topology of uniform convergence on finite intervals.

Satz 4.4 (Distribution of Brownian motion on path space). *The map $B : \Omega \rightarrow C([0, \infty), \mathbb{R}^d)$ is measurable w.r.t. the σ -algebras \mathcal{A}/\mathcal{B} . The distribution $P \circ B^{-1}$ of B is the unique probability measure μ_{x_0} on $(C([0, \infty), \mathbb{R}^d), \mathcal{B})$ with marginals*

$$\begin{aligned} \mu_{x_0} [\{x \in C([0, \infty), \mathbb{R}^d) : x_{t_1} \in A_1, \dots, x_{t_n} \in A_n\}] & \quad (4.1.6) \\ &= \prod_{i=1}^n (2\pi(t_i - t_{i-1}))^{-d/2} \int_{A_1} \cdots \int_{A_n} \exp\left(-\frac{1}{2} \sum_{i=1}^n \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right) dx_n \cdots dx_1 \end{aligned}$$

for any $n \in \mathbb{N}$, $0 < t_1 < \dots < t_n$, and $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^d)$.

Definition (Wiener measure). *The probability measure μ_{x_0} on the path space $C([0, \infty), \mathbb{R}^d)$ determined by (4.1.6) is called **Wiener measure** (with start in x_0).*

Remark (Uniqueness in distribution). The theorem asserts that the path space distribution of a Brownian motion starting at a given point x_0 is the corresponding Wiener measure. In particular, it is uniquely determined by the marginal distributions in (4.1.5).

Proof of Theorem 4.4. For $n \in \mathbb{N}$, $0 < t_1 < \dots < t_n$, and $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\begin{aligned} B^{-1}(\{X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\}) &= \{\omega : X_{t_1}(B(\omega)) \in A_1, \dots, X_{t_n}(B(\omega)) \in A_n\} \\ &= \{B_{t_1} \in A_1, \dots, B_{t_n} \in A_n\} \in \mathcal{A}. \end{aligned}$$

Since the cylinder sets of type $\{X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\}$ generate the σ -algebra \mathcal{B} , the map B is \mathcal{A}/\mathcal{B} -measurable. Moreover, by Corollary 4.3, the probabilities

$$P[B \in \{X_{t_1} \in A_1, \dots, X_{t_n} \in A_n\}] = P[B_{t_1} \in A_1, \dots, B_{t_n} \in A_n]$$

are given by the right hand side of (4.1.6). Finally, the measure μ_{x_0} is uniquely determined by (4.1.6), since the system of cylinder sets as above is stable under intersections and generates the σ -algebra \mathcal{B} . \square

Definition (Canonical model for Brownian motion). *By (4.1.6), the coordinate process*

$$X_t(x) = x_t, \quad t \geq 0,$$

on $C([0, \infty), \mathbb{R}^d)$ is a Brownian motion starting at x_0 w.r.t. Wiener measure μ_{x_0} . We refer to the stochastic process $(X_t)_{t \geq 0}$ on $(C([0, \infty), \mathbb{R}^d), \mathcal{B}, \mu_{x_0})$ as the **canonical model for Brownian motion starting at x_0** .

4.2 Gaussian Processes

We have already verified that Brownian motion is a Gaussian process, i.e., the finite dimensional marginals are multivariate normal distributions. We will now exploit this fact more thoroughly.

4.2.1 Multivariate normals

Let us first recall some basics on normal random vectors:

Definition (Multivariate normal distribution). Suppose that $m \in \mathbb{R}^n$ is a vector and $C \in \mathbb{R}^{n \times n}$ is a symmetric non-negative definite matrix. A random variable $Y : \Omega \rightarrow \mathbb{R}^n$ defined on a probability space (Ω, \mathcal{A}, P) has a **multivariate normal distribution** $N(m, C)$ **with mean m and covariance matrix C** if and only if its characteristic function is given by

$$E[e^{ip \cdot Y}] = e^{ip \cdot m - \frac{1}{2} p \cdot C p} \quad \text{for any } p \in \mathbb{R}^n. \quad (4.2.1)$$

If C is non-degenerate, then a multivariate normal random variable Y is absolutely continuous with density

$$f_Y(x) = (2\pi \det C)^{-1/2} \exp\left(-\frac{1}{2}(x - m) \cdot C^{-1}(x - m)\right).$$

A degenerate normal distribution with vanishing covariance matrix is a Dirac measure:

$$N(m, 0) = \delta_m.$$

Differentiating (4.2.1) w.r.t. p shows that for a random variable $Y \sim N(m, C)$, the mean vector is m and $C_{i,j}$ is the covariance of the components Y_i and Y_j . Moreover, the following important facts hold:

Satz 4.5 (Properties of normal random vectors).

- (1). A random variable $Y : \Omega \rightarrow \mathbb{R}^n$ has a multivariate normal distribution if and only if any linear combination

$$p \cdot Y = \sum_{i=1}^n p_i Y_i, \quad p \in \mathbb{R}^n,$$

of the components Y_i has a one dimensional normal distribution.

(2). Any affine function of a normally distributed random vector Y is again normally distributed:

$$Y \sim N(m, C) \implies AY + b \sim N(Am + b, ACA^\top)$$

for any $d \in \mathbb{N}$, $A \in \mathbb{R}^{d \times n}$ and $b \in \mathbb{R}^d$.

(3). If $Y = (Y_1, \dots, Y_n)$ has a multivariate normal distribution, and the components Y_1, \dots, Y_n are uncorrelated random variables, then Y_1, \dots, Y_n are independent.

Beweis. (1). follows easily from the definition.

(2). For $Y \sim N(m, C)$, $A \in \mathbb{R}^{d \times n}$ and $b \in \mathbb{R}^d$ we have

$$\begin{aligned} E[e^{ip \cdot (AY+b)}] &= e^{ip \cdot b} E[e^{i(A^\top p) \cdot Y}] \\ &= e^{ip \cdot b} e^{i(A^\top p) \cdot m - \frac{1}{2}(A^\top p) \cdot C A^\top p} \\ &= e^{ip \cdot (Am+b) - \frac{1}{2}p \cdot ACA^\top} \quad \text{for any } p \in \mathbb{R}^d, \end{aligned}$$

i.e., $AY + b \sim N(Am + b, ACA^\top)$.

(3). If Y_1, \dots, Y_n are uncorrelated, then the covariance matrix $C_{i,j} = \text{Cov}[Y_i, Y_j]$ is a diagonal matrix. Hence the characteristic function

$$E[e^{ip \cdot Y}] = e^{ip \cdot m - \frac{1}{2}p \cdot C p} = \prod_{k=1}^n e^{im_k p_k - \frac{1}{2}C_{k,k} p_k^2}$$

is a product of characteristic functions of one-dimensional normal distributions. Since a probability measure on \mathbb{R}^n is uniquely determined by its characteristic function, it follows that the adjoint distribution of Y_1, \dots, Y_n is a product measure, i.e. Y_1, \dots, Y_n are independent. □

If Y has a multivariate normal distribution $N(m, C)$ then for any $p, q \in \mathbb{R}^n$, the random variables $p \cdot Y$ and $q \cdot Y$ are normally distributed with means $p \cdot m$ and $q \cdot m$, and covariance

$$\text{Cov}[p \cdot Y, q \cdot Y] = \sum_{i,j=1}^n p_i C_{i,j} q_j = p \cdot C q.$$

In particular, let $\{e_1, \dots, e_n\} \subseteq \mathbb{R}^n$ be an orthonormal basis consisting of eigenvectors of the covariance matrix C . Then the components $e_i \cdot Y$ of Y in this basis are uncorrelated and therefore independent, jointly normally distributed random variables with variances given by the corresponding eigenvalues λ_i :

$$\text{Cov}[e_i \cdot Y, e_j \cdot Y] = \lambda_i \delta_{i,j}, \quad 1 \leq i, j \leq n. \quad (4.2.2)$$

Correspondingly, the contour lines of the density of a non-degenerate multivariate normal distribution $N(m, C)$ are ellipsoids with center at m and principal axes of length $\sqrt{\lambda_i}$ given by the eigenvectors e_i of the covariance matrix C .

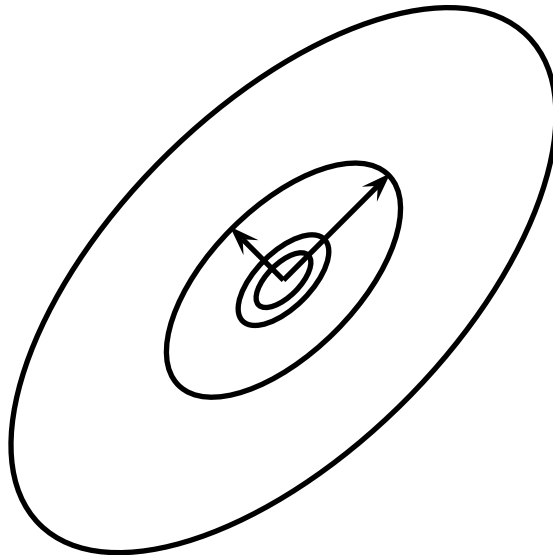


Abbildung 4.2: Level lines of the density of $N\left(\begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}\right)$.

Conversely, we can generate a random vector Y with distribution $N(m, C)$ from independent standard normal random variables Z_1, \dots, Z_n by setting

$$Y = m + \sum_{i=1}^n \sqrt{\lambda_i} Z_i e_i. \quad (4.2.3)$$

More generally, we have:

Korollar 4.6 (Generating normal random vectors). *Suppose that $C = U\Lambda U^\top$ with a matrix $U \in \mathbb{R}^{n \times d}$, $d \in \mathbb{N}$, and a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_d) \in \mathbb{R}^{d \times d}$ with nonnegative entries λ_i . If $Z = (Z_1, \dots, Z_d)$ is a random vector with independent standard normal components Z_1, \dots, Z_d , then*

$$Y = U\Lambda^{1/2}Z + m$$

has distribution $N(m, C)$.

Beweis. Since $Z \sim N(0, I_d)$, the second assertion of Theorem 4.5 implies

$$Y \sim N(m, U\Lambda U^\top).$$

□

Choosing for U the matrix (e_1, \dots, e_n) given by the orthonormal eigenvectors e_1, \dots, e_n of C , we obtain (4.2.3) as a special case of the corollary. For computational purposes it is often more convenient to use the Cholesky decomposition

$$C = LL^\top$$

of the covariance matrix as a product of a lower triangular matrix L and the upper triangular transpose L^\top :

Algorithmus 4.7 (Simulation of multivariate normal random variables).

Given: $m \in \mathbb{R}^n, C \in \mathbb{R}^{n \times n}$ symmetric and non-negative definite.

Output: A sample $y \sim N(m, C)$.

- (1). Compute the Cholesky decomposition $C = LL^\top$.
- (2). Generate independent samples $z_1, \dots, z_n \sim N(0, 1)$ (e.g. by the Box-Muller method).
- (3). Set $y := Lz + m$.

4.2.2 Gaussian processes

Let I be an arbitrary index set, for instance $I = \mathbb{N}, I = [0, \infty)$ or $I = \mathbb{R}^n$.

Definition (Gaussian process). A collection $(Y_t)_{t \in I}$ of random variables $Y_t : \Omega \rightarrow \mathbb{R}^d$ defined on a probability space (Ω, \mathcal{A}, P) is called a **Gaussian process** if and only if the joint distribution of any finite subcollection Y_{t_1}, \dots, Y_{t_n} with $n \in \mathbb{N}$ and $t_1, \dots, t_n \in I$ is a multivariate normal distribution.

The distribution of a Gaussian process $(Y_t)_{t \in I}$ on the path space \mathbb{R}^I or $C(I, \mathbb{R})$ endowed with the σ -algebra generated by the maps $x \mapsto x_t, t \in I$, is uniquely determined by the multinormal distributions of finite subcollections Y_{t_1}, \dots, Y_{t_n} as above, and hence by the expectation values

$$m(t) = E[Y_t], \quad t \in I,$$

and the covariances

$$c(s, t) = \text{Cov}[Y_s, Y_t], \quad s, t \in I.$$

A Gaussian process is called **centered**, if $m(t) = 0$ for any $t \in I$.

Example (AR(1) process). The autoregressive process $(Y_n)_{n=0,1,2,\dots}$ defined recursively by $Y_0 \sim N(0, v_0)$,

$$Y_n = \alpha Y_{n-1} + \varepsilon \eta_n \quad \text{for } n \in \mathbb{N},$$

with parameters $v_0 > 0$, $\alpha, \varepsilon \in \mathbb{R}$, and η_n i.i.d. $\sim N(0, 1)$, is a centered Gaussian process. The covariance function is given by

$$\begin{aligned} c(n, n+k) &= v_0 + \varepsilon^2 n && \text{for any } n, k \geq 0 \quad \text{if } \alpha = 1, \quad \text{and} \\ c(n, n+k) &= \alpha^k \cdot \left(\alpha^{2n} v_0 + (1 - \alpha^{2n}) \cdot \frac{\varepsilon^2}{1 - \alpha^2} \right) && \text{for } n, k \geq 0 \quad \text{otherwise.} \end{aligned}$$

This is easily verified by induction. We now consider some special cases:

$\alpha = 0$: In this case $Y_n = \varepsilon \eta_n$. Hence (Y_n) is a *white noise*, i.e., a sequence of independent normal random variables, and

$$\text{Cov}[Y_n, Y_m] = \varepsilon^2 \cdot \delta_{n,m} \quad \text{for any } n, m \geq 1.$$

$\alpha = 1$: Here $Y_n = Y_0 + \varepsilon \sum_{i=1}^n \eta_i$, i.e., the process (Y_n) is a *Gaussian Random Walk*, and

$$\text{Cov}[Y_n, Y_m] = v_0 + \varepsilon^2 \cdot \min(n, m) \quad \text{for any } n, m \geq 0.$$

We will see a corresponding expression for the covariances of Brownian motion.

$\alpha < 1$: For $\alpha < 1$, the covariances $\text{Cov}[Y_n, Y_{n+k}]$ decay exponentially fast as $k \rightarrow \infty$. If $v_0 = \frac{\varepsilon^2}{1-\alpha^2}$, then the covariance function is translation invariant:

$$c(n, n+k) = \frac{\varepsilon^2 \alpha^k}{1 - \alpha^2} \quad \text{for any } n, k \geq 0.$$

Therefore, in this case the process (Y_n) is *stationary*, i.e., $(Y_{n+k})_{n \geq 0} \sim (Y_n)_{n \geq 0}$ for all $k \geq 0$.

4.2.3 Brownian Motion as a Gaussian Process

Brownian motion is our first example of a nontrivial Gaussian process in continuous time. In fact, we have:

Satz 4.8 (Gaussian characterization of Brownian motion). *A real-valued stochastic process $(B_t)_{t \in [0, \infty)}$ with continuous sample paths $t \mapsto B_t(\omega)$ and $B_0 = 0$ is a Brownian motion if and only if (B_t) is a centered Gaussian process with covariances*

$$\text{Cov}[B_s, B_t] = \min(s, t) \quad \text{for any } s, t \geq 0. \quad (4.2.4)$$

Beweis. For a Brownian motion (B_t) and $0 = t_0 < t_1 < \dots < t_n$, the increments $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq n$, are independent random variables with distribution $N(0, t_i - t_{i-1})$. Hence,

$$(B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}}) \sim \bigotimes_{i=1}^n N(0, t_i - t_{i-1}),$$

which is a multinormal distribution. Since $B_{t_0} = B_0 = 0$, we see that

$$\begin{pmatrix} B_{t_1} \\ \vdots \\ B_{t_n} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & \dots & 0 & 0 \\ & & \ddots & & & \\ & & & \ddots & & \\ 1 & 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & 1 & \dots & 1 & 1 \end{pmatrix} \begin{pmatrix} B_{t_1} - B_{t_0} \\ \vdots \\ B_{t_n} - B_{t_{n-1}} \end{pmatrix}$$

also has a multivariate normal distribution, i.e., (B_t) is a Gaussian process. Moreover, since $B_t = B_t - B_0$, we have $E[B_t] = 0$ and

$$\text{Cov}[B_s, B_t] = \text{Cov}[B_s, B_s] + \text{Cov}[B_s, B_t - B_s] = \text{Var}[B_s] = s$$

for any $0 \leq s \leq t$, i.e., (4.2.4) holds.

Conversely, if (B_t) is a centered Gaussian process satisfying (4.2.4), then for any $0 = t_0 < t_1 < \dots < t_n$, the vector $(B_{t_1} - B_{t_0}, \dots, B_{t_n} - B_{t_{n-1}})$ has a multivariate normal distribution with

$$E[B_{t_i} - B_{t_{i-1}}] = E[B_{t_i}] - E[B_{t_{i-1}}] = 0, \quad \text{and}$$

$$\begin{aligned} \text{Cov}[B_{t_i} - B_{t_{i-1}}, B_{t_j} - B_{t_{j-1}}] &= \min(t_i, t_j) - \min(t_i, t_{j-1}) \\ &\quad - \min(t_{i-1}, t_j) + \min(t_{i-1}, t_{j-1}) \\ &= (t_i - t_{i-1}) \cdot \delta_{i,j} \quad \text{for any } i, j = 1, \dots, n. \end{aligned}$$

Hence by Theorem 4.5 (3), the increments $B_{t_i} - B_{t_{i-1}}$, $1 \leq i \leq n$, are independent with distribution $N(0, t_i - t_{i-1})$, i.e., (B_t) is a Brownian motion. \square

Symmetries of Brownian motion

A first important consequence of the Gaussian characterization of Brownian motion are several symmetry properties of Wiener measure:

Satz 4.9 (Invariance properties of Wiener measure). *Let $(B_t)_{t \geq 0}$ be a Brownian motion starting at 0 defined on a probability space (Ω, \mathcal{A}, P) . Then the following processes are again Brownian motions:*

- (1). $(-B_t)_{t \geq 0}$ (Reflection invariance)
- (2). $(B_{t+h} - B_h)_{t \geq 0}$ for any $h \geq 0$ (Stationarity)
- (3). $(a^{-1/2} B_{at})_{t \geq 0}$ for any $a > 0$ (Scale invariance)
- (4). The time inversion $(\tilde{B}_t)_{t \geq 0}$ defined by

$$\tilde{B}_0 = 0, \quad \tilde{B}_t = t \cdot B_{1/t} \quad \text{for } t > 0.$$

Beweis. The proofs of (1), (2) and (3) are left as an exercise to the reader. To show (4), we first note that for each $n \in \mathbb{N}$ and $0 \leq t_1 < \dots < t_n$, the vector $(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_n})$ has a multivariate normal distribution since it is a linear transformation of $(B_{1/t_1}, \dots, B_{1/t_n}), (B_0, B_{1/t_2}, \dots, B_{1/t_n})$ respectively. Moreover,

$$\begin{aligned} E[\tilde{B}_t] &= 0 \quad \text{for any } t \geq 0, \\ \text{Cov}[\tilde{B}_s, \tilde{B}_t] &= st \cdot \text{Cov}[B_{1/s}, B_{1/t}] \\ &= st \cdot \min\left(\frac{1}{s}, \frac{1}{t}\right) = \min(t, s) \quad \text{for any } s, t > 0, \quad \text{and} \\ \text{Cov}[\tilde{B}_0, \tilde{B}_t] &= 0 \quad \text{for any } t \geq 0. \end{aligned}$$

Hence $(\tilde{B}_t)_{t \geq 0}$ is a centered Gaussian process with the covariance function of Brownian motion. By Theorem 4.8, it only remains to show that P -almost every sample path $t \mapsto \tilde{B}_t(\omega)$ is continuous. This is obviously true for $t > 0$. Furthermore, since the finite dimensional marginals of the processes $(\tilde{B}_t)_{t \geq 0}$ and $(B_t)_{t \geq 0}$ are multivariate normal distributions with the same means and covariances, the distributions of $(\tilde{B}_t)_{t \geq 0}$ and $(B_t)_{t \geq 0}$ on the product space $\mathbb{R}^{(0, \infty)}$ endowed with the product σ -algebra generated by the cylinder sets agree. To prove continuity at 0 we note that the set

$$\left\{ x : (0, \infty) \rightarrow \mathbb{R} : \lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} x_t = 0 \right\}$$

is measurable w.r.t. the product σ -algebra on $\mathbb{R}^{(0, \infty)}$ (in contrast to the set of all paths that are continuous at 0!). Therefore,

$$P \left[\lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} \tilde{B}_t = 0 \right] = P \left[\lim_{\substack{t \searrow 0 \\ t \in \mathbb{Q}}} B_t = 0 \right] = 1.$$

Since \tilde{B}_t is almost surely continuous for $t > 0$, we can conclude that outside a set of measure zero,

$$\sup_{s \in (0,t)} |\tilde{B}_s| = \sup_{s \in (0,t) \cap \mathbb{Q}} |\tilde{B}_s| \longrightarrow 0 \quad \text{as } t \searrow 0,$$

i.e., $t \mapsto \tilde{B}_t$ is almost surely continuous at 0 as well. \square

Remark (Long time asymptotics versus local regularity, Law of large numbers). The time inversion invariance of Wiener measure enables us to translate results on the long time asymptotics of Brownian motion ($t \nearrow \infty$) into local regularity results for Brownian paths ($t \searrow 0$) and vice versa. For example, the continuity of the process (\tilde{B}_t) at 0 is equivalent to the *law of large numbers*:

$$P \left[\lim_{t \rightarrow \infty} \frac{1}{t} B_t = 0 \right] = P \left[\lim_{s \searrow 0} s B_{1/s} = 0 \right] = 1.$$

At first glance, this looks like a simple proof of the LLN. However, the argument is based on the existence of a continuous Brownian motion. The law of large numbers is hidden in the existence proof of Brownian motion.

Wiener measure as a Gaussian measure, path integral heuristics

Wiener measure (with start at 0) is the unique probability measure μ on the continuous path space $C([0, \infty), \mathbb{R}^d)$ such that the coordinate process

$$X_t : C([0, \infty), \mathbb{R}^d) \rightarrow \mathbb{R}^d, \quad X_t(x) = x(t),$$

is a Brownian motion starting at 0. By Theorem 4.8, Wiener measure is a centered **Gaussian measure** on the infinite dimensional vector space $C([0, \infty), \mathbb{R}^d)$, i.e., for any $n \in \mathbb{N}$ and $t_1, \dots, t_n \in \mathbb{R}_+$, $(X_{t_1}, \dots, X_{t_n})$ is normally distributed with mean 0. We now “derive” a heuristic representation of Wiener measure that is not mathematically rigorous but nevertheless useful:

Fix a constant $T > 0$. Then for $0 = t_0 < t_1 < \dots < t_n \leq T$, the distribution of $(X_{t_1}, \dots, X_{t_n})$ w.r.t. Wiener measure is

$$\mu_{t_1, \dots, t_n}(dx_{t_1} \cdots dx_{t_n}) = \frac{1}{Z(t_1, \dots, t_n)} \exp \left(-\frac{1}{2} \sum_{i=1}^n \frac{|x_{t_i} - x_{t_{i-1}}|^2}{t_i - t_{i-1}} \right) \prod_{i=1}^n dx_{t_i}, \quad (4.2.5)$$

where $Z(t_1, \dots, t_n)$ is an appropriate finite normalization constant, and $x_0 := 0$. Now choose a sequence $(\tau_k)_{k \in \mathbb{N}}$ of partitions $0 = t_0^{(k)} < t_1^{(k)} < \dots < t_{n(k)}^{(k)} = T$ of the interval $[0, T]$ such that the mesh size $\max_i |t_{i+1}^{(k)} - t_i^{(k)}|$ tends to zero. Taking informally the limit in (4.2.5), we obtain the *heuristic* asymptotic representation

$$\mu(dx) = \frac{1}{Z_\infty} \exp \left(-\frac{1}{2} \int_0^T \left| \frac{dx}{dt} \right|^2 dt \right) \delta_0(dx_0) \prod_{t \in (0, T]} dx_t \quad (4.2.6)$$

for Wiener measure on continuous paths $x : [0, T] \rightarrow \mathbb{R}^d$ with a “normalizing constant” Z_∞ . Trying to make the informal expression (4.2.6) rigorous fails for several reasons:

- The normalizing constant $Z_\infty = \lim_{k \rightarrow \infty} Z(t_1^{(k)}, \dots, t_{n(k)}^{(k)})$ is infinite.
- The integral $\int_0^T \left| \frac{dx}{dt} \right|^2 dt$ is also infinite for μ -almost every path x , since typical paths of Brownian motion are nowhere differentiable, cf. Section 5.2 below.
- The product measure $\prod_{t \in (0, T]} dx_t$ can be defined on cylinder sets, but an extension to the σ -algebra generated by the coordinate maps on $C([0, \infty), \mathbb{R}^d)$ does not exist.

Hence there are several infinities involved in the informal expression (4.2.6). These infinities magically balance each other such that the measure μ is well defined in contrast to all of the factors on the right hand side. In quantum physics, R. Feynman introduced corresponding integrals w.r.t. “Lebesgue measure on path space”, cf. e.g. the famous Feynman Lecture notes [Feynman], or [Glimm and Jaffe].

Although not mathematically rigorous, the heuristic expression (4.2.5) can be a very helpful guide for intuition. Note for example that (4.2.6) takes the form

$$\mu(dx) \propto \exp(-\|x\|_H^2/2) \lambda(dx), \quad (4.2.7)$$

where $\|x\|_H = (x, x)_H^{1/2}$ is the norm induced by the inner product

$$(x, y)_H = \int_0^T \frac{dx}{dt} \frac{dy}{dt} dt \quad (4.2.8)$$

of functions $x, y : [0, T] \rightarrow \mathbb{R}^d$ vanishing at 0, and λ is a corresponding “infinite dimensional Lebesgue measure” (which does not exist!). The vector space

$$H = \left\{ x : [0, T] \rightarrow \mathbb{R}^d : x(0) = 0, x \text{ is absolutely continuous with } \frac{dx}{dt} \in L^2 \right\}$$

is a Hilbert space w.r.t. the inner product (4.2.8). Therefore, (4.2.7) suggests to consider Wiener measure as a *standard normal distribution on H* . It turns out that this idea can be made rigorous although not as easily as one might think at first glance. The difficulty is that a standard normal distribution on an infinite-dimensional Hilbert space does not exist on the space itself but only on a larger space. We will see in Chapter 5 that Wiener measure μ can indeed be realized on the continuous path space $C([0, T], \mathbb{R}^d)$, but μ -almost every path is not contained in the Hilbert space H !

Remark (Infinite-dimensional standard normal distributions). The fact that a standard normal distribution on an infinite dimensional separable Hilbert space H can not be realized on the space H itself can be easily seen by contradiction: Suppose that μ is a standard normal distribution on H , and $e_n, n \in \mathbb{N}$, are infinitely many orthonormal vectors in H . Then by rotational symmetry, the balls

$$B_n = \left\{ x \in H : \|x - e_n\|_H < \frac{1}{2} \right\}, \quad n \in \mathbb{N},$$

should all have the same measure. On the other hand, the balls are disjoint. Hence by σ -additivity,

$$\sum_{n=1}^{\infty} \mu[B_n] = \mu \left[\bigcup B_n \right] \leq \mu[H] = 1,$$

and therefore $\mu[B_n] = 0$ for all $n \in \mathbb{N}$. A scaling argument now implies that

$$\mu[\{x \in H : \|x - h\| \leq \|h\|/2\}] = 0 \quad \text{for any } h \in H,$$

and hence $\mu \equiv 0$.

4.3 Markovprozesse in stetiger Zeit

Für viele Anwendungsprobleme sind Modelle, die auf Markovprozessen in kontinuierlicher Zeit basieren, natürlicher. Ändert der Prozess nur an abzählbar vielen zufälligen Zeitpunkten seinen Zustand, dann nennt man ihn eine zeitstetige Markovkette. Klassische Anwendungsbereiche zeitstetiger Markovketten sind die Modellierung von Warteschlangen und chemischen Reaktionen. Wir zeigen hier, wie man zeitstetige aus zeitdiskreten Markovketten konstruiert und beschreibt. Viele der Aussagen aus den letzten Abschnitten haben Entsprechungen im zeitstetigen Fall – wir verweisen dazu auf das einführende Lehrbuch [J. Norris: Markov Chains].

4.3.1 Übergangskerne und Markovprozesse

Die Verteilungen von zeitstetigen Markovprozessen werden durch die Übergangswahrscheinlichkeiten $p_{s,t}$ vom Zeitpunkt s zum Zeitpunkt t festgelegt.

Definition (Markovprozess). Seien $p_{s,t}$ ($0 \leq s \leq t < \infty$) stochastische Kerne auf einem messbaren Raum (S, \mathcal{S}) .

- (1). Ein auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) definierter zeitstetiger stochastischer Prozess $X_t : \Omega \rightarrow S$ ($t \in [0, \infty)$) heißt **Markovprozess mit Übergangswahrscheinlichkeiten** $p_{s,t}(x, dy)$, falls

$$P[X_{t+h} \in B \mid \mathcal{F}_t^X] = p_{t,t+h}(X_t, B) \quad P\text{-fast sicher}$$

für alle $B \in \mathcal{S}$ und $t, h \geq 0$ gilt, wobei $\mathcal{F}_t^X = \sigma(X_s : 0 \leq s \leq t)$ die vom Prozess erzeugten σ -Algebren sind.

- (2). Der Markovprozess heißt **zeitlich homogen**, falls die Übergangswahrscheinlichkeiten $p_{t,t+h}$ nur von h abhängen:

$$p_{t,t+h}(x, dy) = p_h(x, dy) \quad \text{für alle } t, h \geq 0.$$

Wenn stochastische Kerne als Übergangswahrscheinlichkeiten eines Markovprozesses auftreten, müssen sie konsistent sein, das heißt es muss

$$p_{s,u} = p_{s,t}p_{t,u} \quad \text{für alle } 0 \leq s \leq t \leq u, \quad (4.3.1)$$

bzw., im zeithomogenen Fall,

$$p_{s+t} = p_s p_t = p_t p_s \quad \text{für alle } s, t \geq 0 \quad (4.3.2)$$

gelten. (4.3.1) und (4.3.2) werden auch als **Chapman-Kolmogorov-Gleichungen** bezeichnet. Im zeithomogenen Fall besagt die Chapman-Kolmogorov-Gleichung (4.3.2), dass die Übergangskerne $p_t, t \geq 0$, eine **Halbgruppe** bilden.

Bemerkung (Diffusions- und Sprungprozesse). Einen Markovprozess mit stetigen Pfaden $t \mapsto X_t(\omega)$ nennt man einen **Diffusionsprozess**. Ein fundamentales Beispiel ist die Brownsche Bewegung. Eine **zeitstetige Markovkette** ist ein Markovprozess, dessen Pfade stückweise konstant sind, und nur an abzählbar vielen (von ω abhängigen) Zeitpunkten springen. Allgemeine Markovprozesse können sowohl stetige als auch Sprunganteile haben – es ist auch möglich, dass sich die Sprünge häufen.

Die zeithomogenen reellwertigen Markovprozesse mit räumlich homogenen Übergangswahrscheinlichkeiten sind gerade die Lévy-Prozesse:

Satz 4.10 (Lévy-Prozesse als Markovprozesse). Ein \mathbb{R}^d -wertiger stochastischer Prozess (X_t, P) mit stationären unabhängigen Inkrementen $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ ($0 \leq t_0 < t_1 < \dots < t_n$) ist ein zeithomogener Markovprozess mit translationsinvarianten Übergangswahrscheinlichkeiten

$$p_h(x, B) = P[X_{t+h} - X_t \in B - x] \quad \text{für alle } t, h \geq 0 \text{ und } B \in \mathcal{B}(\mathbb{R}^d).$$

Beweis. Für jede Partition $0 = t_0 < t_1 < \dots < t_n = t$ eines Intervalls $[0, t]$ sind die Inkremente $X_{t+h} - X_t$ für $h \geq 0$ unabhängig von $\sigma(X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}})$. Wegen

$$X_{t_k} = X_{t_0} + \sum_{i=1}^k (X_{t_i} - X_{t_{i-1}})$$

erzeugen die Zufallsvariablen $X_{t_0}, X_{t_1}, \dots, X_{t_n}$ dieselbe σ -Algebra. Also ist $X_{t+h} - X_t$ auch unabhängig von der σ -Algebra

$$\mathcal{F}_t^X = \sigma(X_s : 0 \leq s \leq t) = \sigma\left(\bigcup_{0=t_0 < t_1 < \dots < t_n=t} \sigma(X_{t_0}, X_{t_1}, \dots, X_{t_n})\right).$$

Damit folgt für P -fast alle ω :

$$\begin{aligned} P[X_{t+h} \in B \mid \mathcal{F}_t^X](\omega) &= P[X_t + (X_{t+h} - X_t) \in B \mid \mathcal{F}_t^X](\omega) \\ &= P[X_{t+h} - X_t \in B - X_t(\omega)] = p_h(X_t(\omega), B). \end{aligned}$$

□

Beispiel. (1). *Poissonprozess:* Ein Poissonprozess mit Parameter $\lambda > 0$ ist eine zeitstetige Markovkette mit Zustandsraum $S = \{0, 1, 2, \dots\}$ und Übergangswahrscheinlichkeiten

$$p_t(x, y) = e^{-\lambda t} \frac{(\lambda t)^{y-x}}{(y-x)!} \quad \text{für } y \geq x, \quad p_t(x, y) = 0 \quad \text{für } y < x.$$

(2). *Brownsche Bewegung:* Eine d -dimensionale Brownsche Bewegung ist ein zeitlich homogener Diffusionsprozess mit Zustandsraum $S = \mathbb{R}^d$ und absolutstetigen Übergangswahrscheinlichkeiten $p_t(x, dy)$ mit Dichte

$$p_t(x, y) = (2\pi t)^{-d/2} \cdot \exp\left(-\frac{\|x-y\|^2}{2t}\right), \quad t > 0, \quad x, y \in \mathbb{R}^d.$$

Ist $(X_t)_{t \in [0, \infty)}$ bzgl. P ein zeitstetiger Markovprozess und (t_n) eine aufsteigende Folge in \mathbb{R}_+ , dann ist (X_{t_n}) eine zeitdiskrete Markovkette mit Übergangskernen p_{t_n-1, t_n} . Insbesondere erhalten wir mit Satz 2.3:

Korollar 4.11 (Endlichdimensionale Randverteilungen). Für jedes $n \geq 0$ und $0 = t_0 < t_1 < \dots < t_n$ hat $(X_{t_0}, X_{t_1}, \dots, X_{t_n})$ die Verteilung

$$\mu(dx_0) p_{t_0, t_1}(x_0, dx_1) p_{t_1, t_2}(x_1, dx_2) \cdots p_{t_{n-1}, t_n}(x_{n-1}, dx_n),$$

wobei $\mu = P \circ X_0^{-1}$ die Startverteilung des Markovprozesses ist.

Bemerkung (Eindeutigkeit in Verteilung, Modifikationen). Nach dem Korollar ist die Verteilung eines Markovprozesses $((X_t)_{t \geq 0}, P)$ auf dem Produktraum $S^{[0, \infty)}$ mit Produkt- σ -Algebra eindeutig durch die Startverteilung und die Übergangswahrscheinlichkeiten festgelegt. Da es überabzählbar viele Zeitpunkte $t \in \mathbb{R}_+$ gibt, ist die Situation allerdings etwas subtiler als im zeitdiskreten Fall. Beispielsweise ist das Ereignis, dass die Pfade $t \mapsto X_t(\omega)$ des Prozesses stetig bzw. rechtsstetig sind, *nicht messbar* bzgl. der Produkt- σ -Algebra! Tatsächlich kann man zu einem Markovprozess (X_t) mit (rechts-)stetigen Pfaden in der Regel einen modifizierten Prozess (\tilde{X}_t) mit

$$P[\tilde{X}_t = X_t] = 1 \quad \text{für jedes } t \in \mathbb{R}_+$$

finden, der keine (rechts-)stetigen Pfade hat. Der Prozess (\tilde{X}_t) hat dann dieselben endlichdimensionalen Randverteilungen wie (X_t) , und ist daher ein Markovprozess mit derselben Startverteilung und denselben Übergangswahrscheinlichkeiten!

4.3.2 Zeitstetige Markovketten

Wir wollen nun (umgekehrt wie oben) aus einer zeitdiskreten Markovkette einen zeitstetigen Markovprozess konstruieren, der dieselben Zustände durchläuft, aber zu zufälligen kontinuierlichen Zeitpunkten von einem Zustand zum nächsten springt. Dazu betrachten wir der Übersichtlichkeit halber nur den Fall eines abzählbaren Zustandsraumes S . Einen zeitstetigen Markovprozess auf S charakterisieren wir dann durch die infinitesimalen Übergangsraten

$$\mathcal{L}_t(x, y) = \lim_{h \searrow 0} \frac{p_{t, t+h}(x, y) - \delta(x, y)}{h}, \quad t \geq 0. \quad (4.3.3)$$

Wir beschränken uns im Folgenden auf den zeithomogenen Fall. Hier hängen die Übergangswahrscheinlichkeiten nicht von t ab, und es gilt

$$\mathcal{L}_t(x, y) = \mathcal{L}(x, y) = \lim_{h \searrow 0} \frac{p_h(x, y) - \delta(x, y)}{h} \quad \text{für alle } t \geq 0. \quad (4.3.4)$$

Wegen

$$\begin{aligned} p_h(x, y) &= h \cdot \mathcal{L}(x, y) + o(h) && \text{für } x \neq y, \\ p_h(x, x) &= 1 + h \cdot \mathcal{L}(x, x) + o(h), \end{aligned}$$

ist $\mathcal{L}(x, y)$ für $x \neq y$ die Sprungrate von x nach y , und $\mathcal{L}(x, x)$ ist die negative Wegsprungrate von x . Erfüllen die Übergangswahrscheinlichkeiten eines zeithomogenen Markovprozesses auf S die Bedingung (4.3.4) bzgl. eines zu spezifizierenden Konvergenzbegriffes, dann heißt die Matrix $\mathcal{L}(x, y)$ ($x, y \in S$) **infinitesimaler Generator** des Markovprozesses. Da $p_h(x, \bullet)$ für alle $h \geq 0$ eine Wahrscheinlichkeitsverteilung ist, sollten die Zeilensummen von \mathcal{L} in diesem Fall gleich null sein, d.h.

$$\mathcal{L}(x, x) = - \sum_{y \neq x} \mathcal{L}(x, y) \quad \text{für alle } x \in S. \quad (4.3.5)$$

Sei nun $\mathcal{L}(x, y)$ ($x, y \in S$) eine gegebene Matrix mit $\mathcal{L}(x, y) \geq 0$ für alle $x, y \in S$ und (4.3.5). Wir setzen zudem voraus, dass die Wegsprungraten $\mathcal{L}(x, x)$ beschränkt sind:

Annahme: Es existiert $\lambda \in (0, \infty)$ mit

$$\sum_{y \neq x} \mathcal{L}(x, y) = -\mathcal{L}(x, x) \leq \lambda \quad \text{für alle } x \in S. \quad (4.3.6)$$

Um einen Markovprozess mit Sprungraten $\mathcal{L}(x, y)$ zu konstruieren, betrachten wir unabhängige, $\text{Exp}(\lambda)$ -verteilte Zufallsvariablen T_1, T_2, \dots auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , die die zeitlichen Abstände zwischen möglichen Sprüngen des Prozesses beschreiben. Sei ferner $(Y_n)_{n=0,1,2,\dots}$ eine von $\sigma(T_1, T_2, \dots)$ unabhängige Markovkette auf (Ω, \mathcal{A}, P) mit Übergangswahrscheinlichkeiten

$$\begin{aligned} \pi(x, y) &= \frac{1}{\lambda} \mathcal{L}(x, y) \quad \text{für } y \neq x, \\ \pi(x, x) &= 1 - \sum_{y \neq x} \pi(x, y). \end{aligned}$$

Die Kette (Y_n) beschreibt die Zustände, die der zu konstruierende zeitstetige Sprungprozess durchläuft. Der Generator \mathcal{L} des zeitstetigen Prozesses und der Generator $\pi - I$ der Kette (Y_n) unterscheiden sich nur um einen Faktor λ :

$$\pi - I = \lambda^{-1} \mathcal{L}. \quad (4.3.7)$$

In der Tat stimmen die Matrizen auf beiden Seiten von (4.3.7) außerhalb der Diagonalen überein. Die Übereinstimmung auf der Diagonalen folgt, da die Zeilensummen beider Matrizen gleich null sind. Mit

$$N_t = \#\{n \in \mathbb{N} : T_1 + T_2 + \dots + T_n \leq t\}$$

erhalten wir:

Satz 4.12 (Konstruktion von zeitstetigen Markovketten). *Der Prozess $X_t := Y_{N_t}$ ist ein zeitstetiger Markovprozess mit Zustandsraum S , Übergangswahrscheinlichkeiten*

$$p_t(x, y) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} \pi^n(x, y), \quad x, y \in S, \quad (4.3.8)$$

und Generator $\mathcal{L}(x, y)$. Genauer gilt

$$\limsup_{h \searrow 0} \sup_{x \in S} \sum_{y \in S} \left| \frac{p_h(x, y) - \delta(x, y)}{h} - \mathcal{L}(x, y) \right| = 0. \quad (4.3.9)$$

Bemerkung. (1). *Matrixexponentialfunktion:* Die Übergangswahrscheinlichkeiten haben die Form

$$p_t = e^{-\lambda t} e^{\lambda t \pi} = e^{\lambda t(\pi - I)} = e^{t\mathcal{L}}, \quad (4.3.10)$$

wobei

$$(e^q)(x, y) = \sum_{n=0}^{\infty} \frac{1}{n!} q^n(x, y)$$

die Matrixexponentialfunktion ist. Hierbei ist $e^{\lambda t \pi}$ auch im abzählbar unendlichen Fall definiert, da die Matrizen $(\lambda t \pi)^n$ für alle $n \geq 0$ nichtnegativ sind. Die Reihe $e^{\lambda t(\pi - I)}$ konvergiert bzgl. der multiplikativen Matrixnorm

$$\|q\| := \sup_{x \in S} \sum_{y \in S} |q(x, y)|, \quad (4.3.11)$$

da $\|\lambda^n t^n (\pi - I)^n\| \leq (2\lambda t)^n$ für alle $n \geq 0$ gilt, und die Identität (4.3.10) folgt wegen

$$e^{-\lambda t} e^{\lambda t \pi} = e^{-\lambda t I} e^{\lambda t \pi} = e^{\lambda t(\pi - I)} = e^{t\mathcal{L}}.$$

(2). *Konvergenzbegriff:* Die Aussage (4.3.9) besagt, dass

$$\lim_{h \searrow 0} \frac{p_h - I}{h} = \mathcal{L}$$

bzgl. der durch (4.3.11) definierten Matrixnorm gilt. Die Voraussetzung (4.3.6) gewährleistet gerade, dass die Norm von \mathcal{L} endlich ist. In anderer Form ausgedrückt bedeutet (4.3.9), dass die signierten Maße $\frac{1}{h}(p_h(x, \bullet) - \delta(x, \bullet))$ für $h \searrow 0$ gleichmäßig in Variationsnorm (ℓ^1 -Norm) gegen $\mathcal{L}(x, \bullet)$ konvergieren. Eine entsprechende Aussage gilt auch (mit analogem Beweis), wenn der Zustandsraum S nicht abzählbar ist und die Wegsprungraten beschränkt sind.

Beweis. Seien $t, h \geq 0$ und $y \in S$. Um die Markoveigenschaft

$$P[X_{t+h} = y \mid \mathcal{F}_t^X] = p_h(X_t, y) \quad P\text{-fast sicher} \quad (4.3.12)$$

zu zeigen, verfahren wir ähnlich wie für Poisson-Prozesse in Satz 1.12. Seien zunächst $k, l \in \{0, 1, 2, \dots\}$ fest, und sei

$$\mathcal{G}_k = \sigma(T_1, \dots, T_k, Y_0, Y_1, \dots, Y_k)$$

die σ -Algebra, die den Verlauf des Prozesses bis zum k -ten Sprung beschreibt. Da die Wartezeiten T_i ($i \in \mathbb{N}$) und die Markovkette (Y_n) unabhängig voneinander sind, und N_t messbar bzgl. $\sigma(T_i : i \in \mathbb{N})$ ist, erhalten wir nach (1.4.3):

$$\begin{aligned} & P[N_t = k, N_{t+h} = k+l, Y_{k+l} = y \mid \mathcal{G}_k] \\ &= P[N_t = k, N_{t+h} = k+l \mid T_1, \dots, T_k] \cdot P[Y_{k+l} = y \mid Y_0, Y_1, \dots, Y_k] \\ &\stackrel{(1.4.3)}{=} P[N_t = k \mid T_1, \dots, T_k] \cdot P[N_h = l] \cdot \pi^l(Y_k, y) \\ &= P[N_t = k \mid \mathcal{G}_k] \cdot e^{-\lambda h} \frac{(\lambda h)^l}{l!} \pi^l(Y_k, y) \quad P\text{-fast sicher.} \end{aligned}$$

Durch Summieren über l folgt:

$$P[N_t = k, X_{t+h} = y \mid \mathcal{G}_k] = P[N_t = k \mid \mathcal{G}_k] \cdot e^{-\lambda h} \sum_{l=0}^{\infty} \frac{(\lambda h)^l}{l!} \pi^l(Y_k, y) \quad P\text{-f.s.} \quad (4.3.13)$$

Sei nun $A \in \mathcal{F}_t^X$. Ähnlich wie im Beweis von Satz 1.12 (3) zeigt man, dass dann ein Ereignis $A_k \in \mathcal{G}_k$ existiert mit

$$A \cap \{N_t = k\} = A_k \cap \{N_t = k\},$$

d.h. für $N_t = k$ hängt der Verlauf von X_s für $0 \leq s \leq t$ nur von den Zufallsvariablen T_1, \dots, T_k und Y_0, \dots, Y_k ab. Nach (4.3.13) folgt dann

$$\begin{aligned} & P[\{N_t = k\} \cap \{X_{t+h} = y\} \cap A] \\ &= E[P[N_t = k, X_{t+h} = y \mid \mathcal{G}_k]; A_k] \\ &= E[P[N_t = k \mid \mathcal{G}_k] \cdot p_h(Y_k, y); A_k] \\ &= E[p_h(X_t, y); A \cap \{N_t = k\}], \end{aligned}$$

wobei p_h wie in (4.3.8) definiert ist. Hierbei haben wir im letzten Schritt benutzt, dass $X_t = Y_k$ auf $\{N_t = k\}$ gilt. Durch Summieren über k erhalten wir schließlich

$$P[\{X_{t+h} = y\} \cap A] = E[p_h(X_t, y); A] \quad \text{für alle } A \in \mathcal{F}_t^X,$$

und damit (4.3.12).

Um den Generator zu identifizieren, verwenden wir die Identität $p_t = e^{t\mathcal{L}} = \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathcal{L}^n$ aus Bemerkung (1). Für $h \downarrow 0$ erhalten wir

$$\|p_h - I - h\mathcal{L}\| = \left\| \sum_{n=2}^{\infty} \frac{h^n}{n!} \mathcal{L}^n \right\| = O(h^2)$$

bezüglich der Norm aus (4.3.11), und damit (4.3.9). \square

4.4 Vorwärts- und Rückwärtsgleichungen

Mithilfe der Chapman-Kolmogorov-Gleichung ergeben sich aus der Definition des Generators Differentialgleichungen für die Zeitentwicklung der Übergangswahrscheinlichkeiten eines Markovprozesses. Wir beschränken uns auf den zeithomogenen Fall.

4.4.1 Vorwärts- und Rückwärtsgleichungen für Markovketten

Wir leiten zunächst Gleichungen für die Zeitentwicklung der Übergangswahrscheinlichkeiten von Markovketten her.

Zeitdiskreter Fall. Für die n -Schritt Übergangswahrscheinlichkeiten einer zeitdiskreten Markovkette mit Übergangskern π gilt

$$\pi^{n+1} - \pi^n = \pi^n(\pi - I) = (\pi - I)\pi^n \quad \text{für alle } n \geq 0. \quad (4.4.1)$$

Hierbei ist $\pi - I$ der Generator der Markovkette.

Zeitstetiger Fall. Im zeitstetigen Fall erhalten wir als infinitesimale Versionen von (4.4.1) Differentialgleichungen für die Zeitentwicklung der Übergangswahrscheinlichkeiten. Aus (4.3.9) und der Chapman-Kolmogorov-Gleichung (4.3.1) folgt unter der Annahme (4.3.6):

Satz 4.13 (Kolmogorovsche Vorwärts- und Rückwärtsgleichung). Für die Übergangsmatrizen $p_t(x, y)$ des in Satz 4.12 konstruierten Markovprozesses gilt

$$\lim_{h \rightarrow 0} \frac{p_{t+h} - p_t}{h} = p_t \mathcal{L} = \mathcal{L} p_t \quad \text{für alle } t \geq 0$$

mit Konvergenz bzgl. der in 4.3.11 definierten Matrixnorm $\|\bullet\|$. Insbesondere erfüllen die Übergangswahrscheinlichkeiten die **Kolmogorovsche Vorwärtsgleichung (Mastergleichung)**

$$\frac{d}{dt} p_t(x, y) = \sum_{z \in S} p_t(x, z) \mathcal{L}(z, y), \quad t \geq 0, \quad (4.4.2)$$

sowie die **Kolmogorovsche Rückwärtsgleichung**

$$\frac{d}{dt}p_t(x, y) = \sum_{z \in S} \mathcal{L}(x, z)p_t(z, y), \quad t \geq 0. \quad (4.4.3)$$

Beweis. Nach (4.3.9) gilt $\lim_{h \searrow 0} h^{-1}(p_h - I) = \mathcal{L}$ bzgl. der Matrixnorm $\|\cdot\|$. Da die Norm multiplikativ mit $\|p_t\| \leq 1$ ist, folgt für $t \geq 0$ und $h \searrow 0$ nach der Chapman-Kolmogorov-Gleichung

$$\begin{aligned} \left\| \frac{p_{t+h} - p_t}{h} - p_t \mathcal{L} \right\| &= \left\| p_t \left(\frac{p_h - I}{h} - \mathcal{L} \right) \right\| \leq \|p_t\| \cdot \left\| \frac{p_h - I}{h} - \mathcal{L} \right\| \\ &\leq \left\| \frac{p_h - I}{h} - \mathcal{L} \right\| \xrightarrow{h \searrow 0} 0. \end{aligned}$$

Entsprechend konvergiert auch

$$\begin{aligned} \left\| \frac{p_{t-h} - p_t}{-h} - p_t \mathcal{L} \right\| &= \left\| p_{t-h} \left(\frac{p_h - I}{h} - p_h \mathcal{L} \right) \right\| \leq \|p_{t-h}\| \cdot \left\| \frac{p_h - I}{h} - p_h \mathcal{L} \right\| \\ &\leq \left\| \frac{p_h - I}{h} - \mathcal{L} \right\| + \|I - p_h\| \cdot \|\mathcal{L}\| \end{aligned}$$

für $t \geq 0$ und $h \searrow 0$ gegen 0. Damit haben wir die Vorwärtsgleichung

$$\lim_{h \rightarrow 0} h^{-1}(p_{t+h} - p_t) = p_t \mathcal{L}$$

gezeigt, wobei der Grenzwert für $t = 0$ rechtsseitig ist. Der Beweis der Rückwärtsgleichung verläuft ähnlich. \square

Anschaulich können wir die *Vorwärtsgleichung* folgendermaßen interpretieren: Sei $x \in S$ ein fester Zustand. Dann beschreibt die Funktion

$$u(t, y) = p_t(x, y) = P[X_t = y \mid X_0 = x], \quad t \geq 0, \quad y \in S,$$

die Zeitentwicklung der Aufenthaltswahrscheinlichkeiten der Markovkette in Zuständen $y \in S$. Die Vorwärtsgleichung besagt, dass u das Anfangswertproblem

$$\begin{aligned} \frac{\partial u}{\partial t}(t, y) &= \sum_{z \in S} u(t, z) \mathcal{L}(z, y) \quad \text{für } t \geq 0, \\ u(0, y) &= \delta_x(y), \end{aligned}$$

löst. Die Wahrscheinlichkeitsmasse im Punkt y ändert sich also dadurch, dass Übergänge von anderen Zuständen z nach y mit den Raten $\mathcal{L}(z, y)$, bzw. Übergänge von y in andere Zustände mit der Rate $-\mathcal{L}(y, y)$ stattfinden. Bei der Analyse chemischer Reaktionen spielt die Vorwärtsgleichung eine wichtige Rolle – sie wird in den Naturwissenschaften meist als Mastergleichung

bezeichnet.

Für die *Rückwärtsgleichung* ergibt sich eine ähnliche, aber andere Interpretation: Seien hier $y \in S$ und $t \geq 0$ fest, und

$$v(s, x) = p_{t-s}(x, y) = P[X_t = y \mid X_s = x], \quad s \in [0, t], \quad x \in S.$$

Die Funktion v beschreibt die Abhängigkeit der Aufenthaltswahrscheinlichkeiten von dem zurückliegenden Startzeitpunkt und Anfangszustand des Markovprozesses. Die Rückwärtsgleichung besagt dann, dass v das folgende „Endwertproblem“ löst:

$$\begin{aligned} -\frac{\partial v}{\partial s}(s, x) &= \sum_{z \in S} \mathcal{L}(x, z)v(s, z), & s \in [0, t], \\ v(t, x) &= \delta_y(x). \end{aligned}$$

4.4.2 Gleichgewichte zeitstetiger Markovketten

Allgemeiner ergeben sich aus der Vorwärtsgleichung Zeitentwicklungsgleichungen für die Verteilungen μp_t des Markovprozesses mit beliebiger Startverteilung μ , und aus der Rückwärtsgleichung Zeitentwicklungsgleichungen für die Erwartungswerte $E[f(X_t) \mid X_s = x]$ von Funktionen $f : S \rightarrow \mathbb{R}$. Die Rückwärtsgleichung liefert auch eine infinitesimale Charakterisierung von Gleichgewichtsverteilungen zeitstetiger Markovketten:

Korollar 4.14 (Gleichgewichte zeitstetiger Markovketten). *Ist die Annahme (4.3.6) erfüllt, dann sind für eine Wahrscheinlichkeitsverteilung μ auf S die folgenden Aussagen äquivalent:*

(1). μ ist ein Gleichgewicht der Übergangshalbgruppe $(p_t)_{t \geq 0}$ aus (4.3.8), d.h.

$$\mu p_t = \mu \quad \text{für alle } t \geq 0.$$

(2). $\mu \mathcal{L} = 0$, d.h.

$$\sum_{x \in S} \mu(x) \mathcal{L}(x, y) = 0 \quad \text{für alle } y \in S.$$

Hierbei gewährleistet die Annahme (4.3.6) unter anderem, dass $\mu \mathcal{L}$ auch im abzählbar unendlichen Fall definiert ist.

Beweis. Anschaulich folgt aus der Rückwärtsgleichung

$$\frac{d}{dt} \mu p_t = \mu \mathcal{L} p_t \quad \text{für } t \geq 0, \quad \mu p_0 = \mu, \quad (4.4.4)$$

und damit die Aussage. Um dies auch im abzählbar unendlichen Fall zu rechtfertigen, verwenden wir die Variationsnorm (ℓ^1 -Norm) $\|\nu\|_{TV} = \frac{1}{2} \sum_{x \in S} |\nu(x)|$ von signierten Maßen. Für eine Matrix $q(x, y)$ ($x, y \in S$) und eine Wahrscheinlichkeitsverteilung μ gilt bezüglich der in (4.3.11) definierten Matrixnorm:

$$\|\mu q\|_{TV} \leq \|\mu\|_{TV} \cdot \|q\| = \|q\|/2.$$

Nach Satz 4.12 erhalten wir

$$\lim_{h \searrow 0} \left\| \frac{\mu p_{t+h} - \mu p_t}{h} - \mu \mathcal{L} p_t \right\|_{TV} \leq \|\mu\|_{TV} \cdot \lim_{h \searrow 0} \left\| \frac{p_{t+h} - p_t}{h} - \mathcal{L} p_t \right\| = 0,$$

und somit (4.4.4), wobei die Ableitung als Grenzwert der Differenzenquotienten in Variationsnorm definiert ist. \square

Aufbauend auf den obigen Resultaten kann man nun, ähnlich wie im zeitdiskreten Fall in Kapitel 6, die Rekurrenz und Transienz von zeitstetigen Markovketten untersuchen, mittlere Rückkehrzeiten und Trefferwahrscheinlichkeiten berechnen, und einen Ergodensatz beweisen. Unter der Annahme (4.3.6) können Rekurrenz und Transienz vollständig auf den zeitdiskreten Fall zurückgeführt werden, da der zeitstetige Prozess $X_t = Y_{N_t}$ dieselben Zustände durchläuft wie die zeitdiskrete Markovkette $(Y_n)_{n=0,1,2,\dots}$. Für die Herleitung von Differenzgleichungen für mittlere Rückkehrzeiten, Trefferwahrscheinlichkeiten usw., sowie den Beweis des Gesetzes der großen Zahlen im zeitstetigen Fall verweisen wir auf das Buch [Markov Chains] von J. R. Norris.

Wir betrachten abschließend ein Beispiel, das einen wichtigen Anwendungsbereich zeitstetiger Markovketten kurz anreißt:

Beispiel (M/M/1-Warteschlangenmodell). Im einfachsten Modell einer Warteschlange gibt es nur einen Server. Die Aufträge kommen jeweils nach unabhängigen, mit einem Parameter $\lambda > 0$ exponentialverteilten Wartezeiten beim Server an, und die Abstände zwischen den Bearbeitungszeiten zweier Aufträge sind ebenfalls unabhängig, und mit einem Parameter ν exponentialverteilt. Die beiden „M“s in M/M/1 stehen für gedächtnislose (engl. memoryless) Ankunfts- und Bearbeitungszeiten, die „1“ für die Anzahl der Server. Unter diesen (sehr restriktiven) Annahmen wird die Warteschlange durch eine zeitstetige Markovkette mit Zustandsraum $S = \{0, 1, 2, \dots\}$ und Übergangsraten

$$\mathcal{L}(x, x+1) = \lambda, \quad \mathcal{L}(x, x-1) = \nu,$$

beschrieben, d.h. durch einen zeitstetigen Birth-Death-Prozess.

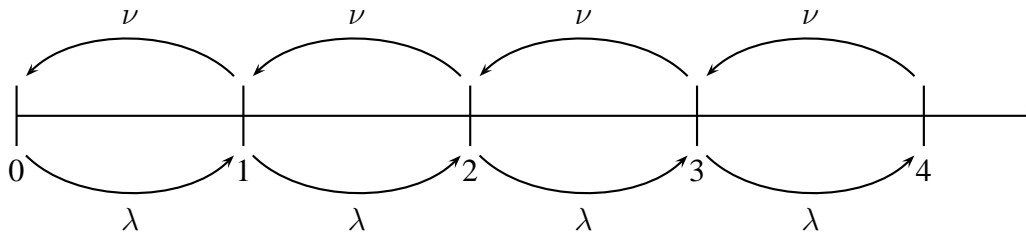


Abbildung 4.3: Übergangsraten einer M/M/1-Warteschlange.

Die Wegsprungraten $-\mathcal{L}(x, x)$ sind durch $\lambda + \nu$ beschränkt, und die Sprungkette (Y_n) hat die Übergangswahrscheinlichkeiten

$$\pi(x, x+1) = \frac{\lambda}{\lambda + \nu}, \quad \pi(x, x-1) = \frac{\nu}{\lambda + \nu} \quad \text{für } x > 0,$$

und $\pi(0, 0) = \frac{\nu}{\lambda + \nu}$. Insbesondere sind die Sprungkette, und damit auch die zeitstetige Markovkette, genau dann rekurrent, wenn $\lambda \leq \nu$ gilt. Die Gleichgewichtsbedingung $\mu \mathcal{L} = 0$ für den zeitstetigen Prozess lautet

$$\begin{aligned} -\mu(0) \cdot \lambda + \mu(1) \cdot \nu &= 0, \\ \mu(x-1) \cdot \lambda - \mu(x)(\lambda + \nu) + \mu(x+1) \cdot \nu &= 0 \quad \text{für } x \in \mathbb{N}. \end{aligned}$$

Für $\lambda \geq \nu$ existiert keine Gleichgewichtsverteilung, für $\lambda < \nu$ ist die geometrische Verteilung

$$\mu(x) = \left(1 - \frac{\lambda}{\nu}\right) \cdot \left(\frac{\lambda}{\nu}\right)^x, \quad x = 0, 1, 2, \dots,$$

das eindeutige Gleichgewicht. Aus dem Ergodensatz für zeitstetige Markovketten folgt dann beispielsweise, dass die mittlere Länge $\frac{1}{t} \int_0^t X_s ds$ der Warteschlange sich asymptotisch wie der Erwartungswert $\frac{\lambda}{\nu - \lambda}$ der Gleichgewichtsverteilung verhält.

4.4.3 Vorwärts- und Rückwärtsgleichung für die Brownsche Bewegung

Für allgemeine Markovprozesse ist die Herleitung von Vorwärts- und Rückwärtsgleichungen technisch häufig deutlich aufwändiger, da der infinitesimale Generator \mathcal{L} im Allgemeinen ein unbeschränkter linearer Operator ist. Dies ist bereits bei zeitstetigen Markovketten der Fall, wenn die Wegsprungraten nicht beschränkt sind. Für die Brownsche Bewegung erhalten wir die Kolmogorovschen Gleichungen unmittelbar aus der expliziten Form der Übergangsdichten

$$p_t(x, y) = (2\pi t)^{-d/2} \cdot \exp\left(-\frac{|x-y|^2}{2t}\right).$$

Als infinitesimaler Generator ergibt sich der Laplaceoperator:

Satz 4.15 (Brownsche Bewegung und Wärmeleitungsgleichung). Die Übergangsdichten $p_t(x, y)$ der Brownschen Bewegung bilden die Fundamentallösung der Wärmeleitungsgleichung, d.h. es gilt

$$\frac{\partial}{\partial t} p_t(x, y) = \frac{1}{2} \Delta_x p_t(x, y) = \frac{1}{2} \Delta_y p_t(x, y) \quad (4.4.5)$$

mit Anfangsbedingung

$$\lim_{t \searrow 0} \int p_t(x, y) f(y) dy = f(x) \quad \text{für alle } f \in C_b(\mathbb{R}^d) \text{ und } x \in \mathbb{R}^d, \quad (4.4.6)$$

bzw.

$$\lim_{t \searrow 0} \int g(x) p_t(x, y) dy = g(y) \quad \text{für alle } g \in C_b(\mathbb{R}^d) \text{ und } y \in \mathbb{R}^d. \quad (4.4.7)$$

Hierbei ist $\Delta_x = \sum_{i=1}^d \frac{\partial^2}{\partial x_i^2}$ der Laplace-Operator in der x -Variable.

Beweis. Die Gleichung (4.4.5) verifiziert man durch Nachrechnen. Für $x \in \mathbb{R}^d$ ist $p_t(x, y) dy$ eine Normalverteilung mit Mittelwertvektor x und Kovarianzmatrix $t \cdot I_d$. Hieraus folgt (4.4.6), da diese Wahrscheinlichkeitsverteilung für $t \searrow 0$ analog zu Beispiel 2 in Abschnitt ?? schwach gegen das Diracmaß δ_x konvergiert. Die Identität (4.4.7) folgt aus (4.4.6) wegen $p_t(x, y) = p_t(y, x)$. \square

Die Gleichung

$$\frac{\partial}{\partial t} p_t(x, y) = \frac{1}{2} \Delta_y p_t(x, y)$$

ist die *Vorwärtsgleichung*, und die Gleichung

$$\frac{\partial}{\partial t} p_t(x, y) = \frac{1}{2} \Delta_x p_t(x, y)$$

die *Rückwärtsgleichung* der Brownschen Bewegung. Anschaulich können wir die Vorwärtsgleichung auch folgendermaßen interpretieren: Für jedes Gebiet $D \subset \mathbb{R}^d$ mit glattem Rand gilt:

$$\begin{aligned} \frac{\partial}{\partial t} p_t(x, D) &= \int_D \frac{\partial}{\partial t} p_t(x, y) dy \\ &= \frac{1}{2} \int_D \Delta_y p_t(x, y) dy \\ &= \frac{1}{2} \int_{\partial D} n(y) \cdot \nabla_y p_t(x, y) \nu(dy), \end{aligned}$$

wobei n der äußere Normalenvektor und ν das Oberflächenmaß auf ∂D ist.

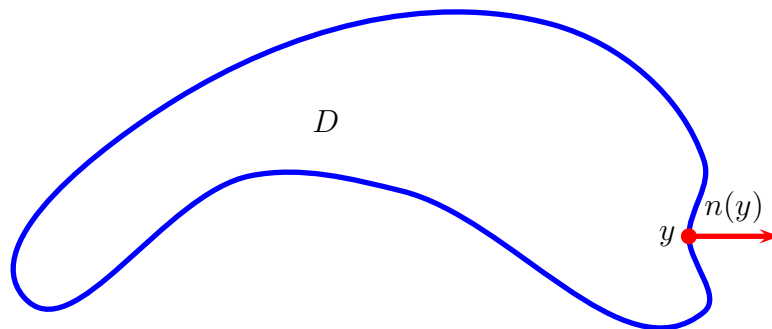


Abbildung 4.4: Äußerer Normalenvektor der Menge D im Punkt y .

Also beschreibt $\frac{1}{2}n(y) \cdot \nabla_y p_t(x, y)$ den Nettozufluss von Wahrscheinlichkeitsmasse pro Flächeneinheit durch ein infinitesimales Flächenstück mit Ausrichtung $n(y)$ am Punkt y .

Für Funktionen $f \in C_b^2(\mathbb{R}^d)$ ergeben sich aus (4.4.5) die Zeitentwicklungsgleichungen

$$\frac{\partial}{\partial t} p_t f = \frac{1}{2} \Delta p_t f = \frac{1}{2} p_t \Delta f$$

für die Erwartungswerte

$$(p_t f)(x) = \int_{\mathbb{R}^d} p_t(x, y) f(y) dy = E_x[f(B_t)].$$

Teil II

**Pfadeigenschaften von stochastischen
Prozessen**

Kapitel 5

Pfadeigenschaften der Brownschen Bewegung

In diesem Kapitel zeigen wir, dass Brownsche Bewegungen sich als zufällige Überlagerungen deterministischer Pfade darstellen und konstruieren lassen. Wir untersuchen die Regularität typischer Brownscher Pfade, und beweisen eine starke Markov-Eigenschaft und ein Reflektionsprinzip für die Brownsche Bewegung.

5.1 The Wiener-Lévy Construction

In this section we discuss how to construct Brownian motion as a random superposition of deterministic paths. The idea already goes back to N. Wiener, who constructed Brownian motion as a random Fourier series. The approach described here is slightly different and due to P. Lévy: The idea is to approximate the paths of Brownian motion on a finite time interval by their piecewise linear interpolations w.r.t. the sequence of dyadic partitions. This corresponds to a development of the Brownian paths w.r.t. Schauder functions which turns out to be very useful for applications including numerical simulation methods.

Our aim is to construct a one-dimensional Brownian motion B_t starting at 0 for $t \in [0, 1]$. By stationarity and independence of the increments, a Brownian motion defined for all $t \in [0, \infty)$ can then easily be obtained from infinitely many independent copies of Brownian motion on $[0, 1]$. We are hence looking for a random variable

$$B = (B_t)_{t \in [0,1]} : \Omega \longrightarrow C([0, 1])$$

defined on a probability space (Ω, \mathcal{A}, P) such that the distribution $P \circ B^{-1}$ is Wiener measure μ on the space of all continuous paths $x : [0, 1] \rightarrow \mathbb{R}$ such that $x(0) = 0$.

5.1.1 A first attempt

Recall that μ should be a kind of standard normal distribution w.r.t. the inner product

$$(x, y)_H = \int_0^1 \frac{dx}{dt} \frac{dy}{dt} dt \quad (5.1.1)$$

on functions $x, y : [0, 1] \rightarrow \mathbb{R}$ vanishing at 0. Therefore, we could try to define

$$B_t(\omega) := \sum_{i=1}^{\infty} Z_i(\omega) e_i(t) \quad \text{for } t \in [0, 1] \text{ and } \omega \in \Omega, \quad (5.1.2)$$

where $(Z_i)_{i \in \mathbb{N}}$ is a sequence of independent standard normal random variables, and $(e_i)_{i \in \mathbb{N}}$ is an orthonormal basis in the Hilbert space

$$H = \{x : [0, 1] \rightarrow \mathbb{R} : x(0) = 0, x \text{ is absolutely continuous with } (x, x)_H < \infty\}. \quad (5.1.3)$$

However, the resulting series approximation does not converge in H :

Satz 5.1. *Suppose $(e_i)_{i \in \mathbb{N}}$ is a sequence of orthonormal vectors in a Hilbert space H , and $(Z_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. random variables with $P[Z_i \neq 0] > 0$. Then the series $\sum_{i=1}^{\infty} Z_i(\omega) e_i$ diverges with probability 1 w.r.t. the norm on H .*

Beweis. By orthonormality and by the law of large numbers,

$$\left\| \sum_{i=1}^n Z_i(\omega) e_i \right\|_H^2 = \sum_{i=1}^n Z_i(\omega)^2 \longrightarrow \infty$$

P -almost surely as $n \rightarrow \infty$. □

The theorem again reflects the fact that a standard normal distribution on an infinite-dimensional Hilbert space can not be realized on the space itself.

To obtain a positive result, we will replace the norm

$$\|x\|_H = \left(\int_0^1 \left| \frac{dx}{dt} \right|^2 dt \right)^{1/2}$$

on H by the supremum norm

$$\|x\|_{\text{sup}} = \sup_{t \in [0, 1]} |x(t)|,$$

and, correspondingly, the Hilbert space H by the Banach space $C([0, 1])$. Note that the supremum norm is weaker than the H -norm. In fact, for $x \in H$ and $t \in [0, 1]$, the Cauchy-Schwarz inequality implies

$$|x(t)|^2 = \left| \int_0^t x'(s) ds \right|^2 \leq t \cdot \int_0^t |x'(s)|^2 ds \leq \|x\|_H^2,$$

and, therefore,

$$\|x\|_{\text{sup}} \leq \|x\|_H \quad \text{for any } x \in H.$$

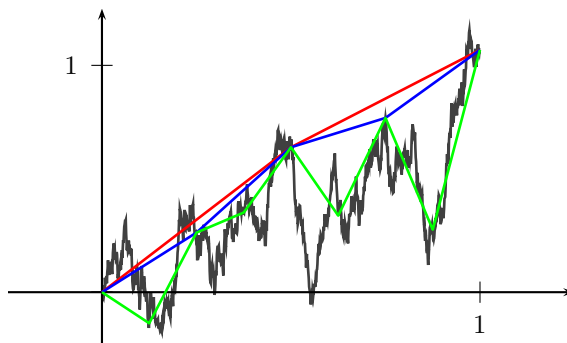
There are two choices for an orthonormal basis of the Hilbert space H that are of particular interest: The first is the Fourier basis given by

$$e_0(t) = t, \quad e_n(t) = \frac{\sqrt{2}}{\pi n} \sin(\pi n t) \quad \text{for } n \geq 1.$$

With respect to this basis, the series in (5.1.2) is a Fourier series with random coefficients. Wiener's original construction of Brownian motion is based on a *random Fourier series*. A second convenient choice is the basis of *Schauder functions* that has been used by P. Lévy to construct Brownian motion. Below, we will discuss Lévy's construction in detail. In particular, we will prove that for the Schauder functions, the series in (5.1.2) converges almost surely w.r.t. the supremum norm towards a continuous (but not absolutely continuous) random path $(B_t)_{t \in [0,1]}$. It is then not difficult to conclude that $(B_t)_{t \in [0,1]}$ is indeed a Brownian motion.

5.1.2 The Wiener-Lévy representation of Brownian motion

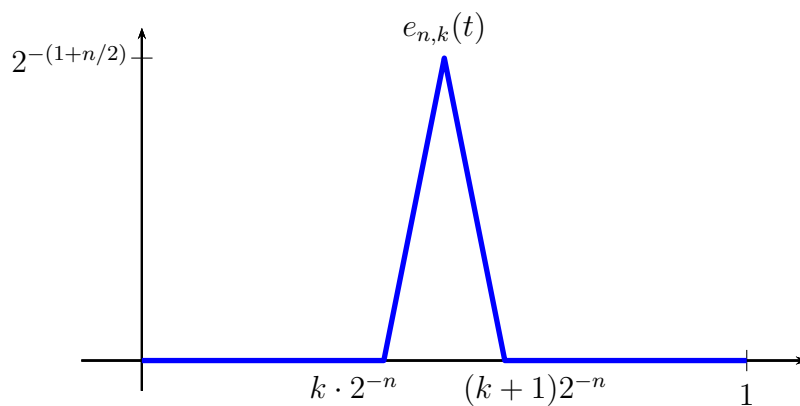
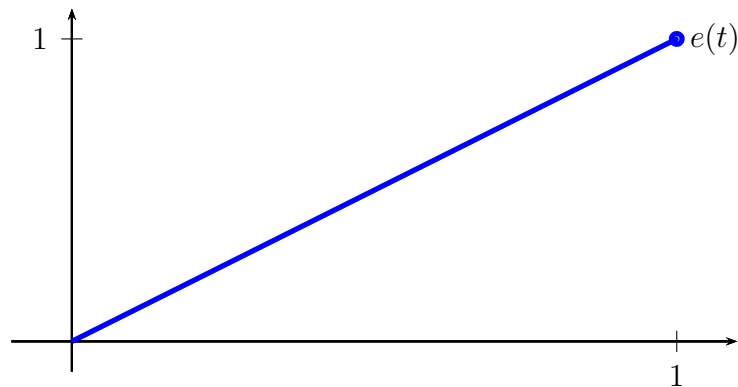
Before carrying out Lévy's construction of Brownian motion, we introduce the Schauder functions, and we show how to expand a given Brownian motion w.r.t. this basis of function space. Suppose we would like to approximate the paths $t \mapsto B_t(\omega)$ of a Brownian motion by their piecewise linear approximations adapted to the sequence of dyadic partitions of the interval $[0, 1]$.

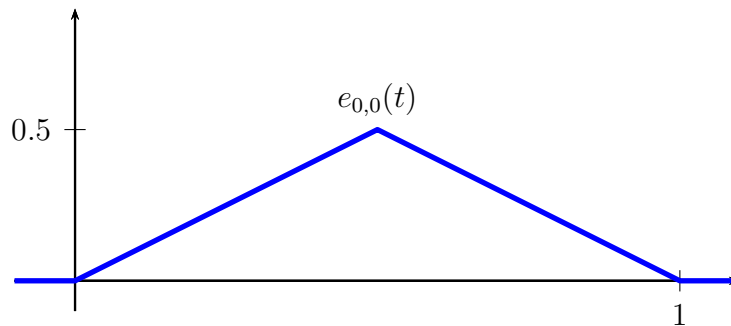


An obvious advantage of such an approximation over a Fourier expansion is that the values of the approximating functions at the dyadic points remain fixed once the approximating partition is fine enough. Moreover, piecewise linear approximations of a continuous function on $[0, 1]$ w.r.t. dyadic subdivisions can also be interpreted as an expansion in an appropriate basis on function space. For this purpose, we consider the base functions

$$e(t) = t, \quad \text{and} \\ e_{n,k}(t) = 2^{-n/2} e_{0,0}(2^n t - k), \quad n = 0, 1, 2, \dots, k = 0, 1, 2, \dots, 2^{n-1},$$

$$\text{where } e_{0,0}(t) = \min(t, 1-t)^+ = \begin{cases} t & \text{for } t \in [0, 1/2] \\ 1-t & \text{for } t \in (1/2, 1] \\ 0 & \text{for } t \in \mathbb{R} \setminus [0, 1] \end{cases} .$$





The functions $e_{n,k}$ ($n \geq 0, 0 \leq k < 2^n$) are called **Schauder functions**. It is easy to verify that piecewise linear approximation w.r.t. the m th dyadic partition of the unit interval corresponds to the expansion of a function $x \in C([0, 1])$ with $x(0) = 0$ in the basis given by $e(t)$ and the Schauder functions up to order $n < m$.

Above, the normalization constants in defining the functions $e_{n,k}$ have been chosen in such a way that the $e_{n,k}$ are orthonormal w.r.t. the H -inner product introduced before.

Definition (Orthonormal basis in an infinite dimensional Hilbert space). A sequence $(e_i)_{i \in \mathbb{N}}$ of vectors in an infinite-dimensional Hilbert space H is called an **orthonormal basis** (or **complete orthonormal system**) of H if and only if

(1). *Orthonormality:* $(e_i, e_j)_H = \delta_{ij}$ for any $i, j \in \mathbb{N}$, and

(2). *Completeness:* Any $h \in H$ can be expressed as

$$h = \sum_{i=1}^{\infty} (h, e_i)_H e_i.$$

Remark (Equivalent characterizations of orthonormal bases). Let e_i ($i \in \mathbb{N}$) be orthonormal vectors in a Hilbert space H . Then the following conditions are equivalent:

(1). $(e_i)_{i \in \mathbb{N}}$ is an orthonormal basis of H .

(2). The linear span

$$\text{span}\{e_i : i \in \mathbb{N}\} = \left\{ \sum_{i=1}^k c_i e_i : k \in \mathbb{N}, c_1, \dots, c_k \in \mathbb{R} \right\}$$

is a dense subset of H .

(3). There does not exist a non-zero element $x \in H$ such that $(x, e_i)_H = 0$ for every $i \in \mathbb{N}$.

(4). For any element $x \in H$, Parseval's relation

$$\|x\|_H^2 = \sum_{i=1}^{\infty} (x, e_i)_H^2 \quad (5.1.4)$$

holds.

(5). For any $x, y \in H$,

$$(x, y)_H = \sum_{i=1}^{\infty} (x, e_i)_H (y, e_i)_H. \quad (5.1.5)$$

For the proofs we refer to any book on functional analysis, cf. e.g. [Reed and Simon: Methods of modern mathematical physics, Vol. I].

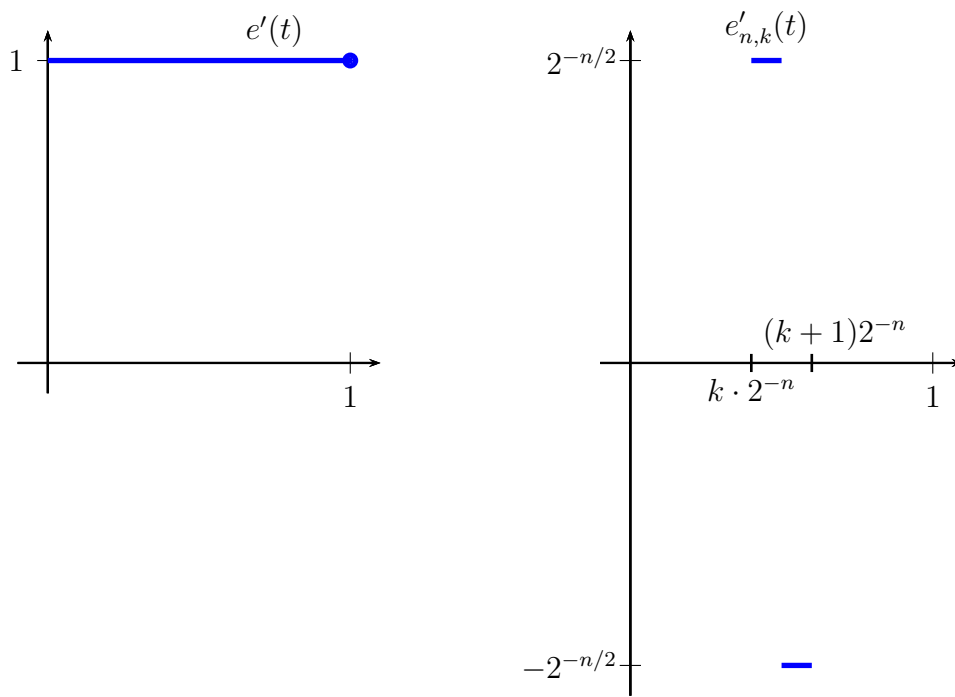
Lemma 5.2. *The Schauder functions e and $e_{n,k}$ ($n \geq 0, 0 \leq k < 2^n$) form an orthonormal basis in the Hilbert space H defined by (5.1.3).*

Beweis. By definition of the inner product on H , the linear map d/dt which maps an absolutely continuous function $x \in H$ to its derivative $x' \in L^2(0, 1)$ is an isometry from H onto $L^2(0, 1)$, i.e.,

$$(x, y)_H = (x', y')_{L^2(0,1)} \quad \text{for any } x, y \in H.$$

The derivatives of the Schauder functions are the Haar functions

$$e'(t) = 1, \quad e'_{n,k}(t) = 2^{n/2} (I_{(k \cdot 2^{-n}, (k+1/2) \cdot 2^{-n})}(t) - I_{((k+1/2) \cdot 2^{-n}, (k+1) \cdot 2^{-n})}(t)) \quad \text{for a.e. } t.$$



It is easy to see that these functions form an orthonormal basis in $L^2(0, 1)$. In fact, orthonormality w.r.t. the L^2 inner product can be verified directly. Moreover, the linear span of the functions e' and $e'_{n,k}$ for $n = 0, 1, \dots, m$ and $k = 0, 1, \dots, 2^n - 1$ consists of all step functions that are constant on each dyadic interval $[j \cdot 2^{-(m+1)}, (j+1) \cdot 2^{-(m+1)})$. An arbitrary function in $L^2(0, 1)$ can be approximated by dyadic step functions w.r.t. the L^2 norm. This follows for example directly from the L^2 martingale convergence Theorem, cf. [Introduction to Stochastic Analysis]. Hence the linear span of e' and the Haar functions $e'_{n,k}$ is dense in $L^2(0, 1)$, and therefore these functions form an orthonormal basis of the Hilbert space $L^2(0, 1)$. Since $x \mapsto x'$ is an isometry from H onto $L^2(0, 1)$, we can conclude that e and the Schauder functions $e_{n,k}$ form an orthonormal basis of H . \square

The expansion of a function $x : [0, 1] \rightarrow \mathbb{R}$ in the basis of Schauder functions can now be made explicit. The coefficients of a function $x \in H$ in the expansion are

$$(x, e)_H = \int_0^1 x' e' dt = \int_0^1 x' dt = x(1) - x(0) = x(1)$$

$$\begin{aligned}
(x, e_{n,k})_H &= \int_0^1 x' e'_{n,k} dt = 2^{n/2} \int_0^1 x'(t) e'_{0,0}(2^n t - k) dt \\
&= 2^{n/2} \left[x\left(\left(k + \frac{1}{2}\right) \cdot 2^{-n}\right) - x\left(k \cdot 2^{-n}\right) - \left(x\left((k+1) \cdot 2^{-n}\right) - x\left(\left(k + \frac{1}{2}\right) \cdot 2^{-n}\right)\right) \right].
\end{aligned}$$

Satz 5.3 (Series expansion of continuous paths). *Let $x \in C([0, 1])$ with $x(0) = 0$. Then the expansion*

$$x(t) = x(1)e(t) - \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} 2^{n/2} \Delta_{n,k} x \cdot e_{n,k}(t),$$

$$\Delta_{n,k} x = \left[x\left(\left(k+1\right) \cdot 2^{-n}\right) - x\left(\left(k + \frac{1}{2}\right) \cdot 2^{-n}\right) - \left(x\left(\left(k + \frac{1}{2}\right) \cdot 2^{-n}\right) - x\left(k \cdot 2^{-n}\right)\right) \right]$$

holds w.r.t. uniform convergence on $[0, 1]$. For $x \in H$ the series also converges w.r.t. the stronger H -norm.

Beweis. It can be easily verified that by definition of the Schauder functions, for each $m \in \mathbb{N}$ the partial sum

$$x^{(m)}(t) := x(1)e(t) - \sum_{n=0}^m \sum_{k=0}^{2^n-1} 2^{n/2} \Delta_{n,k} x \cdot e_{n,k}(t) \quad (5.1.6)$$

is the polygonal interpolation of $x(t)$ w.r.t. the $(m+1)$ -th dyadic partition of the interval $[0, 1]$. Since the function x is uniformly continuous on $[0, 1]$, the polygonal interpolations converge uniformly to x . This proves the first statement. Moreover, for $x \in H$, the series is the expansion of x in the orthonormal basis of H given by the Schauder functions, and therefore it also converges w.r.t. the H -norm. \square

Applying the expansion to the paths of a Brownian motions, we obtain:

Korollar 5.4 (Wiener-Lévy representation). *For a Brownian motion $(B_t)_{t \in [0,1]}$ the series representation*

$$B_t(\omega) = Z(\omega)e(t) + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} Z_{n,k}(\omega) e_{n,k}(t), \quad t \in [0, 1], \quad (5.1.7)$$

holds w.r.t. uniform convergence on $[0, 1]$ for P -almost every $\omega \in \Omega$, where

$$Z := B_1, \quad \text{and} \quad Z_{n,k} := -2^{n/2} \Delta_{n,k} B \quad (n \geq 0, 0 \leq k \leq 2^n - 1)$$

are independent random variables with standard normal distribution.

Beweis. It only remains to verify that the coefficients Z and $Z_{n,k}$ are independent with standard normal distribution. A vector given by finitely many of these random variables has a multivariate normal distribution, since it is a linear transformation of increments of the Brownian motion B_t . Hence it suffices to show that the random variables are uncorrelated with variance 1. This is left as an exercise to the reader. \square

5.1.3 Lévy's construction of Brownian motion

The series representation (5.1.7) can be used to construct Brownian motion from a sequence of independent standard normal random variables. The resulting construction does not only prove existence of Brownian motion but it is also very useful for numerical implementations:

Satz 5.5 (Construction of Brownian motion; P. Lévy 1948). *Let Z and $Z_{n,k}$ ($n \geq 0, 0 \leq k \leq 2^n - 1$) be independent standard normally distributed random variables on a probability space (Ω, \mathcal{A}, P) . Then the series in (5.1.7) converges uniformly on $[0, 1]$ with probability 1. The limit process $(B_t)_{t \in [0,1]}$ is a Brownian motion.*

The convergence proof relies on a combination of the Borel-Cantelli Lemma and the Weierstrass criterion for uniform convergence of series of functions. Moreover, we will need the following result to identify the limit process as a Brownian motion:

Lemma 5.6 (Parseval relation for Schauder functions). *For any $s, t \in [0, 1]$,*

$$e(t)e(s) + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} e_{n,k}(t)e_{n,k}(s) = \min(t, s).$$

Beweis. Note that for $g \in H$ and $s \in [0, 1]$, we have

$$g(s) = g(s) - g(0) = \int_0^1 g' \cdot I_{(0,s)} = (g, h^{(s)})_H,$$

where $h^{(s)}(t) := \int_0^t I_{(0,s)} = \min(s, t)$. Hence the Parseval relation (5.1.4) applied to the functions $h^{(s)}$ and $h^{(t)}$ yields

$$\begin{aligned} e(t)e(s) + \sum_{n,k} e_{n,k}(t)e_{n,k}(s) &= (e, h^{(t)})(e, h^{(s)}) + \sum_{n,k} (e_{n,k}, h^{(t)})(e_{n,k}, h^{(s)}) \\ &= (h^{(t)}, h^{(s)}) = \int_0^1 I_{(0,t)} I_{(0,s)} = \min(t, s). \end{aligned}$$

□

Proof of Theorem 5.5. We proceed in four steps:

- (1). *Uniform convergence for P -a.e. ω :* By the Weierstrass criterion, a series of functions converges uniformly if the sum of the supremum norms of the summands is finite. To apply the criterion, we note that for any fixed $t \in [0, 1]$ and $n \in \mathbb{N}$, only one of the functions $e_{n,k}$, $k = 0, 1, \dots, 2^n - 1$, does not vanish at t . Moreover, $|e_{n,k}(t)| \leq 2^{-n/2}$. Hence

$$\sup_{t \in [0,1]} \left| \sum_{k=0}^{2^n-1} Z_{n,k}(\omega) e_{n,k}(t) \right| \leq 2^{-n/2} \cdot M_n(\omega), \quad (5.1.8)$$

where

$$M_n := \max_{0 \leq k < 2^n} |Z_{n,k}|.$$

We now apply the Borel-Cantelli Lemma to show that with probability 1, M_n grows at most linearly. Let Z denote a standard normal random variable. Then we have

$$\begin{aligned} P[M_n > n] &\leq 2^n \cdot P[|Z| > n] \leq \frac{2^n}{n} \cdot E[|Z|; |Z| > n] \\ &= \frac{2 \cdot 2^n}{n \cdot \sqrt{2\pi}} \int_n^\infty x e^{-x^2/2} dx = \sqrt{\frac{2}{\pi}} \frac{2^n}{n} \cdot e^{-n^2/2} \end{aligned}$$

for any $n \in \mathbb{N}$. Since the sequence on the right hand side is summable, $M_n \leq n$ holds eventually with probability one. Therefore, the sequence on the right hand side of (5.1.8) is also summable for P -almost every ω . Hence, by (5.1.8) and the Weierstrass criterion, the partial sums

$$B_t^{(m)}(\omega) = Z(\omega)e(t) + \sum_{n=0}^m \sum_{k=0}^{2^n-1} Z_{n,k}(\omega) e_{n,k}(t), \quad m \in \mathbb{N},$$

converge almost surely uniformly on $[0, 1]$. Let

$$B_t = \lim_{m \rightarrow \infty} B_t^{(m)}$$

denote the almost surely defined limit.

- (2). *L^2 convergence for fixed t :* We now want to prove that the limit process (B_t) is a Brownian motion, i.e., a continuous Gaussian process with $E[B_t] = 0$ and $\text{Cov}[B_t, B_s] = \min(t, s)$ for any $t, s \in [0, 1]$. To compute the covariances we first show that for a given $t \in [0, 1]$ the

series approximations $B_t^{(m)}$ of B_t converge also in L^2 . Let $l, m \in \mathbb{N}$ with $l < m$. Since the $Z_{n,k}$ are independent (and hence uncorrelated) with variance 1, we have

$$E[(B_t^{(m)} - B_t^{(l)})^2] = E \left[\left(\sum_{n=l+1}^m \sum_{k=0}^{2^n-1} Z_{n,k} e_{n,k}(t) \right)^2 \right] = \sum_{n=l+1}^m \sum_k e_{n,k}(t)^2.$$

The right hand side converges to 0 as $l, m \rightarrow \infty$ since $\sum_{n,k} e_{n,k}(t)^2 < \infty$ by Lemma 5.6. Hence $B_t^{(m)}, m \in \mathbb{N}$, is a Cauchy sequence in $L^2(\Omega, \mathcal{A}, P)$. Since $B_t = \lim_{m \rightarrow \infty} B_t^{(m)}$ almost surely, we obtain

$$B_t^{(m)} \xrightarrow{m \rightarrow \infty} B_t \quad \text{in } L^2(\Omega, \mathcal{A}, P).$$

(3). *Expectations and Covariances:* By the L^2 convergence we obtain for any $s, t \in [0, 1]$:

$$E[B_t] = \lim_{m \rightarrow \infty} E[B_t^{(m)}] = 0, \quad \text{and}$$

$$\begin{aligned} \text{Cov}[B_t, B_s] &= E[B_t B_s] = \lim_{m \rightarrow \infty} E[B_t^{(m)} B_s^{(m)}] \\ &= e(t)e(s) + \lim_{m \rightarrow \infty} \sum_{n=0}^m \sum_{k=0}^{2^n-1} e_{n,k}(t)e_{n,k}(s). \end{aligned}$$

Here we have used again that the random variables Z and $Z_{n,k}$ are independent with variance 1. By Parseval's relation (Lemma 5.6), we conclude

$$\text{Cov}[B_t, B_s] = \min(t, s).$$

Since the process $(B_t)_{t \in [0,1]}$ has the right expectations and covariances, and, by construction, almost surely continuous paths, it only remains to show that (B_t) is a Gaussian process in order to complete the proof:

(4). $(B_t)_{t \in [0,1]}$ is a Gaussian process: We have to show that $(B_{t_1}, \dots, B_{t_l})$ has a multivariate normal distribution for any $0 \leq t_1 < \dots < t_l \leq 1$. By Theorem 4.5, it suffices to verify that any linear combination of the components is normally distributed. This holds by the next lemma since

$$\sum_{j=1}^l p_j B_{t_j} = \lim_{m \rightarrow \infty} \sum_{j=1}^l p_j B_{t_j}^{(m)} \quad P\text{-a.s.}$$

is an almost sure limit of normally distributed random variables for any $p_1, \dots, p_l \in \mathbb{R}$.

Combining Steps 3, 4 and the continuity of sample paths, we conclude that $(B_t)_{t \in [0,1]}$ is indeed a Brownian motion. \square

Lemma 5.7 (Limits of normal random variables). *Suppose that $(X_n)_{n \in \mathbb{N}}$ is a sequence of normally distributed random variables defined on a joint probability space (Ω, \mathcal{A}, P) , and X_n converges almost surely to a random variable X . Then X is also normally distributed.*

Beweis. Suppose $X_n \sim N(m_n, \sigma_n^2)$ with $m_n \in \mathbb{R}$ and $\sigma_n \in [0, \infty)$. By the Dominated Convergence Theorem,

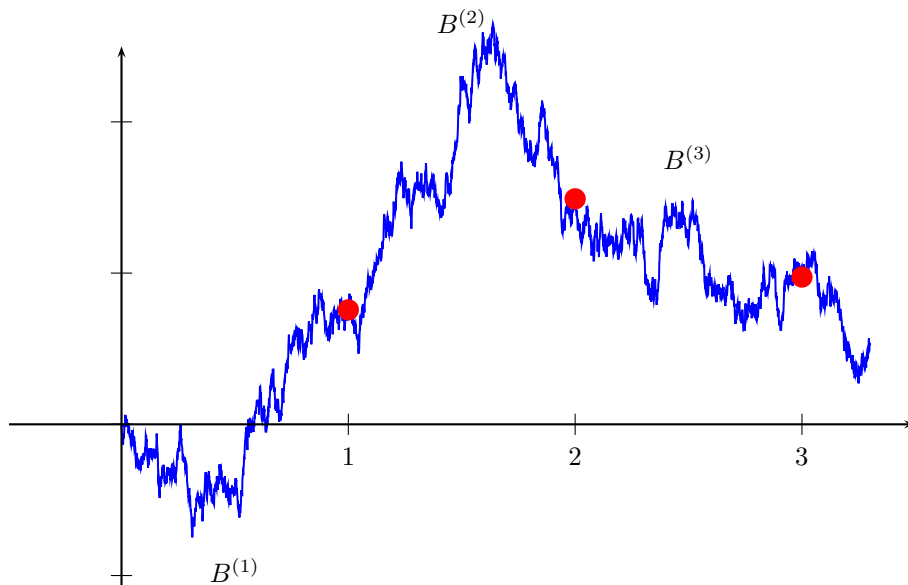
$$E[e^{ipX}] = \lim_{n \rightarrow \infty} E[e^{ipX_n}] = \lim_{n \rightarrow \infty} e^{ipm_n} e^{-\frac{1}{2}\sigma_n^2 p^2}.$$

The limit on the right hand side only exists for all p , if either $\sigma_n \rightarrow \infty$, or the sequences σ_n and m_n both converge to finite limits $\sigma \in [0, \infty)$ and $m \in \mathbb{R}$. In the first case, the limit would equal 0 for $p \neq 0$ and 1 for $p = 0$. This is a contradiction, since characteristic functions are always continuous. Hence the second case occurs, and, therefore

$$E[e^{ipX}] = e^{ipm - \frac{1}{2}\sigma^2 p^2} \quad \text{for any } p \in \mathbb{R},$$

i.e., $X \sim N(m, \sigma^2)$. □

So far, we have constructed Brownian motion only for $t \in [0, 1]$. Brownian motion on any finite time interval can easily be obtained from this process by rescaling. Brownian motion defined for all $t \in \mathbb{R}_+$ can be obtained by joining infinitely many Brownian motions on time intervals of length 1:



Satz 5.8 (Brownian motion for $t \in \mathbb{R}_+$). *Suppose that $B_t^{(1)}, B_t^{(2)}, \dots$ are independent Brownian motions starting at 0 defined for $t \in [0, 1]$. Then the process*

$$B_t := B_{t - \lfloor t \rfloor}^{|\lfloor t \rfloor + 1} + \sum_{i=1}^{\lfloor t \rfloor} B_1^{(i)}, \quad t \geq 0,$$

is a Brownian motion defined for $t \in [0, \infty)$.

The proof is left as an exercise.

5.2 Regularity of Brownian paths

In this section we study some properties of Brownian sample paths in dimension one. We show that a typical Brownian path is nowhere differentiable, whereas it is Hölder-continuous with parameter α if and only if $\alpha < 1/2$. Furthermore, the set $\Lambda_a = \{t \geq 0 : B_t = a\}$ of all passage times of a given point $a \in \mathbb{R}$ is a fractal. We will show that almost surely, Λ_a has Lebesgue measure zero but any point in Λ_a is an accumulation point of Λ_a .

Throughout this section, we consider a one-dimensional Brownian motion $(B_t)_{t \geq 0}$ with $B_0 = 0$ defined on a probability space (Ω, \mathcal{A}, P) .

5.2.1 Typical Brownian sample paths are nowhere differentiable

For any $t \geq 0$ and $h > 0$, the difference quotient $\frac{B_{t+h} - B_t}{h}$ is normally distributed with mean 0 and standard deviation

$$\sigma[(B_{t+h} - B_t)/h] = \sigma[B_{t+h} - B_t]/h = 1/\sqrt{h}.$$

This suggests that the derivative

$$\frac{d}{dt} B_t = \lim_{h \searrow 0} \frac{B_{t+h} - B_t}{h}$$

does not exist. Indeed, we have the following stronger statement.

Satz 5.9 (Nowhere differentiability of Brownian paths; Paley, Wiener, Zygmund 1933). *Almost surely, the Brownian sample path $t \mapsto B_t$ is nowhere differentiable, and*

$$\limsup_{s \searrow t} \left| \frac{B_s - B_t}{s - t} \right| = \infty \quad \text{for any } t \geq 0.$$

Note that, since there are uncountably many $t \geq 0$, the statement is stronger than claiming only the almost sure non-differentiability for any given $t \geq 0$.

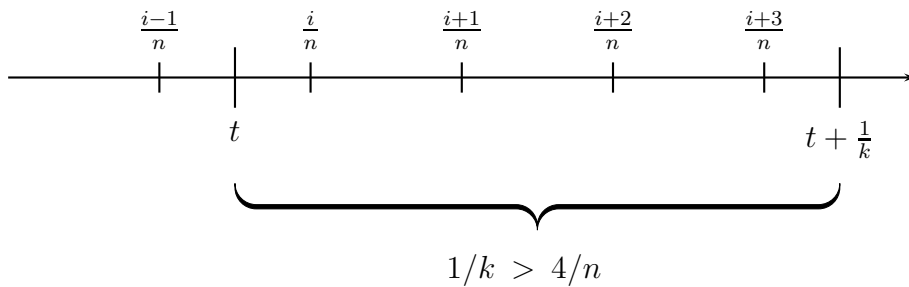
Beweis. It suffices to show that the set

$$N = \left\{ \omega \in \Omega : \exists t \in [0, T], k, L \in \mathbb{N} : \text{for all } s \in (t, t + \frac{1}{k}), |B_s(\omega) - B_t(\omega)| \leq L|s - t| \right\}$$

is a null set for any $T \in \mathbb{N}$. Hence fix $T \in \mathbb{N}$, and consider $\omega \in N$. Then there exist $k, L \in \mathbb{N}$ and $t \in [0, T]$ such that

$$|B_s(\omega) - B_t(\omega)| \leq L \cdot |s - t| \quad \text{holds for } s \in (t, t + \frac{1}{k}). \quad (5.2.1)$$

To make use of the independence of the increments over disjoint intervals, we note that for any $n > 4k$, we can find an $i \in \{1, 2, \dots, nT\}$ such that the intervals $(\frac{i}{n}, \frac{i+1}{n})$, $(\frac{i+1}{n}, \frac{i+2}{n})$, and $(\frac{i+2}{n}, \frac{i+3}{n})$ are all contained in $(t, t + \frac{1}{k})$:



Hence by (5.2.1), the bound

$$\begin{aligned} \left| B_{\frac{i+1}{n}}(\omega) - B_{\frac{i}{n}}(\omega) \right| &\leq \left| B_{\frac{i+1}{n}}(\omega) - B_t(\omega) \right| + \left| B_t(\omega) - B_{\frac{i}{n}}(\omega) \right| \\ &\leq L \cdot \left(\frac{i+1}{n} - t \right) + L \cdot \left(t - \frac{i}{n} \right) \leq \frac{8L}{n} \end{aligned}$$

holds for $j = i, i+1, i+2$. Thus we have shown that N is contained in the set

$$\tilde{N} := \bigcup_{k, L \in \mathbb{N}} \bigcap_{n > 4k} \bigcup_{i=1}^{nT} \left\{ \left| B_{\frac{j+1}{n}} - B_{\frac{j}{n}} \right| \leq \frac{8L}{n} \quad \text{for } j = i, i+1, i+2 \right\}.$$

We now prove $P[\tilde{N}] = 0$. By independence and stationarity of the increments we have

$$\begin{aligned} &P \left[\left\{ \left| B_{\frac{j+1}{n}} - B_{\frac{j}{n}} \right| \leq \frac{8L}{n} \quad \text{for } j = i, i+1, i+2 \right\} \right] \\ &= P \left[\left| B_{\frac{1}{n}} \right| \leq \frac{8L}{n} \right]^3 = P \left[|B_1| \leq \frac{8L}{\sqrt{n}} \right]^3 \\ &\leq \left(\frac{1}{\sqrt{2\pi}} \frac{16L}{\sqrt{n}} \right)^3 = \frac{16^3}{\sqrt{2\pi}^3} \cdot \frac{L^3}{n^{3/2}} \end{aligned} \quad (5.2.2)$$

for any i and n . Here we have used that the standard normal density is bounded from above by $1/\sqrt{2\pi}$. By (5.2.2) we obtain for $k, L \in \mathbb{N}$,

$$\begin{aligned} P \left[\bigcap_{n>4k} \bigcup_{i=1}^{nT} \left\{ \left| B_{\frac{i+1}{n}} - B_{\frac{i}{n}} \right| \leq \frac{8L}{n} \text{ for } j = i, i+1, i+2 \right\} \right] \\ \leq \frac{16^3}{\sqrt{2\pi}^3} \cdot \inf_{n>4k} (nTL^3/n^{3/2}) = 0. \end{aligned}$$

Hence, $P[\tilde{N}] = 0$, and therefore N is a null set. \square

5.2.2 Hölder continuity

The statement of Theorem 5.9 says that a typical Brownian path is not Lipschitz continuous on any non-empty open interval. On the other hand, the Wiener-Lévy construction shows that the sample paths are continuous. We can almost close the gap between these two statements by arguing in both cases slightly more carefully:

Satz 5.10 (Hölder continuity of typical Brownian paths). *The following statements hold almost surely:*

(1). For any $\alpha > 1/2$,

$$\limsup_{s \searrow t} \frac{|B_s - B_t|}{|s - t|^\alpha} = \infty \quad \text{for all } t \geq 0.$$

(2). For any $\alpha < 1/2$,

$$\sup_{\substack{s, t \in [0, T] \\ s \neq t}} \frac{|B_s - B_t|}{|s - t|^\alpha} < \infty \quad \text{for all } T > 0.$$

Hence a typical Brownian path is nowhere Hölder continuous with parameter $\alpha > 1/2$, but it is Hölder continuous with parameter $\alpha < 1/2$ on any finite interval. The critical case $\alpha = 1/2$ is more delicate, and will be briefly discussed below.

Proof of Theorem 5.10. The first statement can be shown by a similar argument as in the proof of Theorem 5.9. The details are left to the reader.

To prove the second statement for $T = 1$, we use the Wiener-Lévy representation

$$B_t = Z \cdot t + \sum_{n=0}^{\infty} \sum_{k=0}^{2^n-1} Z_{n,k} e_{n,k}(t) \quad \text{for any } t \in [0, 1]$$

with independent standard normal random variables $Z, Z_{n,k}$. For $t, s \in [0, 1]$ we obtain

$$|B_t - B_s| \leq |Z| \cdot |t - s| + \sum_n M_n \sum_k |e_{n,k}(t) - e_{n,k}(s)|,$$

where $M_n := \max_k |Z_{n,k}|$ as in the proof of Theorem 5.5. We have shown above that by the Borel-Cantelli Lemma, $M_n \leq n$ eventually with probability one, and hence

$$M_n(\omega) \leq C(\omega) \cdot n$$

for some almost surely finite constant $C(\omega)$. Moreover, note that for each s, t and n , at most two summands in $\sum_k |e_{n,k}(t) - e_{n,k}(s)|$ do not vanish. Since $|e_{n,k}(t)| \leq \frac{1}{2} \cdot 2^{-n/2}$ and $|e'_{n,k}(t)| \leq 2^{n/2}$, we obtain the estimates

$$|e_{n,k}(t) - e_{n,k}(s)| \leq 2^{-n/2}, \quad \text{and} \quad (5.2.3)$$

$$|e_{n,k}(t) - e_{n,k}(s)| \leq 2^{n/2} \cdot |t - s|. \quad (5.2.4)$$

For given $s, t \in [0, 1]$, we now choose $N \in \mathbb{N}$ such that

$$2^{-N} \leq |t - s| < 2^{1-N}. \quad (5.2.5)$$

By applying (5.2.3) for $n > N$ and (5.2.4) for $n \leq N$, we obtain

$$|B_t - B_s| \leq |Z| \cdot |t - s| + 2C \cdot \left(\sum_{n=1}^N n 2^{n/2} \cdot |t - s| + \sum_{n=N+1}^{\infty} n 2^{-n/2} \right).$$

By (5.2.5) the sums on the right hand side can both be bounded by a constant multiple of $|t - s|^\alpha$ for any $\alpha < 1/2$. This proves that $(B_t)_{t \in [0,1]}$ is almost surely Hölder-continuous of order α . \square

5.2.3 Law of the iterated logarithm

Khintchine's version of the law of the iterated logarithm is a much more precise statement on the local regularity of a typical Brownian path at a fixed time $s \geq 0$. It implies in particular that almost every Brownian path is not Hölder continuous with parameter $\alpha = 1/2$. We state the result without proof:

Satz 5.11 (Law of the iterated logarithm for Brownian motion; Khintchine 1933). *For $s \geq 0$, the following statements hold almost surely:*

$$\limsup_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log \log(1/t)}} = +1, \quad \text{and} \quad \liminf_{t \searrow 0} \frac{B_{s+t} - B_s}{\sqrt{2t \log \log(1/t)}} = -1.$$

For the proof cf. e.g. Breiman, Probability, Section 12.9. By a time inversion, the theorem translates into a statement on the global asymptotics of Brownian paths:

Korollar 5.12 (Global law of the iterated logarithm for Brownian motion). *The following statements hold almost surely:*

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} = +1, \quad \text{and} \quad \liminf_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log t}} = -1.$$

Beweis. This follows by applying the Theorem above to the Brownian motion $\widehat{B}_t = t \cdot B_{1/t}$. For example, substituting $h = 1/t$, we have

$$\limsup_{t \rightarrow \infty} \frac{B_t}{\sqrt{2t \log \log(t)}} = \limsup_{h \searrow 0} \frac{h \cdot B_{1/h}}{\sqrt{2h \log \log 1/h}} = +1$$

almost surely. □

The corollary is a continuous time analogue of the law of the iterated logarithm for random walks stating that for $S_n = \sum_{i=1}^n \eta_i$, η_i i.i.d. with $E[\eta_i] = 0$ and $\text{Var}[\eta_i] = 1$, one has

$$\limsup_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = +1 \quad \text{and} \quad \liminf_{n \rightarrow \infty} \frac{S_n}{\sqrt{2n \log \log n}} = -1$$

almost surely. In fact, one way to prove the LIL for random walks is to embed a random walk into a Brownian motion, cf. e.g. Rogers and Williams, Vol. I, Ch. 7.

5.2.4 Level sets of Brownian paths

We now study the set of passage times to a given level a for a one-dimensional Brownian motion $(B_t)_{t \geq 0}$. This set has interesting properties – in particular it is a *random fractal*. Fix $a \in \mathbb{R}$, and let

$$\Lambda_a(\omega) = \{t \geq 0 : B_t(\omega) = a\} \subseteq [0, \infty).$$

Assuming that every path is continuous, the random set $\Lambda_a(\omega)$ is *closed* for every ω . Moreover, scale invariance of Brownian motion implies a *statistical self similarity* property for the level sets: Since the rescaled process $(c^{-1/2} B_{ct})_{t \geq 0}$ has the same distribution as $(B_t)_{t \geq 0}$ for any $c > 0$, we can conclude that the set valued random variable $c \cdot \Lambda_{a/\sqrt{c}}$ has the same distribution as Λ_a . In particular, Λ_0 is a *fractal* in the sense that

$$\Lambda_0 \sim c \cdot \Lambda_0 \quad \text{for any } c > 0.$$

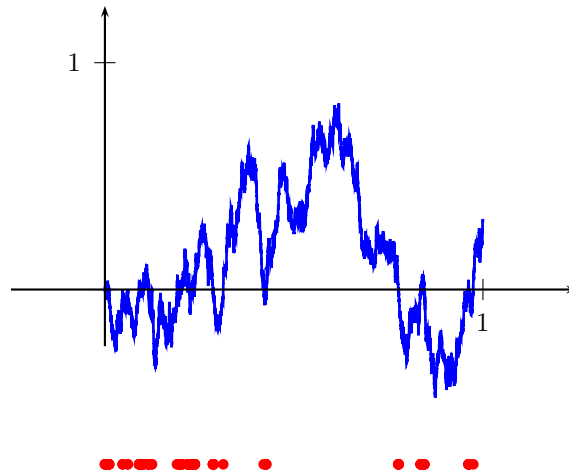


Abbildung 5.1: Brownian motion with corresponding level set Λ_0 .

Moreover, by Fubini's Theorem one easily verifies that Λ_a has almost surely Lebesgue measure zero. In fact, continuity of $t \mapsto B_t(\omega)$ for any ω implies that $(t, \omega) \mapsto B_t(\omega)$ is product measurable (Exercise). Hence $\{(t, \omega) : B_t(\omega) = a\}$ is contained in the product σ -algebra, and

$$E[\lambda(\Lambda_a)] = E\left[\int_0^\infty I_{\{a\}}(B_t) dt\right] = \int_0^\infty P[B_t = a] dt = 0.$$

It is also not difficult to see that the level sets are unbounded:

Satz 5.13 (Unbounded oscillations, recurrence).

$$P\left[\sup_{t \geq 0} B_t = +\infty\right] = P\left[\inf_{t \geq 0} B_t = -\infty\right] = 1.$$

In particular, for any $a \in \mathbb{R}$, the random set Λ_a is almost surely unbounded, i.e. Brownian motion is recurrent.

Beweis. By scale invariance of Brownian motion,

$$\sup_{t \geq 0} B_t \sim c^{-1/2} \sup_{t \geq 0} B_{ct} = c^{-1/2} \sup_{t \geq 0} B_t \quad \text{for any } c > 0.$$

Hence,

$$P\left[\sup_{t \geq 0} B_t \geq a\right] = P\left[\sup_{t \geq 0} B_t \geq a \cdot \sqrt{c}\right]$$

for any $c > 0$, and, therefore, $\sup B_t \in \{0, \infty\}$ almost surely. The first part of the assertion now follows since $\sup B_t$ is almost surely strictly positive. By reflection symmetry, we also obtain $\inf B_t = -\infty$ with probability one. \square

The last theorem makes a statement on the global structure of the set Λ_a . By invariance w.r.t. time inversion this again translates into a local regularity result:

Satz 5.14 (Fine structure of level sets). *The set Λ_a is almost surely a **perfect set**, i.e., any $t \in \Lambda_a$ is an accumulation point of Λ_a .*

Beweis. We prove the statement for $a = 0$, the general case being left as an exercise. We proceed in three steps:

STEP 1: *0 is almost surely an accumulation point of Λ_0 :* This holds by time-reversal. Setting $\widehat{B}_t = t \cdot B_{1/t}$, we see that 0 is an accumulation point of Λ_0 if and only if for any $n \in \mathbb{N}$ there exists $t > n$ such that $\widehat{B}_t = 0$, i.e., if and only if the zero set of \widehat{B}_t is unbounded. By Theorem 5.13, this holds almost surely.

STEP 2: *For any $s \geq 0$, $T_s := \min(\Lambda_0 \cap [s, \infty)) = \min\{t \geq s : B_t = 0\}$ is almost surely an accumulation point of Λ_0 :* For the proof we need the strong Markov property of Brownian motion which will be proved in the next section. By Theorem 5.13, the random variable T_s is almost surely finite. Hence, by continuity, $B_{T_s} = 0$ almost surely. The strong Markov property says that the process

$$\widetilde{B}_t := B_{T_s+t} - B_{T_s} = B_{T_s+t}, \quad t \geq 0,$$

is again a Brownian motion starting at 0. Therefore, with probability one, 0 is an accumulation point of the zero set of \widetilde{B}_t by Step 1. The claim follows since $\widetilde{B}_t = 0$ implies $B_{T_s+t} = 0$.

STEP 3: To complete the proof note that we have shown that the following properties hold with probability one:

- (1). Λ_0 is closed.
- (2). $\min(\Lambda_0 \cap [s, \infty))$ is an accumulation point of Λ_0 for any $s \in \mathbb{Q}_+$.

Since \mathbb{Q}_+ is a dense subset of \mathbb{R}_+ , (1) and (2) imply that any $t \in \Lambda_0$ is an accumulation point of Λ_0 . In fact, for any $s \in [0, t] \cap \mathbb{Q}$, there exists an accumulation point of Λ_0 in $(s, t]$ by (2), and hence t itself is an accumulation point. \square

Remark (Hausdorff dimension). It can be shown that the random set Λ_a has almost surely Hausdorff dimension $1/2$.

5.3 Strong Markov property and reflection principle

In this section we prove a strong Markov property for Brownian motion. Before, we give another motivation for our interest in an extension of the Markov property to random times.

5.3.1 Running maximum of Brownian motion

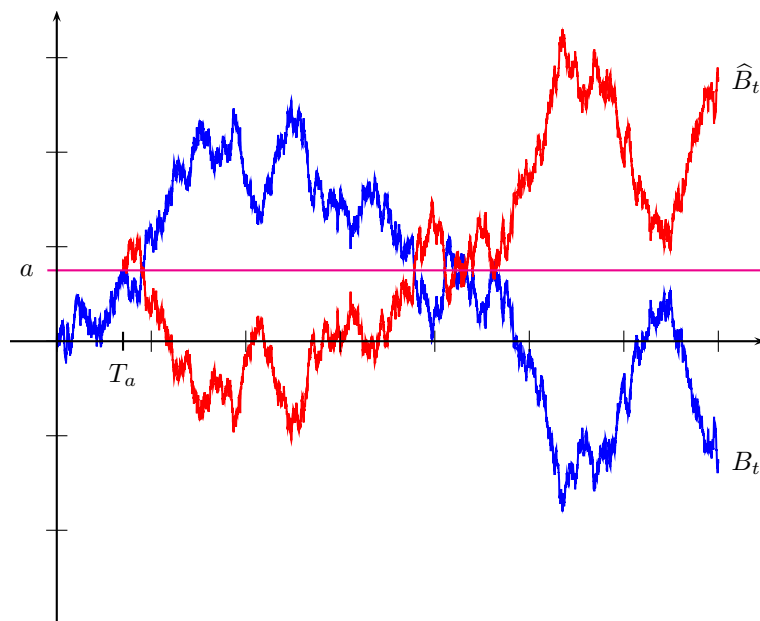
Suppose that $(B_t)_{t \geq 0}$ is a one-dimensional continuous Brownian motion starting at 0 defined on a probability space (Ω, \mathcal{A}, P) . We would like to compute the distribution of the maximal value

$$M_s = \max_{t \in [0, s]} B_t$$

attained before a given time $s \in \mathbb{R}_+$. The idea is to proceed similarly as for random walks, and to reflect the Brownian path after the first passage time

$$T_a = \min\{t \geq 0 : B_t = a\}$$

to a given level $a > 0$:



It seems plausible (e.g. by the heuristic path integral representation of Wiener measure, or by a random walk approximation) that the reflected process $(\widehat{B}_t)_{t \geq 0}$ defined by

$$\widehat{B}_t := \begin{cases} B_t & \text{for } t \leq T_a \\ a - (B_t - a) & \text{for } t > T_a \end{cases}$$

is again a Brownian motion. At the end of this section, we will prove this reflection principle rigorously by the strong Markov property. Assuming the reflection principle is true, we can compute the distribution of M_s in the following way:

$$\begin{aligned} P[M_s \geq a] &= P[M_s \geq a, B_s \leq a] + P[M_s \geq a, B_s > a] \\ &= P[\widehat{B}_s \geq a] + P[B_s > a] \\ &= 2 \cdot P[B_s \geq a] \\ &= P[|B_s| \geq a]. \end{aligned}$$

Thus M_s has the same distribution as $|B_s|$. Furthermore, since $M_s \geq a$ if and only if $\widehat{M}_s = \max\{\widehat{B}_t : t \in [0, s]\} \geq a$, we obtain the stronger statement

$$\begin{aligned} P[M_s \geq a, B_s \leq c] &= P[\widehat{M}_s \geq a, \widehat{B}_s \geq 2a - c] = P[\widehat{B}_s \geq 2a - c] \\ &= \frac{1}{\sqrt{2\pi s}} \int_{2a-c}^{\infty} \exp(-x^2/2s) dx \end{aligned}$$

for any $a \geq 0$ and $c \leq a$. As a consequence, we have:

Satz 5.15 (Distribution of running maxima).

(1). For any $s \geq 0$, the distribution of M_s is absolutely continuous with density

$$f_{M_s}(x) = \frac{2}{\sqrt{2\pi s}} \exp(-x^2/2s) \cdot I_{(0,\infty)}(x).$$

(2). The joint distribution of M_s and B_s is absolutely continuous with density

$$f_{M_s, B_s}(x, y) = 2 \frac{2x - y}{\sqrt{2\pi s^3}} \exp\left(-\frac{(2x - y)^2}{2s}\right) I_{(0,\infty)}(x) I_{(-\infty, x)}(y).$$

Beweis. (1) holds since $M_s \sim |B_s|$. For the proof of (2) we assume w.l.o.g. $s = 1$. The general case can be reduced to this case by the scale invariance of Brownian motion (Exercise). For $a \geq 0$ and $c \leq a$ let

$$G(a, c) := P[M_1 \geq a, B_1 \leq c].$$

By the reflection principle,

$$G(a, c) = P[B_1 \geq 2a - c] = 1 - \Phi(2a - c),$$

where Φ denotes the standard normal distribution function. Since $\lim_{a \rightarrow \infty} G(a, c) = 0$ and $\lim_{c \rightarrow -\infty} G(a, c) = 0$, we obtain

$$\begin{aligned} P[M_1 \geq a, B_1 \leq c] &= G(a, c) = - \int_{x=a}^{\infty} \int_{y=-\infty}^c \frac{\partial^2 G}{\partial x \partial y}(x, y) dy dx \\ &= \int_{x=a}^{\infty} \int_{y=-\infty}^c 2 \cdot \frac{2x-y}{\sqrt{2\pi}} \cdot \exp\left(-\frac{(2x-y)^2}{2}\right) dy dx. \end{aligned}$$

This implies the claim for $s = 1$, since $M_1 \geq 0$ and $B_1 \leq M_1$ by definition of M_1 . \square

The theorem enables us to compute the distributions of the first passage times T_a . In fact, for $a > 0$ and $s \in [0, \infty)$ we obtain

$$\begin{aligned} P[T_a \leq s] &= P[M_s \geq a] = 2 \cdot P[B_s \geq a] = 2 \cdot P[B_1 \geq a/\sqrt{s}] \\ &= \sqrt{\frac{2}{\pi}} \int_{a/\sqrt{s}}^{\infty} e^{-x^2/2} dx. \end{aligned} \tag{5.3.1}$$

Korollar 5.16 (Distribution of first passage times). *For any $a \in \mathbb{R} \setminus \{0\}$, the distribution of T_a is absolutely continuous with density*

$$f_{T_a}(s) = \frac{|a|}{\sqrt{2\pi s^3}} \cdot e^{-a^2/2s}.$$

Beweis. For $a > 0$, we obtain

$$f_{T_a}(s) = F'_{T_a}(s) = \frac{a}{\sqrt{2\pi s^3}} e^{-a^2/2s}$$

by (5.3.1). For $a < 0$ the assertion holds since $T_a \sim T_{-a}$ by reflection symmetry of Brownian motion. \square

We now prove a strong Markov property for Brownian motion. Below we will then complete the proof of the reflection principle and the statements above by applying the strong Markov property to the passage time T_a .

5.3.2 Strong Markov property for Brownian motion

Suppose that $(B_t)_{t \geq 0}$ is a d -dimensional continuous Brownian motion starting at 0 on a probability space (Ω, \mathcal{A}, P) , and let

$$\mathcal{F}_t^B = \sigma(B_s : 0 \leq s \leq t), \quad t \geq 0,$$

denote the filtration generated by the process. It turns out that usually, the first hitting time of an open set in \mathbb{R}^d is *not* a stopping time w.r.t. (\mathcal{F}_t^B) . The reason is that when reaching the boundary, a Brownian path may enter the set or return to the complement. Therefore, given only the information in \mathcal{F}_t^B , it is not possible to decide whether $T \leq t$. Instead, for knowing if $T \leq t$, we need information about the infinitesimal future development of the Brownian paths after time t . This information is encoded in the enlarged σ -algebra

$$\mathcal{F}_t := \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon}^B.$$

The family $(\mathcal{F}_t)_{t \geq 0}$ is again a filtration, and, moreover, it is **right-continuous**, i.e.,

$$\mathcal{F}_t = \bigcap_{\varepsilon > 0} \mathcal{F}_{t+\varepsilon} \quad \text{for any } t \geq 0.$$

Definition (Stopping time). A random variable $T : \Omega \rightarrow [0, \infty]$ is called an (\mathcal{F}_t) **stopping time** if and only if

$$\{T \leq t\} \in \mathcal{F}_t \quad \text{for any } t \geq 0.$$

Example (Hitting times). The proofs of the following facts are left as an exercise:

- (1). In the one-dimensional case, the first passage time

$$T_a = \min\{t \geq 0 : B_t = a\}$$

to a level $a \in \mathbb{R}$ is a stopping time w.r.t. (\mathcal{F}_t^B) , and hence also w.r.t. the larger filtration (\mathcal{F}_t) .

- (2). In general, both hitting times of open and closed sets are (\mathcal{F}_t) stopping times. However, hitting times of open sets are not necessarily (\mathcal{F}_t^B) stopping times.

The σ -algebra \mathcal{F}_T describing the information about the process up to a stopping time T (and an infinitesimal bit beyond) is defined by

$$\mathcal{F}_T = \{A \in \mathcal{A} : A \cap \{T \leq t\} \in \mathcal{F}_t \text{ for any } t \geq 0\}.$$

Note that for (\mathcal{F}_t) stopping times S and T with $S \leq T$ we have $\mathcal{F}_S \subseteq \mathcal{F}_T$, since for $t \geq 0$,

$$A \cap \{S \leq t\} \in \mathcal{F}_t \quad \implies \quad A \cap \{T \leq t\} = A \cap \{S \leq t\} \cap \{T \leq t\} \in \mathcal{F}_t.$$

For any constant $s \in \mathbb{R}_+$, the process $(B_{s+t} - B_s)_{t \geq 0}$ is a Brownian motion independent of \mathcal{F}_s . Indeed, by continuity, we almost surely have

$$B_{s+t} - B_s = \lim_{\varepsilon > 0} (B_{s+t+\varepsilon} - B_{s+\varepsilon}),$$

and the process $B_{s+t+\varepsilon} - B_{s+\varepsilon}$ is a Brownian motion that is independent of $\mathcal{F}_{s+\varepsilon}^B$, and hence of \mathcal{F}_s . A corresponding statement holds for stopping times:

Satz 5.17 (Strong Markov property). *Suppose that T is an almost surely finite (\mathcal{F}_t) stopping time. Then the process $(\tilde{B}_t)_{t \geq 0}$ defined by*

$$\tilde{B}_t = B_{T+t} - B_T \quad \text{if } T < \infty, \quad 0 \quad \text{otherwise,}$$

is a Brownian motion that is independent of \mathcal{F}_T .

Beweis. We first assume that T is an (\mathcal{F}_t) stopping time that takes values only in $C \cup \{\infty\}$ where C is a countable subset of $[0, \infty)$. Then one verifies that for $A \in \mathcal{F}_T$ and $s \in C$, we have $A \cap \{T = s\} \in \mathcal{F}_s$ and $\tilde{B}_t = B_{t+s} - B_s$ on $A \cap \{T = s\}$. As remarked above, the process $(B_{t+s} - B_s)_{t \geq 0}$ is a Brownian motion independent of \mathcal{F}_s . Hence for any measurable subset Γ of $C([0, \infty], \mathbb{R}^d)$, we have

$$\begin{aligned} P[\{(\tilde{B}_t)_{t \geq 0} \in \Gamma\} \cap A] &= \sum_{s \in C} P[\{(B_{t+s} - B_s)_{t \geq 0} \in \Gamma\} \cap A \cap \{T = s\}] \\ &= \sum_{s \in C} \mu_0[\Gamma] \cdot P[A \cap \{T = s\}] = \mu_0[\Gamma] \cdot P[A] \end{aligned}$$

where μ_0 denotes the distribution of Brownian motion starting at 0. This proves the assertion for discrete (\mathcal{F}_t) stopping times.

For an arbitrary (\mathcal{F}_t) stopping time T that is almost surely finite and $n \in \mathbb{N}$, we set

$$T_n = \frac{k}{n} \quad \text{on} \quad \left\{ \frac{k-1}{n} < T \leq \frac{k}{n} \right\} \quad \text{for any } k \in \mathbb{N}.$$

Since the event $\{T_n = k/n\}$ is $\mathcal{F}_{k/n}$ -measurable for any $k \in \mathbb{N}$, T_n is a discrete (\mathcal{F}_t) stopping time. Therefore, $(B_{T_n+t} - B_{T_n})_{t \geq 0}$ is a Brownian motion that is independent of \mathcal{F}_{T_n} , and hence of the smaller σ -algebra \mathcal{F}_T . As $n \rightarrow \infty$, $T_n \rightarrow T$, and thus, by continuity,

$$\tilde{B}_t = B_{T+t} - B_T = \lim_{n \rightarrow \infty} (B_{T_n+t} - B_{T_n}) \quad \text{almost surely.}$$

Now it is easy to verify that $(\tilde{B}_t)_{t \geq 0}$ is again a Brownian motion that is independent of \mathcal{F}_T . \square

The assertion of Theorem 5.17 is even of interest for the trivial stopping time $T \equiv 0$!

Korollar 5.18 (Blumenthal's 0 – 1 law). *The events in \mathcal{F}_0 have either probability zero or one.*

Beweis. By Theorem 5.17, the random variables $B_t = B_t - B_0$ are independent of \mathcal{F}_0 for any $t \geq 0$. On the other hand, $\mathcal{F}_0 \subseteq \sigma(B_t : t \geq 0)$, and hence any event in \mathcal{F}_0 is independent of itself. \square

Blumenthal's 0 – 1 law says that events that only depend on the asymptotics of Brownian motion B_t as $t \downarrow 0$ have probability zero or one. Note that these events correspond to events depending on the asymptotics as $t \uparrow \infty$ of the time-inverted Brownian motion $\hat{B}_t = tB_{1/t}$.

5.3.3 A rigorous reflection principle

We now apply the strong Markov property to prove a reflection principle for Brownian motion. Consider a one-dimensional continuous Brownian motion $(B_t)_{t \geq 0}$ starting at 0. For $a \in \mathbb{R}$ let

$$\begin{aligned} T_a &= \min\{t \geq 0 : B_t = a\} && \text{(first passage time),} \\ B_t^{T_a} &= B_{\min(t, T_a)} && \text{(process stopped at } T_a), \quad \text{and} \\ \tilde{B}_t &= B_{T_a+t} - B_{T_a} && \text{(process after } T_a). \end{aligned}$$

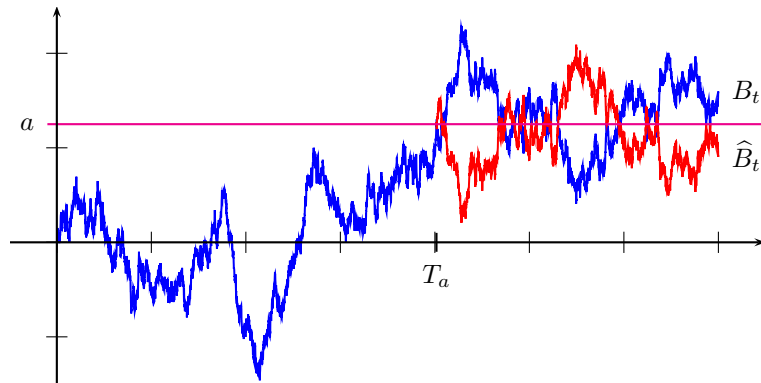
Satz 5.19 (Reflection principle). *The joint distributions of the following random variables with values in $\mathbb{R}_+ \times C([0, \infty)) \times C([0, \infty))$ agree:*

$$(T_a, (B_t^{T_a})_{t \geq 0}, (\tilde{B}_t)_{t \geq 0}) \sim (T_a, (B_t^{T_a})_{t \geq 0}, (-\tilde{B}_t)_{t \geq 0})$$

Beweis. By the strong Markov property, the process \tilde{B} is a Brownian motion starting at 0 that is independent of \mathcal{F}_{T_a} , and hence of T_a and $B^{T_a} = (B_t^{T_a})_{t \geq 0}$. Therefore,

$$P \circ (T_a, B^{T_a}, \tilde{B})^{-1} = P \circ (T_a, B^{T_a})^{-1} \otimes \mu_0 = P \circ (T_a, B^{T_a}, -\tilde{B})^{-1}.$$

□



As a consequence of the theorem, we can complete the argument given at the beginning of this section: The “shadow path” \hat{B}_t of a Brownian path B_t with reflection when reaching the level a is given by

$$\hat{B}_t = \begin{cases} B_t^{T_a} & \text{for } t \leq T_a \\ a - \tilde{B}_{t-T_a} & \text{for } t > T_a \end{cases},$$

whereas

$$B_t = \begin{cases} B_t^{T_a} & \text{for } t \leq T_a \\ a + \tilde{B}_{t-T_a} & \text{for } t > T_a \end{cases}.$$

By Theorem 5.19, $(\widehat{B}_t)_{t \geq 0}$ has the same distribution as $(B_t)_{t \geq 0}$. Therefore, and since $\max_{t \in [0, s]} B_t \geq a$ if and only if $\max_{t \in [0, s]} \widehat{B}_t \geq a$, we obtain for $a \geq c$:

$$\begin{aligned} P \left[\max_{t \in [0, s]} B_t \geq a, B_s \leq c \right] &= P \left[\max_{t \in [0, s]} \widehat{B}_t \geq a, \widehat{B}_s \geq 2a - c \right] \\ &= P \left[\widehat{B}_s \geq 2a - c \right] \\ &= \frac{1}{\sqrt{2\pi s}} \int_{2a-c}^{\infty} e^{-x^2/2s} dx. \end{aligned}$$

Kapitel 6

Langzeitasymptotik von Markovketten

In diesem Abschnitt betrachten wir Rekurrenz, Transienz und Ergodizität von stochastischen Prozessen. Wir beschränken meist auf zeithomogene Markovketten mit abzählbarem Zustandsraum. Einige grundlegende Aussagen beweisen wir auch für allgemeine stationäre Prozesse.

6.1 Rekurrenz und Transienz von Markovketten

Sei $p(x, y)$ ($x, y \in S$) eine stochastische Matrix auf einer abzählbaren Menge S . Wir betrachten eine zeithomogene Markovkette (X_n, P_x) mit Übergangsmatrix p im kanonischen Modell, d.h.

$$\Omega = S^{\{0,1,2,\dots\}}, \quad X_n(\omega) = \omega_n, \quad \mathcal{A} = \sigma(X_n : n \geq 0),$$

und P_x ist die Verteilung der Markovkette bei Start in x . Für $y \in S$ sei

$$B_y(\omega) = \sum_{n=0}^{\infty} I_{\{y\}}(X_n(\omega))$$

die Anzahl der Besuche (Aufenthaltszeit) der Markovkette im Punkt y . Wir wollen untersuchen, ob die Markovkette immer wieder zu ihrem Startpunkt zurückkehrt.

Definition (Rekurrenz und Transienz). Ein Punkt $x \in S$ heißt **transient**, falls $P_x[B_x = \infty] = 0$ gilt, und **rekurrent**, falls $P_x[B_x = \infty] = 1$. Die Markovkette (X_n, P_x) heißt **transient** bzw. **rekurrent**, falls alle Punkte $x \in S$ transient bzw. rekurrent sind.

Sei nun

$$G(x, y) = E_x[B_y] = \sum_{n=0}^{\infty} p^n(x, y)$$

die mittlere Anzahl der Besuche der Markovkette im Punkt y bei Start in x . Offensichtlich ist x transient, wenn

$$G(x, x) = E_x[B_x] < \infty$$

gilt. Wir werden in Korollar 6.2 zeigen, dass umgekehrt x rekurrent ist, wenn $G(x, x) = \infty$ gilt. Insbesondere ergibt sich ein 0-1-Gesetz: Jeder Punkt ist entweder transient oder rekurrent. Allgemeiner werden wir sehen, dass bei irreduziblen Markovketten sogar entweder alle Punkte transient oder alle Punkte rekurrent sind.

Intuitiv können wir diese Dichotomie folgendermaßen erklären: Jedes Mal, wenn die Markovkette zum Startpunkt x zurückkehrt, startet sie aufgrund der Markoveigenschaft wieder neu in diesem Punkt – unabhängig vom vorherigen Verlauf. Kehrt die Kette also mit Wahrscheinlichkeit 1 wieder zum Startpunkt zurück, dann kehrt sie auch mit Wahrscheinlichkeit 1 immer wieder, also unendlich oft nach x zurück. Ist die Markovkette zudem irreduzibel, dann erreicht sie jeden festen Punkt y auf jeder Exkursion mit einer konstanten strikt positiven Wahrscheinlichkeit – trifft also insgesamt den Punkt y mit Wahrscheinlichkeit 1 unendlich oft.

Kehrt die Kette dagegen mit einer strikt positiven Wahrscheinlichkeit $\varepsilon > 0$ nicht zum Startpunkt x zurück, dann wird sie auch bei jedem weiteren Erreichen von x mit derselben Wahrscheinlichkeit ε nicht wieder zurückkehren – unabhängig vom vorherigen Verlauf. Also wird sie mit Wahrscheinlichkeit 1 schließlich nicht mehr nach x zurückkehren – sie durchläuft also jeden Punkt nur endlich oft.

Um diese *Dichotomie von Rekurrenz und Transienz* rigoros zu beweisen, werden wir die starke Markov-Eigenschaft benutzen. Zuvor betrachten wir schon mal eine Anwendung auf mehrdimensionale Random Walks.

6.1.1 Rekurrenz und Transienz von d -dimensionalen Random Walks

Sei (X_n, P_x) der klassische Random Walk auf \mathbb{Z}^d mit Übergangswahrscheinlichkeiten $p(x, y) = \frac{1}{2d}$ falls $|x - y| = 1$, $p(x, y) = 0$ sonst. Wir untersuchen Rekurrenz und Transienz in Abhängigkeit von der Dimension d :

$d = 1$: Im eindimensionalen Fall erhalten wir für die Rückkehrwahrscheinlichkeiten zum Ausgangspunkt x mithilfe der Stirling-Approximation:

$$\begin{aligned} p^{2n}(x, x) &= \binom{2n}{n} \cdot 2^{-2n} = \frac{(2n)!}{(n!)^2} 2^{-2n} \\ &\sim \frac{\sqrt{4\pi n}}{2\pi n} \frac{(2n)^{2n}}{n^{2n}} \cdot 2^{-2n} = \frac{1}{\sqrt{\pi n}}. \end{aligned}$$

Also gilt $G(x, x) = \sum_{n=0}^{\infty} p^n(x, x) = \infty$, d.h. jeder Punkt $x \in \mathbb{Z}$ ist *rekurrent*.

$d = 2$: Beim klassischen Random Walk $X_n = (X_n^{(1)}, X_n^{(2)})$ auf \mathbb{Z}^2 sind die Komponenten $X_n^{(1)}$ und $X_n^{(2)}$ nicht unabhängig.

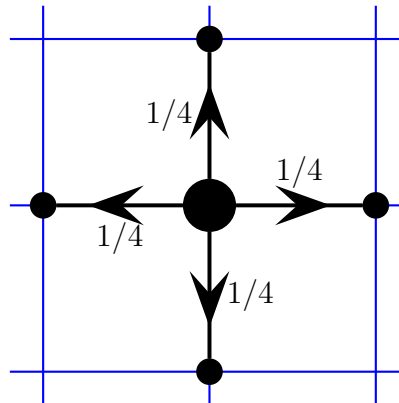


Abbildung 6.1: Klassischer Random Walk.

Durch eine 45° Drehung des Koordinatensystems, können wir den Prozess aber in einen zweidimensionalen Random Walk

$$Y_n = (X_n^{(1)} + X_n^{(2)}, X_n^{(1)} - X_n^{(2)})$$

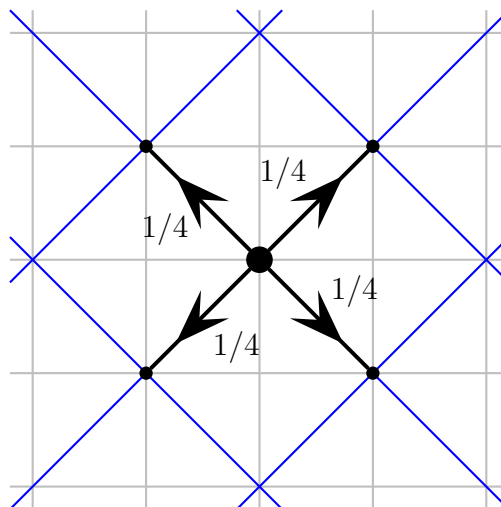


Abbildung 6.2: Um 45° gedrehter Random Walk.

überführen, dessen Komponenten $Y_n^{(1)}$ und $Y_n^{(2)}$ unabhängige eindimensionale Random Walks sind. Offensichtlich ist der Prozess X_n genau dann rekurrent, wenn Y_n rekurrent ist. Die Rückkehrwahrscheinlichkeiten für Y_n sind

$$\begin{aligned} p^{2n}(x, x) &= P_x[Y_{2n}^{(1)} = x_1, Y_{2n}^{(2)} = x_2] = P_{x_1}[Y_{2n}^{(1)} = x_1] \cdot P_{x_2}[Y_{2n}^{(2)} = x_2] \\ &= \left(\binom{2n}{n} \cdot 2^{-2n} \right)^2 \sim \frac{1}{\pi n}. \end{aligned}$$

Also gilt erneut $G(x, x) = \infty$, d.h. jedes $x \in \mathbb{Z}^2$ ist *rekurrent*.

$d = 3$: Betrachten wir einen dreidimensionalen Random Walk

$$Y_n = (X_n^{(1)}, X_n^{(2)}, X_n^{(3)}),$$

dessen Komponenten $X_n^{(i)}$ unabhängige klassische Random Walks auf \mathbb{Z}^1 sind, dann gilt entsprechend

$$p^{2n}(x, x) = \left(\binom{2n}{n} \cdot 2^{-2n} \right)^3 \sim \frac{1}{(\pi n)^{3/2}},$$

und damit

$$G(x, x) = \sum_{n=0}^{\infty} p^{2n}(x, x) < \infty$$

Der Prozess ist also *transient*. Auch der klassische Random Walk auf \mathbb{Z}^3 ist transient – der Beweis erfordert allerdings etwas mehr Kombinatorik, da sich der Prozess in Dimension 3 nicht durch eine Drehung in einen Prozess mit unabhängigen Komponenten überführen lässt. Die Details werden in einer Übungsaufgabe ausgeführt. Analog folgt Transienz in höheren Dimensionen. Zwischen Dimension 2 und 3 gibt es also einen Übergang von rekurrentem zu transientem Verhalten. Anschaulich steht in Dimension $d > 2$ soviel Raum zur Verfügung, dass der Random Walk der Startpunkt schließlich nicht mehr trifft.

6.1.2 Rekurrenz und Transienz von einzelnen Zuständen

Mithilfe der starken Markoveigenschaft können wir die Verteilung der Aufenthaltszeit B_y einer Markovkette in einem Punkt $y \in S$ aus den Treffer- bzw. Rückkehrwahrscheinlichkeiten

$$f(x, y) := P_x[T_y < \infty]$$

berechnen. Hierbei bezeichnen wir mit

$$T_y = \min\{n \geq 1 : X_n = y\}$$

die erste Passierzeit des Zustandes y . Man beachte, dass diese im Gegensatz zur ersten Trefferzeit mit der ersten Rückkehrzeit nach y übereinstimmt, falls die Markovkette in y startet.

Satz 6.1 (Anzahl der Besuche in einem Zustand). Für alle $x, y \in S$ und $n \in \mathbb{N}$ gilt

$$P_x[B_y \geq n] = \begin{cases} f(x, y) \cdot f(y, y)^{n-1} & \text{falls } y \neq x \\ f(y, y)^{n-1} & \text{falls } y = x \end{cases}.$$

Insbesondere ist jedes $y \in S$ entweder rekurrent oder transient, und es gilt:

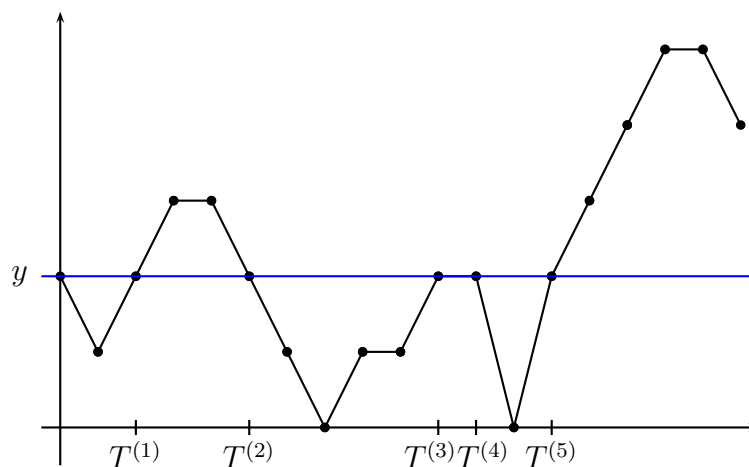
$$y \text{ rekurrent} \iff f(y, y) = 1,$$

$$y \text{ transient} \iff f(y, y) < 1.$$

Beweis. Sei $T^{(0)} := 0$, und sei

$$T^{(n)} := T^{(n-1)} + T_y \circ \theta^{T^{(n-1)}}$$

die n -te Besuchszeit (bei Start außerhalb von y) bzw. Rückkehrzeit (bei Start in y) des Zustands y . Hierbei ist $\theta(\omega_0, \omega_1, \dots) = (\omega_1, \omega_2, \dots)$ der Shift-Operator auf dem Folgenraum Ω .



Mithilfe der starken Markoveigenschaft erhalten wir für $n \in \mathbb{N}$:

$$\begin{aligned} P_x[T^{(n)} < \infty] &= P_x[T^{(n-1)} < \infty \text{ und } T_y \circ \theta^{T^{(n-1)}} < \infty] \\ &= E_x[P_x[T_y \circ \theta^{T^{(n-1)}} < \infty \mid \mathcal{F}_{T^{(n-1)}}]; T^{(n-1)} < \infty] \\ &= E_x[P_{X_{T^{(n-1)}}}[T_y < \infty]; T^{(n-1)} < \infty] \\ &= P_y[T_y < \infty] \cdot P_x[T^{(n-1)} < \infty]. \end{aligned}$$

Durch Induktion nach n ergibt sich:

$$P_x[T^{(n)} < \infty] = P_x[T_y < \infty] \cdot P_y[T_y < \infty]^{n-1} = f(x, y) \cdot f(y, y)^{n-1}.$$

Die Aussage folgt wegen

$$P_x[B_y \geq n] = \begin{cases} P_x[T^{(n)} < \infty] & \text{falls } x \neq y, \\ P_x[T^{(n-1)} < \infty] & \text{falls } x = y. \end{cases}$$

□

Bemerkung (Dirichletproblem für Trefferwahrscheinlichkeiten). Die Wahrscheinlichkeiten $f(x, y)$ (und damit die Verteilungen der Aufenthaltszeiten) kann man im Prinzip durch Lösen eines Dirichletproblems berechnen: Nach Korollar 2.13 bilden die Trefferwahrscheinlichkeiten

$$h(x) = P_x[\tilde{T}_y < \infty], \quad \tilde{T}_y := \min\{n \geq 0 : X_n = y\},$$

die minimale nichtnegative Lösung von

$$ph = h \quad \text{auf } S \setminus \{y\}; \quad h(y) = 1.$$

Wegen $\tilde{T}_y = T_y$ P_x -fast sicher für alle $y \neq x$ folgt $f(x, y) = h(x)$ für $x \neq y$, und

$$f(y, y) = \sum_{x \in S} p(y, x) f(x, y) = (ph)(y).$$

Beispiel (Kartenhaus, Maschinenerneuerung). Wir betrachten eine Markovkette mit Zustandsraum $S = \{0, 1, 2, \dots\}$ und Übergangswahrscheinlichkeiten

$$p(x, x+1) = 1 - \varrho(x), \quad p(x, 0) = \varrho(x), \quad \varrho(x) \in (0, 1).$$

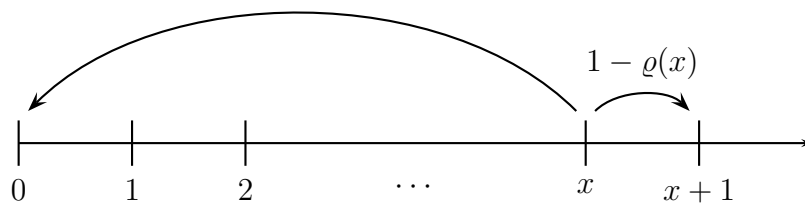


Abbildung 6.3: Kartenhaus-Prozess.

Hier gilt

$$P_0[T_0 > n] = \prod_{x=0}^{n-1} (1 - \varrho(x)),$$

also:

$$0 \text{ rekurrent} \iff P_0[T_0 = \infty] = \prod_{x=0}^{\infty} (1 - \varrho(x)) = 0 \iff \sum_{x=0}^{\infty} \varrho(x) = \infty$$

Aus Satz 6.1 folgt unmittelbar die schon oben erwähnte Charakterisierung rekurrenter Zustände über die Greensche Funktion:

Korollar 6.2 (Rekurrenz Kriterium). Für alle $x \in S$ gilt

$$G(x, x) = \frac{1}{1 - f(x, x)} = \frac{1}{P_x[T_x = \infty]} \quad \text{falls } P_x[T_x = \infty] > 0,$$

bzw. $G(x, x) = \infty$ falls $P_x[T_x = \infty] = 0$. Insbesondere ist x genau dann rekurrent, wenn $G(x, x)$ unendlich ist.

Beweis. Für $x \in S$ gilt nach Satz 6.1:

$$G(x, x) = E_x[B_x] = \sum_{n=1}^{\infty} P_x[B_x \geq n] = \sum_{n=1}^{\infty} f(x, x)^{n-1} = \sum_{n=0}^{\infty} f(x, x)^n.$$

□

Leider ist das Kriterium zwar für die Theorie wichtig, aber praktisch nur selten einsetzbar. Leichtere verifizierbare hinreichende Bedingungen für Rekurrenz und Transienz basieren auf stochastischen Lyapunovfunktionen und dem Martingalkonvergenzsatz, siehe zum Beispiel [Markov Processes].

6.1.3 Kommunikationsklassen und globale Rekurrenz

Wir wollen nun untersuchen, wie die Rekurrenz verschiedener Zustände $x, y \in S$ miteinander zusammenhängt.

Definition (Erreichbarkeit von Zuständen). Der Zustand y heißt **erreichbar** von x für die Markovkette (X_n, P_x) , falls $P_x[T_y < \infty] > 0$ gilt.

Bemerkung. (1). Ein Zustand y ist genau dann von x erreichbar, wenn es ein $n \in \mathbb{N}$ gibt mit $p^n(x, y) > 0$. Insbesondere gilt für $y \neq x$:

$$y \text{ ist erreichbar von } x \iff G(x, y) > 0.$$

(2). Ist y erreichbar von x und z erreichbar von y , dann ist z erreichbar von x .

(3). Die Übergangsmatrix ist genau dann irreduzibel, wenn jeder Zustand von jedem anderen Zustand aus erreichbar ist.

Wir wollen zeigen, dass mit einem Zustand $x \in S$ auch jeder von x aus erreichbare Zustand y rekurrent ist. Dazu bemerken wir zunächst, dass wir die Werte $G(x, y)$ der Greenschen Matrix außerhalb der Diagonalen durch die Werte auf der Diagonalen abschätzen können:

Lemma 6.3. Für $x, y \in S$ mit $y \neq x$ gilt

$$G(x, y) = P_x[T_y < \infty] \cdot G(y, y).$$

Beweis. Für $y \neq x$ gilt P_x -fast sicher $X_0 \neq y$, also

$$B_y = B_y \circ \theta^{T_y} \quad \text{auf } \{T_y < \infty\}.$$

Mit der starken Markoveigenschaft folgt

$$\begin{aligned} E_x[B_y] &= E_x[B_y ; T_y < \infty] = E_x[B_y \circ \theta^{T_y} ; T_y < \infty] \\ &= E_y[B_y] \cdot P_x[T_y < \infty]. \end{aligned}$$

□

Satz 6.4 (Globale Rekurrenz). Ist x rekurrent, und y von x aus erreichbar, dann ist auch x von y aus erreichbar, y ist rekurrent, und es gilt

$$B_y = \infty \quad P_x\text{-fast sicher} \quad \text{und} \quad B_x = \infty \quad P_y\text{-fast sicher.}$$

Insbesondere gilt also

$$G(x, y) = G(y, x) = G(y, y) = G(x, x) = \infty.$$

Beweis. (1). y ist rekurrent: Da y von x aus erreichbar ist, existiert $m \geq 0$ mit $p^m(x, y) > 0$.

Nach Lemma 6.3 folgt:

$$\begin{aligned} G(y, y) &\geq G(x, y) \geq \sum_{n=0}^{\infty} p^{n+m}(x, y) \\ &\geq \sum_{n=0}^{\infty} p^n(x, x) p^m(x, y) \\ &= \underbrace{G(x, x)}_{=\infty} \cdot \underbrace{p^m(x, y)}_{>0} = \infty. \end{aligned}$$

(2). Wir zeigen $P_y[B_x = \infty] = 1$: Da y von x aus erreichbar und x rekurrent ist, gilt nach der starken Markoveigenschaft

$$\begin{aligned} 0 < P_x[T_y < \infty] &\stackrel{x \text{ rek.}}{=} P_x[T_y < \infty, T_x \circ \theta^{T_y} < \infty] \\ &\stackrel{\text{SME}}{=} P_x[T_y < \infty] \cdot P_y[T_x < \infty]. \end{aligned}$$

Hierau folgt $f(y, x) = P_y[T_x < \infty] = 1$. Da x rekurrent ist, gilt zudem $f(x, x) = 1$, also nach Satz 6.1

$$P_y[B_x = \infty] = \lim_{n \rightarrow \infty} P_y[B_x \geq n] = \lim_{n \rightarrow \infty} (f(y, x) \cdot f(x, x)^{n-1}) = 1.$$

Insbesondere ist x von y aus erreichbar.

(3). Analog erhalten wir $P_x[B_y = \infty] = 1$ durch Vertauschen der Rolle von x und y .

□

Der Satz zeigt, dass für eine Markovkette mit irreduzibler Übergangsmatrix und einem rekurrenten Zustand alle Zustände rekurrent sind, und jeder Zustand bei beliebiger Startverteilung mit Wahrscheinlichkeit 1 unendlich oft durchlaufen wird:

Korollar 6.5 (Dichotomie von Rekurrenz und Transienz). *Für eine zeithomogene Markovkette mit irreduzibler Übergangsmatrix gilt entweder*

(1). *Alle $x \in S$ sind rekurrent, und $P_x[B_y = \infty] = 1$ für alle $x, y \in S$, oder*

(2). *Alle $x \in S$ sind transient, und $E_x[B_y] < \infty$ für alle $x, y \in S$.*

Ist S endlich, dann kann nur der erste Fall eintreten.

Beweis. Existiert ein rekurrenter Zustand, dann sind nach Satz 6.4 alle Zustände rekurrent, und $P_x[B_y = \infty] = 1$ für alle $x, y \in S$. Andernfalls sind nach Satz 6.1 alle $x \in S$ transient, und nach Korollar 6.2 gilt $G(x, x) < \infty$. Nach Lemma 6.3 folgt dann $E_x[B_y] < \infty$ für alle $x, y \in S$. Ist S endlich, dann kann der zweite Fall wegen

$$\sum_{y \in S} E_x[B_y] = E_x \left[\sum_{y \in S} B_y \right] = \infty$$

nicht eintreten.

□

Was können wir aussagen, wenn die Übergangsmatrix nicht irreduzibel ist?

Allgemein ist die Relation

$$x \rightsquigarrow y \quad \text{„}y \text{ ist von } x \text{ aus erreichbar“}$$

eine Äquivalenzrelation auf der Menge S_{rek} der rekurrenten Zustände in S . Die zugehörigen Äquivalenzklassen $S_i, i \in I$, heißen **Rekurrenzklassen**. Wir erhalten also eine disjunkte Zerlegung

$$S = S_{\text{trans}} \dot{\cup} \bigcup_{i \in I} S_i$$

des Zustandsraums in die Menge S_{trans} der transienten Zustände, und die verschiedenen Rekurrenzklassen.

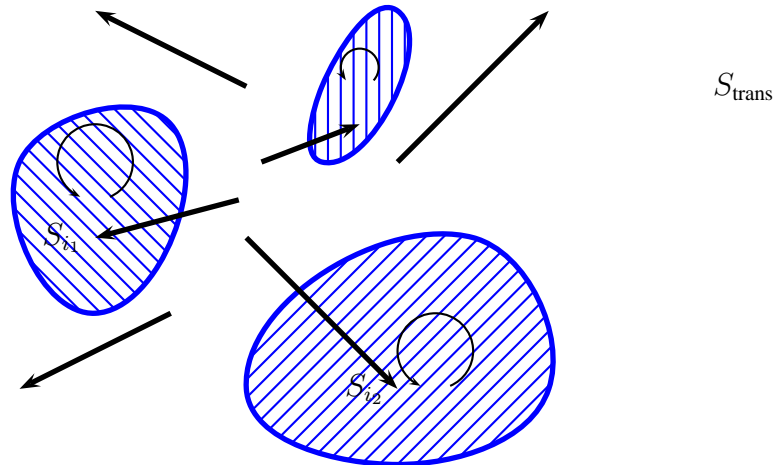


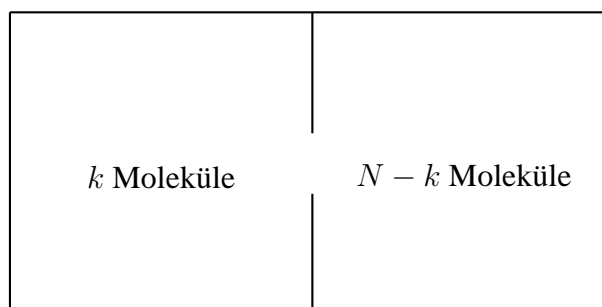
Abbildung 6.4: Zerlegung der Menge S in transienten Zustände und Rekurrenzklassen

Gelangt die Markovkette in eine Rekurrenzklass, dann bleibt sie dort mit Wahrscheinlichkeit 1, und durchläuft alle Zustände der Rekurrenzklass unendlich oft. Startet die Markovkette in einem transienten Zustand, dann läuft sie entweder in eine Rekurrenzklass, oder sie verbleibt im transienten Bereich, verlässt aber jede endliche Teilmenge von S_{trans} schließlich mit Wahrscheinlichkeit 1.

Beispiel. (1). *Ehrenfestmodell:* Die Markovkette aus dem Ehrenfestmodell ist rekurrent, da der Zustandsraum $S = \{0, 1, \dots, N\}$ endlich, und die Übergangsmatrix

$$\begin{aligned} p(k, k-1) &= k/N \\ p(k, k+1) &= (N-k)/N \end{aligned}$$

irreduzibel ist.



Jeder Zustand wird also unendlich oft durchlaufen, was der thermodynamischen Irreversibilität zunächst zu widersprechen scheint (Einwand von Zermelo, vgl. die Bemerkung unter Satz ??). Tatsächlich kann man zeigen, dass die mittlere Zeit $E_0[T_{N/2}]$ für den Übergang vom geordneten Zustand $k = 0$ in den ungeordneten Zustand $k = N/2$ von der Größenordnung $N \log N$ ist, die mittlere Zeit $E_{N/2}[T_0]$ für den umgekehrten Übergang dagegen von der Größenordnung $\frac{1}{2^N} 2^{2N}$. Da N zum Beispiel gleich 10^{23} ist, ist die Rekurrenz jenseits des ungeordneten Zustandes de facto nicht beobachtbar – im makroskopischen Skalierungslimes $N \rightarrow \infty$ ergibt sich bei geeigneter Zeitreskalierung eine irreversible Dynamik.

(2). *Kartenhaus/Maschinenerneuerung*: Im Fall $\sum_{x=0}^{\infty} \varrho_x = \infty$ sind alle Zustände der Markovkette aus dem Beispiel von oben rekurrent, da 0 rekurrent und die Übergangsmatrix irreduzibel ist. Andernfalls sind alle Zustände transient.

(3). *Galton-Watson-Prozess*: Für den Galton-Watson-Verzweigungsprozess mit Nachkommensverteilung ν ist 0 ein *absorbierender* Zustand, d.h. kein anderer Zustand ist von 0 aus erreichbar. Insbesondere ist $\{0\}$ eine Rekurrenzklasse. Gilt $\nu(0) \neq 0$, dann ist umgekehrt 0 von jedem Zustand $x \in \mathbb{N}$ aus erreichbar, also sind alle $x \neq 0$ transient. Es folgt dann:

$$P_x[Z_n = 0 \text{ schließlich oder } Z_n \rightarrow \infty] = 1 \quad \text{für alle } x \in \mathbb{Z}_+.$$

6.2 Stationäre stochastische Prozesse

In vielen Fällen nähert sich die Verteilung eines zeitlich verschobenen stochastischen Prozesses (Y_n, Y_{n+1}, \dots) mit Zustandsraum (S, \mathcal{S}) für $n \rightarrow \infty$ einer Grenzverteilung P auf dem Produktraum $\Omega = S^{\{0,1,2,\dots\}}$ mit Produkt- σ -Algebra \mathcal{A} an („asymptotische Stationarität“). Die Grenzverteilung P sollte dann selbst invariant unter Verschiebungen sein, d.h. für den Koordinatenprozess $X_n(\omega) = \omega_n$ sollte gelten:

$$(X_n, X_{n+1}, \dots) \sim (X_0, X_1, \dots) \quad \text{unter } P \text{ für alle } n \geq 0. \quad (6.2.1)$$

Wir wollen stochastische Prozesse mit der Eigenschaft (6.2.1) nun genauer untersuchen. Dabei setzen wir zunächst keine Markoveigenschaft voraus !

6.2.1 Stationarität und Reversibilität

Definition (Stationäre und reversible Prozesse). (1). Eine Wahrscheinlichkeitsverteilung P auf dem Folgenraum (Ω, \mathcal{A}) bzw. ein beliebiger stochastischer Prozess $((X_n), P)$ heißt **stationär**, falls (6.2.1) gilt.

(2). Der Prozess $((X_n), P)$ heißt **reversibel**, falls die endlichdimensionalen Verteilungen invariant unter Zeitumkehr sind, d.h. falls

$$(X_0, X_1, \dots, X_n) \sim (X_n, X_{n-1}, \dots, X_0) \quad \text{unter } P \text{ für alle } n \geq 0. \quad (6.2.2)$$

Bemerkung (Stationarität und maßerhaltende Abbildungen). Eine Wahrscheinlichkeitsverteilung P auf (Ω, \mathcal{A}) ist genau dann stationär, wenn die Shiftabbildung $\theta : \Omega \rightarrow \Omega$ maßerhaltend ist, d.h. wenn $P = P \circ \theta^{-1}$ gilt.

Beispiele. (1). *IID Folgen:* Eine Folge $(X_n)_{n \geq 0}$ unabhängiger, identisch verteilter Zufallsvariablen ist ein stationärer und reversibler stochastischer Prozess.

(2). *Gaußprozesse:* Ein Gaußprozess ist ein reellwertiger stochastischer Prozess (X_n) , dessen Randverteilungen Normalverteilungen sind. Beispielsweise ist ein $AR(p)$ -Prozess ein Gaußprozess, wenn die Startwerte normalverteilt sind. Gaußprozesse sind eindeutig durch die Mittelwerte $E[X_n]$ und die Kovarianzen $\text{Cov}[X_n, X_m]$ festgelegt. Stationarität gilt genau dann, wenn $E[X_n] = \alpha$ nicht von n abhängt, und $\text{Cov}[X_n, X_m] = c_{n-m}$ nur von der Differenz $n - m$ abhängt.

(3). *Deterministische Rotationen:* Ist X_0 gleichverteilt auf dem Einheitskreis S^1 , und $X_{n+1} = e^{i\phi} X_n$ mit $\phi \in \mathbb{R}$, dann ist $(X_n)_{n \geq 0}$ stets ein stationärer Prozess. Reversibilität gilt dagegen nur für $\phi = k\pi$ mit $k \in \mathbb{Z}$.

Satz 6.6 (Reversibilität impliziert Stationarität). *Jeder reversible Prozess ist stationär.*

Beweis. Aus der Reversibilität folgt durch Zeitumkehr auf $\{0, 1, \dots, n+1\}$ und $\{0, 1, \dots, n\}$:

$$P \circ (X_1, X_2, \dots, X_{n+1})^{-1} = P \circ (X_n, X_{n-1}, \dots, X_0)^{-1} = P \circ (X_0, \dots, X_n)^{-1}$$

für alle $n \geq 0$. Also gilt

$$P[(X_1, X_2, \dots) \in A] = P[(X_0, X_1, \dots) \in A]$$

für alle Zylindermengen $A \in \mathcal{A}$, und damit für alle $A \in \mathcal{A}$. □

Stationarität bzw. Reversibilität zeithomogener Markovketten ist durch die Startverteilung und den Übergangskern charakterisierbar:

Satz 6.7 (Stationarität und Reversibilität von Markovketten). *Für eine zeithomogene Markovkette (X_n, P_μ) im kanonischen Modell gilt:*

- (1). $P_\mu \circ (X_n, X_{n+1}, \dots)^{-1} = P_{\mu p^n}$ für alle $n \geq 0$.
- (2). P_μ ist genau dann stationär, wenn μ ein Gleichgewicht des Übergangskerns p ist.
- (3). P_μ ist genau dann reversibel, wenn μ die Detailed-Balance-Bedingung

$$\mu(dx)p(x, dy) = \mu(dy)p(y, dx) \quad (6.2.3)$$

erfüllt, d.h. wenn die Wahrscheinlichkeitsverteilung $\mu \otimes p$ auf $S \times S$ invariant unter der Abbildung $(x, y) \mapsto (y, x)$ ist.

Beweis. (1). Für $A \in \mathcal{A}$ und $n \geq 0$ gilt nach der Markoveigenschaft

$$\begin{aligned} P_\mu[(X_n, X_{n+1}, \dots) \in A] &= E_\mu[I_A \circ \theta^n] = E_\mu[P_{X_n}[A]] \\ &= \int P_x[A] (\mu p^n)(dx) = P_{\mu p^n}[A]. \end{aligned}$$

- (2). folgt unmittelbar aus (1).
- (3). Aus der Reversibilität von (X_n, P_μ) folgt, dass

$$\mu \otimes p = P_\mu \circ (X_0, X_1)^{-1}$$

invariant unter Koordinatentausch ist.

Umgekehrt folgt aus der Detailed-Balance-Bedingung durch Induktion

$$\begin{aligned} \mu(dx_0)p(x_0, dx_1) \cdots p(x_{n-1}, dx_n) &= \mu(dx_1)p(x_1, dx_2) \cdots p(x_{n-1}, dx_n)(x_1, dx_0) \\ &= \dots = \mu(dx_n)p(x_n, dx_{n-1}) \cdots p(x_1, dx_0) \end{aligned}$$

für alle $n \geq 0$, also $P_\mu \circ (X_0, \dots, X_n)^{-1} = P_\mu \circ (X_n, \dots, X_0)^{-1}$.

□

6.2.2 Rekurrenz von stationären Prozessen

Stationäre stochastische Prozesse haben starke Rekurrenzeigenschaften. Die folgende Aussage zeigt unter Anderem, dass die mittlere Rückkehrzeit in eine Menge B endlichen Erwartungswert hat, wenn der Prozess mit positiver Wahrscheinlichkeit in B startet:

Satz 6.8 (Wiederkehrsatz von Kac). Sei (X_n, P) ein stationärer stochastischer Prozess mit Zustandsraum (S, \mathcal{S}) , und sei

$$T_B = \min\{n \geq 1 : X_n \in B\}$$

die erste Eintritts- bzw. Rückkehrzeit in eine Menge $B \in \mathcal{S}$. Dann gilt

$$E[T_B; X_0 \in B] = P[T_B < \infty], \quad (6.2.4)$$

also mit anderen Worten

$$E[T_B | X_0 \in B] = \frac{P[T_B < \infty]}{\mu[B]} \quad \text{falls } \mu[B] > 0, \quad \text{und} \quad (6.2.5)$$

$$P[T_B < \infty] = 0 \quad \text{falls } \mu[B] = 0, \quad (6.2.6)$$

wobei $\mu = P \circ X_0^{-1}$ die Verteilung des Prozesses zu einem festen Zeitpunkt ist.

Bemerkung. (1). Nach (6.2.5) ist die mittlere Rückkehrzeit in die Menge B der Kehrwert des Quotienten $\frac{P[X_0 \in B]}{P[T_B < \infty]}$, also des Anteils von $\{X_0 \in B\}$ an allen Pfaden, die B treffen.

(2). Allgemeiner gilt für jede messbare Teilmenge $A \in \mathcal{A}$ des Pfadraumes:

$$E[\tau_A; A] = P[\tau_A < \infty],$$

wobei $\tau_A = \min\{n \geq 1 : (X_n, X_{n+1}, \dots) \in A\}$ die erste Zeit ist, zu der der verschobene Pfad in A liegt.

Beweis. Für $n \in \mathbb{N}$ gilt wegen der Stationarität des Prozesses:

$$\begin{aligned} E[\min(T_B, n); X_0 \in B] &= \sum_{k=0}^{n-1} P[T_B > k \text{ und } X_0 \in B] \\ &= \sum_{k=0}^{n-1} P[X_0 \in B, X_1 \notin B, \dots, X_k \notin B] \\ &= \sum_{k=0}^{n-1} P[X_{n-k} \in B, X_{n-k+1} \notin B, \dots, X_n \notin B] \\ &= P[T_B \leq n]. \end{aligned}$$

Hierbei haben wir verwendet, dass $T_B \leq n$ genau dann gilt, wenn zu einer der Zeiten $n - k$ ($k = 0, 1, \dots, n - 1$) ein letzter Besuch in B vor der Zeit n stattfindet. Die Aussage folgt für $n \rightarrow \infty$. \square

Nach dem Wiederkehrrsatz von Kac kehrt der Prozess (X_n) auf der Menge $\{X_0 \in B\}$ P -fast sicher nach B zurück. Durch Anwenden dieser Aussage auf die Teilfolgen $(X_{nk})_{n \geq 0}$, $k \in \mathbb{N}$, die alle wieder stationäre Prozesse unter P sind, erhalten wir sogar:

Korollar 6.9 (Rekurrenz von stationären Prozessen). *Jeder stationäre Prozess (X_n, P) ist rekurrent in folgendem Sinne: Für alle $B \in \mathcal{S}$ gilt $X_n \in B$ unendlich oft P -fast sicher auf $\{X_0 \in B\}$.*

Bemerkung (Wiederkehrsatz von Poincaré). Allgemeiner gilt für $A \in \mathcal{A}$:

$$(X_n(\omega), X_{n+1}(\omega), \dots) \in A \quad \text{unendlich oft für } P\text{-fast alle } \omega \in A.$$

6.2.3 Anwendung auf Markovketten

Wir betrachten nun eine zeithomogene Markovkette (X_n, P_x) mit abzählbarem Zustandsraum S im kanonischen Modell.

Definition (Positive Rekurrenz). *Ein Zustand $x \in S$ heißt **positiv rekurrent**, falls die mittlere Rückkehrzeit $E_x[T_x]$ endlich ist. Die Markovkette heißt **positiv rekurrent**, falls alle Zustände positiv rekurrent sind.*

Aus dem Wiederkehrsatz von Kac folgt unmittelbar:

Korollar 6.10 (Gleichgewichte und mittlere Rückkehrzeiten). *Sei μ ein Gleichgewicht der Markovkette.*

(1). *Für alle $x \in S$ gilt*

$$\mu(x) \cdot E_x[T_x] = P_\mu[T_x < \infty].$$

Insbesondere sind alle Zustände x mit $\mu(x) > 0$ positiv rekurrent.

(2). *Ist die Übergangsmatrix irreduzibel, dann sind sogar alle $x \in S$ positiv rekurrent mit*

$$\mu(x) = \frac{1}{E_x[T_x]}. \quad (6.2.7)$$

Insbesondere ist das Gleichgewicht in diesem Fall eindeutig.

Beweis. (1). Da die Markovkette mit Startverteilung μ ein stationärer Prozess ist, gilt nach dem Satz von Kac:

$$\mu(x) \cdot E_x[T_x] = E_\mu[T_x; X_0 = x] = P_\mu[T_x < \infty] \quad \text{für alle } x \in S.$$

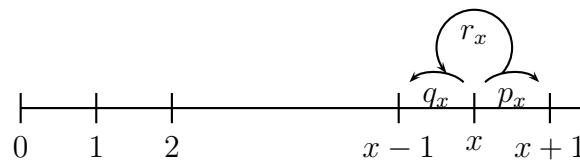
(2). Bei Irreduzibilität folgt globale Rekurrenz, also $P_y[T_x < \infty] = 1$ für alle $x, y \in S$. Nach (1) erhalten wir $\mu(x) \cdot E_x[T_x] = 1$, und damit $\mu(x) > 0$ und $E_x[T_x] < \infty$ für alle x .

□

Beispiel (Birth-Death-Process). Wir betrachten erneut die zeithomogene Markovkette mit Zustandsraum $S = \{0, 1, 2, \dots\}$ und Übergangswahrscheinlichkeiten

$$p(x, x+1) = p_x, \quad p(x, x-1) = q_x, \quad p(x, x) = r_x,$$

$p_x, q_x, r_x > 0$ mit $p_x + q_x + r_x = 1$, $q_0 = 0$, und $p_x, q_x > 0$ für alle $x \geq 1$.



Offensichtlich gilt Irreduzibilität. In Abschnitt 2.2 haben wir gezeigt, dass das eindeutige Gleichgewicht μ durch (2.2.7) mit $\mu(0) = 1/Z$ gegeben ist, falls

$$Z := \sum_{x=0}^{\infty} \frac{p_0 \cdot p_1 \cdot \dots \cdot p_{x-1}}{q_1 \cdot q_2 \cdot \dots \cdot q_x} < \infty$$

gilt. Für die mittleren Rückkehrzeiten folgt dann

$$E_x[T_x] = 1/\mu(x) \quad \text{für alle } x \geq 0.$$

Gilt dagegen $Z = \infty$, dann existiert keine Gleichgewichtsverteilung. Wir werden in Satz 6.11 sehen, dass in diesem Fall auch keiner der Zustände $x \in S$ positiv rekurrent ist. Durch Lösen des Dirichletproblems kann man zudem zeigen, dass die Markovkette genau dann rekurrent ist, wenn

$$\sum_{x=0}^{\infty} \frac{q_1 \cdot q_2 \cdot \dots \cdot q_x}{p_1 \cdot p_2 \cdot \dots \cdot p_x} = \infty \quad \text{gilt.}$$

6.3 Ergodizität

In diesem Abschnitt werden wir ein Gesetz der großen Zahlen für positiv rekurrente Markovketten beweisen. Dabei verwenden wir, dass die Verläufe der Markovkette während verschiedener Exkursionen von einem Punkt aus unabhängig voneinander und identisch verteilt sind. Langzeitmittelwerte verhalten sich daher asymptotisch wie der Erwartungswert des zeitlichen Mittelwerts über eine Exkursion. Als Vorbereitung überlegen wir uns, dass der Anteil der mittleren Exkursionszeit, den die Markovkette in bestimmten Bereichen verbringt, eine Gleichgewichtsverteilung definiert.

Wie zuvor sei (X_n, P_x) eine zeithomogene Markovkette mit abzählbarem Zustandsraum S und Übergangsmatrix $p(x, y)$ im kanonischen Modell. Ferner sei

$$T_x = \min\{n \geq 1 : X_n = x\}$$

die erste Treffer- bzw. Rückkehrzeit zum Punkt x .

6.3.1 Positive Rekurrenz und Gleichgewichte

Für einen Zustand $x \in S$ sei

$$\mu_x[B] := E_x \left[\sum_{n=0}^{T_x-1} I_B(X_n) \right] = \sum_{n=0}^{\infty} P[X_n \in B ; n < T_x] \quad (6.3.1)$$

die mittlere Anzahl der Besuche in einer Menge $B \subset S$ während einer Exkursion von x .

Definition (Invariantes Maß). Ein positives Maß ν auf S heißt **invariant** bzgl. der Übergangsmatrix p , falls

$$\sum_{x \in S} \nu(x) p(x, y) = \nu(y) \quad \text{für alle } y \in S \quad \text{gilt.}$$

Ein Gleichgewicht ist also eine invariante Wahrscheinlichkeitsverteilung.

Satz 6.11 (Exkursionen und invariante Maße). (1). Ist $x \in S$ ein rekurrenter Zustand der Markovkette, dann ist μ_x ein invariantes Maß mit Gesamtmasse $\mu_x[S] = E_x[T_x]$.

(2). Ist x positiv rekurrent, dann ist das normierte Maß

$$\bar{\mu}_x[B] = \frac{\mu_x[B]}{E_x[T_x]} \quad \left(= \frac{\text{mittlere Aufenthaltszeit in } B}{\text{mittlere Exkursionsdauer}} \right)$$

ein Gleichgewicht der Markovkette.

Bei positiver Rekurrenz existiert also stets ein Gleichgewicht. Umgekehrt haben wir in Korollar 6.9 bereits gezeigt, dass Gleichgewichtsverteilungen nur positiv rekurrenten Zuständen eine strikt positive Gesamtmasse zuordnen. Ist die Markovkette zudem irreduzibel, dann ist die Gleichgewichtsverteilung nach Korollar 6.9 eindeutig, d.h. die Verteilung $\bar{\mu}_x$ hängt nicht vom Startpunkt x ab.

Beweis. (1). Ist x rekurrent, dann gilt P_x -fast sicher $T_x < \infty$, und damit $X_{T_x} = x = X_0$. Für $B \subseteq S$ folgt

$$\sum_{n=0}^{T_x-1} I_B(X_n) = \sum_{n=0}^{T_x-1} I_B(X_{n+1}).$$

Mit der Markoveigenschaft erhalten wir damit

$$\begin{aligned}
 \mu_x[B] &= E_x \left[\sum_{n=0}^{T_x-1} I_B(X_{n+1}) \right] \\
 &= \sum_{n=0}^{\infty} P_x[X_{n+1} \in B; n < T_x] \\
 &\stackrel{\text{ME}}{=} \sum_{n=0}^{\infty} E_x [p(X_n, B); n < T_x] \\
 &= \int p(y, B) \mu_x(dy) = (\mu_x p)[B],
 \end{aligned}$$

d.h. μ_x ist ein invariantes Maß. Die Gesamtmasse ist

$$\mu_x[S] = E_x \left[\sum_{n=0}^{T_x-1} I_S(X_n) \right] = E_x[T_x].$$

- (2). Ist x positiv rekurrent, dann hat μ_x endliche Gesamtmasse, also erhält man durch Normieren ein Gleichgewicht. □

6.3.2 Ein Gesetz der großen Zahlen für Markovketten

Wir können nun das Hauptresultat dieses Abschnitts formulieren. Dazu betrachten wir eine Markovkette (X_n, P) mit beliebiger Startverteilung. Die entsprechende Kette mit Startwert x im kanonischen Modell bezeichnen wir mit (X_n, P_x) . Für $n \in \mathbb{N}$ und $y \in S$ sei

$$B_y(n) := \sum_{i=0}^{n-1} I_{\{X_i=y\}}$$

die Anzahl der Besuche der Markovkette im Zustand y vor der Zeit n .

Satz 6.12 (Ergodensatz für Markovketten, 1. Version). *Sei (X_n, P) eine irreduzible homogene Markovkette mit abzählbarem Zustandsraum S .*

- (1). *Ist die Markovkette rekurrent, dann gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{B_y(n)} \sum_{i=0}^{n-1} f(X_i) = E_y \left[\sum_{i=0}^{T_y-1} f(X_i) \right] = \int f d\mu_y$$

P -fast sicher für jede Funktion $f : S \rightarrow \mathbb{R}_+$ und alle $y \in S$. Hierbei ist μ_y das durch (6.3.1) definierte invariante Maß.

(2). Existiert eine Gleichgewichtsverteilung $\bar{\mu}$, dann folgt $\bar{\mu}_y = \bar{\mu}$ für alle $y \in S$ und

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_i) = \int f d\bar{\mu} \quad P\text{-fast sicher.}$$

Die letzte Aussage ist ein *Gesetz der großen Zahlen* für irreduzible, positiv rekurrente Markovketten, und eine erste Version eines *Ergodensatzes* für Markovketten: Die „zeitlichen“ Mittelwerte $\frac{1}{n} \sum_{i=0}^{n-1} f(X_i)$ konvergieren fast sicher gegen den „räumlichen“ Mittelwert $\int f d\bar{\mu}$ der Funktion f bzgl. der Gleichgewichtsverteilung. Insbesondere ergibt sich

$$\bar{\mu}(x) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} I_{\{x\}}(X_i) \quad P\text{-fast sicher für alle } x \in S,$$

d.h. die Gewichte der Gleichgewichtsverteilung sind die asymptotischen relativen Häufigkeiten der Zustände $x \in S$. Dieser Zusammenhang kann in beide Richtungen verwendet werden:

- (1). Berechnung der asymptotischen relativen Häufigkeiten durch Lösen des linearen Gleichungssystems $\bar{\mu} = \bar{\mu}p$.
- (2). Schätzen der Gleichgewichtsverteilung:

$$\bar{\mu} \approx \frac{1}{n} \sum_{i=0}^{n-1} \delta_{X_i} \quad \text{für große } n.$$

Beweis von Satz 6.12. Da die Zufallsvariablen X_i nicht unabhängig sind, können wir nicht wie im Beweis des klassischen Gesetzes der großen Zahlen verfahren. Stattdessen nutzen wir aus, dass die Markovkette jedes mal, wenn sie den Punkt x trifft, neu startet – unabhängig vom vorherigen Verlauf. Durch Zerlegen der Summe in Teilsummen über diese verschiedenen Zyklen erhalten wir eine Summe von unabhängigen Zufallsvariablen, auf die sich das klassische Gesetz der großen Zahlen anwenden lässt:

- (1). Wir betrachten die Markovkette o.B.d.A. im kanonischen Modell. Sei $T^{(k)}$ die k -te Besuchszeit bzw. Rückkehrzeit zu einem festen Zustand $y \in S$, d.h. $T^{(0)} = 0$, und

$$T^{(k+1)} = T^{(k)} + T_y \circ \theta^{T^{(k)}} \quad \text{für alle } k \geq 0.$$

Da die Kette irreduzibel und rekurrent ist, gilt $T^{(k)} < \infty$ P -fast sicher für alle k , und damit

$$\sum_{i=1}^{T^{(l)}} f(X_i) = \sum_{k=0}^{l-1} Y_k \quad \text{mit} \quad Y_k := \sum_{i=T^{(k)}+1}^{T^{(k+1)}} f(X_i).$$

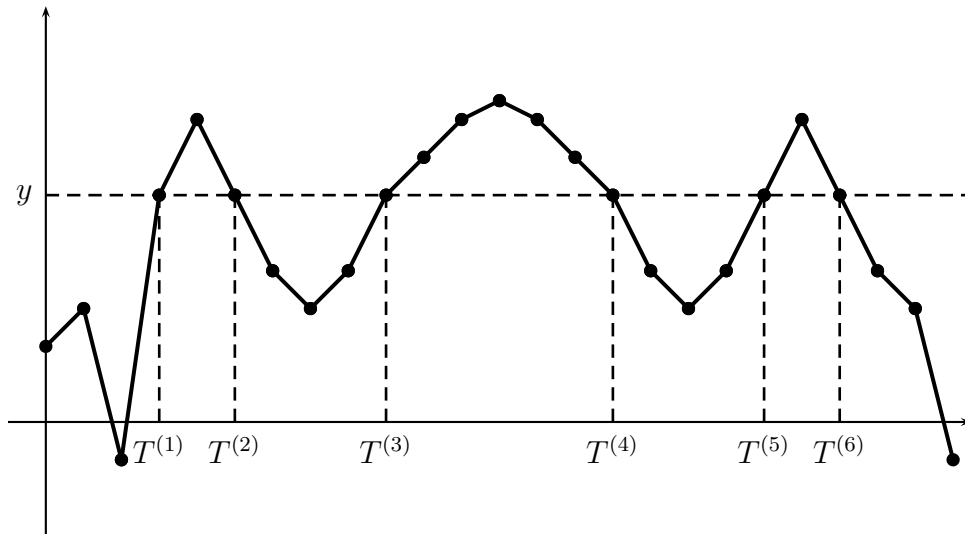


Abbildung 6.5: Regenerative Zyklen.

Wir zeigen nun, dass aufgrund der starken Markoveigenschaft die Zufallsvariablen Y_k unter P unabhängig und identisch verteilt sind. Es gilt nämlich

$$Y_k = \sum_{i=T^{(k)}+1}^{T^{(k)}+T_y \circ \theta^{T^{(k)}}} f(X_i) = \sum_{j=1}^{T_y} f(X_j \circ \theta^{T^{(k)}}) = Y_0 \circ \theta^{T^{(k)}},$$

also

$$P[Y_k \in B \mid \mathcal{F}_{T^{(k)}}] \stackrel{\text{SME}}{=} P_y[Y_0 \in B] \quad \text{für alle } B \subset S,$$

d.h. Y_k ist unabhängig von $\mathcal{F}_{T^{(k)}}$ mit Verteilung $P_y \circ Y_0^{-1}$. Da die Zufallsvariablen Y_0, \dots, Y_{k-1} $\mathcal{F}_{T^{(k)}}$ -messbar sind, folgt die Unabhängigkeit der $Y_k, k \geq 0$, unter P . Zudem erhalten wir für $k \geq 1$:

$$E[Y_k] = E_y[Y_0] = E_y \left[\sum_{i=1}^{T_y} f(X_i) \right] = \int f d\mu_y.$$

Nach dem *Gesetz der großen Zahlen* folgt dann:

$$\lim_{l \rightarrow \infty} \frac{1}{l} \sum_{i=1}^{T^{(l)}} f(X_i) = \lim_{l \rightarrow \infty} \frac{1}{l} \sum_{k=0}^{l-1} Y_k = \int f d\mu_y \quad P\text{-fast sicher.} \quad (6.3.2)$$

Ist die Anzahl $B_y(n)$ der Besuche in y vor der Zeit n gleich l , dann gilt

$$T^{(l-1)} < n \leq T^{(l+1)},$$

also

$$\frac{1}{l} \sum_{i=1}^{T^{(l-1)}} f(X_i) \leq \frac{1}{B_y(n)} \sum_{i=1}^n f(X_i) \leq \frac{1}{l} \sum_{i=1}^{T^{(l+1)}} f(X_i). \quad (6.3.3)$$

Für $n \rightarrow \infty$ konvergiert auch $l = B_y(n)$ gegen unendlich, da die Markovkette rekurrent ist. Da die linke und rechte Seite von (6.3.3) nach (6.3.2) für $l \rightarrow \infty$ gegen $\int f d\mu_y$ konvergieren, folgt

$$\lim_{n \rightarrow \infty} \frac{1}{B_y(n)} \sum_{i=0}^{n-1} f(X_i) = \lim_{n \rightarrow \infty} \frac{1}{B_y(n)} \sum_{i=1}^n f(X_i) = \int f d\mu_y \quad P\text{-fast sicher.}$$

(2). Anwenden von Aussage (1) mit der konstanten Funktion $f \equiv 1$ liefert

$$\frac{n}{B_y(n)} \xrightarrow{n \rightarrow \infty} \mu_y[S] \quad P\text{-fast sicher.}$$

Da eine invariante Verteilung existiert, ist die Kette positiv rekurrent, d.h. $\mu_y[S] < \infty$.

Daher folgt für $f \geq 0$:

$$\frac{1}{n} \sum_{i=0}^{n-1} f(X_i) = \frac{B_y(n)}{n} \cdot \frac{1}{B_y(n)} \sum_{i=0}^{n-1} f(X_i) \xrightarrow{n \rightarrow \infty} \frac{\int f d\mu_y}{\mu_y[S]} = \int f d\bar{\mu}_y$$

P -fast sicher für $n \rightarrow \infty$. Da die Markovkette nach Voraussetzung irreduzibel ist, ist die Gleichgewichtsverteilung nach Korollar 6.9 eindeutig. Also gilt $\bar{\mu}_y = \bar{\mu}$ für alle $y \in S$.

□

Beispiel (Kartenhaus / Maschinenerneuerung). Wir betrachten die Markovkette aus dem Beispiel von oben.

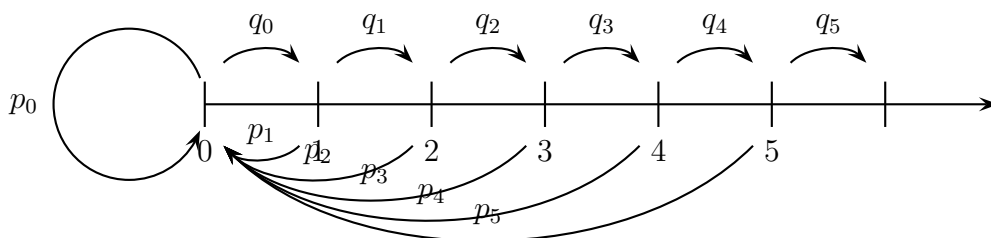


Abbildung 6.6: Übergangswahrscheinlichkeiten Kartenhaus.

Mit den Übergangswahrscheinlichkeiten $p(i, 0) = p_i \in (0, 1)$ und $p(i, i + 1) = q_i = 1 - p_i$ erhalten wir

$$P_0[T_0 > n] = q_0 \cdot q_1 \cdot \dots \cdot q_{n-1}, \quad \text{und damit}$$

$$E_0[T_0] = \sum_{n=0}^{\infty} \prod_{i=0}^{n-1} q_i.$$

Gilt $E_0[T_0] < \infty$, dann ist die Kette irreduzibel und positiv rekurrent. Für die asymptotische relative Häufigkeit des Zusammenfallens des Kartenhauses folgt dann nach Satz 6.12 und Korollar 6.9:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} I_{\{0\}}(X_i) = \bar{\mu}(0) = \frac{1}{E_0[T_0]} \quad P_x\text{-fast sicher für alle } x \in S.$$

Beispiel (Markov Chain Monte Carlo Verfahren (MCMC)). Sei μ eine Wahrscheinlichkeitsverteilung auf eine abzählbaren Menge S , deren Gewichte wir bis auf eine Normierungskonstante kennen bzw. berechnen können. Um Erwartungswerte von Funktionen $f : S \rightarrow \mathbb{R}_+$ bzgl μ approximativ zu berechnen, können wir dann wie in Kapitel ?? beschrieben eine irreduzible Übergangsmatrix p mit Gleichgewicht μ bestimmen, und eine Markovkette (X_n, P) mit dieser Übergangsmatrix simulieren. Nach Satz 6.13 liefern die empirischen Mittelwerte

$$\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n f(X_i)$$

dann eine konsistente Folge von Schätzern für den gesuchten Erwartungswert

$$\theta = \int f d\mu.$$

Für praktische Anwendungen ist es wichtig, den Schätzfehler zu quantifizieren. Eine erste Aussage in diese Richtung liefert ein zentraler Grenzwertsatz für Markovketten, siehe z.B. [T. Komorowski, C. Landim, S. Olla: Fluctuations in Markov Processes].

6.3.3 Allgemeiner Ergodensätze

Die Aussage von Satz 6.12 lässt sich wesentlich allgemeiner formulieren. Wir notieren zunächst eine elementare, aber wichtige Erweiterung:

Satz 6.13 (Ergodensatz für Markovketten, 2. Version). *Ist (X_n, P) eine irreduzible, zeitlich homogene Markovkette, und $\bar{\mu}$ ein Gleichgewicht des Übergangskerns p , dann gilt*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} f(X_i, X_{i+1}, \dots, X_{i+r}) = \int \cdots \int f(x_0, x_1, \dots, x_r) \bar{\mu}(dx_0) p(x_0, dx_1) \cdots p(x_{r-1}, dx_r)$$

P -fast sicher für alle $r \geq 0$ und $f : S^{r+1} \rightarrow \mathbb{R}_+$.

Wir geben nur die Beweisidee an, und überlassen die Ausführung der Details dem Leser als Übung:

Beweis-Skizze. Der Prozess $\tilde{X}_i := (X_i, X_{i+1}, \dots, X_{i+r})$ ist wieder eine Markovkette mit Zustandsraum

$$\tilde{S} = \{(x_0, \dots, x_r) \in S^{r+1} : p(x_i, x_{i+1}) > 0 \quad \forall 0 \leq i < r\},$$

Übergangsmatrix

$$\tilde{p}((x_0, \dots, x_r), (y_0, \dots, y_r)) = \delta_{x_1}(y_0) \delta_{x_2}(y_1) \cdots \delta_{x_r}(y_{r-1}) p(x_r, y_r),$$

und Gleichgewichtsverteilung

$$\tilde{\mu}(x_0, \dots, x_r) = \bar{\mu}(x_0) \cdot p(x_0, x_1) \cdots p(x_{r-1}, x_r).$$

Ist (X_n) irreduzibel, so auch (\tilde{X}_n) . Die Behauptung folgt daher aus Satz 6.12. □

Eine wichtige Anwendung von Satz 6.13 ist das *Schätzen der Übergangsmatrix* einer Markovkette: Für $x, y \in S$ gilt

$$p(x, y) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} I_{\{X_i=x, X_{i+1}=y\}} \quad P\text{-fast sicher,}$$

d.h. die Übergangswahrscheinlichkeiten sind die asymptotischen relativen Häufigkeiten der Übergänge.

Beispiel (Neues im I.I.D. Fall). Auch im i.i.d. Fall liefert Satz 6.13 eine neue Aussage: Ist X_0, X_1, \dots eine Folge unabhängiger, identisch verteilter Zufallsvariablen („Buchstaben“) mit Werten in einer endlichen oder abzählbaren Menge S („Alphabet“), dann ergibt sich für die asymptotische relative Häufigkeit eines Wortes $(a_0, a_1, \dots, a_k) \in S^{k+1}$:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} I_{\{X_i=a_0, X_{i+1}=a_1, \dots, X_{i+k}=a_k\}} = \prod_{j=0}^k \mu(a_j) \quad P\text{-fast sicher,}$$

wobei $\mu(a) = P[X_i = a]$ die Wahrscheinlichkeit des Buchstabens a ist.

Mit abstrakteren Argumenten kann man Ergodensätze im allgemeinen Rahmen dynamischer Systeme beweisen. Zum Abschluss dieses Abschnittes geben wir kurz ein entsprechendes zentrales Resultat ohne Beweis wieder. Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $\theta : \Omega \rightarrow \Omega$ eine *maßerhaltende Abbildung*, d.h. $P \circ \theta^{-1} = P$. Den Raum (Ω, \mathcal{A}, P) zusammen mit der maßerhaltenden Abbildung θ nennt man auch ein *dynamisches System*. Beispielsweise ist der Shiftoperator

θ auf dem Pfadraum maßerhaltend bzgl. der Verteilung P eines stationären stochastischen Prozesses. Die σ -Algebra \mathcal{J} der θ -invarianten Ereignisse ist definiert als

$$\mathcal{J} = \{A \in \mathcal{A} : \theta^{-1}(A) = A\}.$$

Beispielsweise sind die Zufallsvariablen $\liminf \frac{1}{n} \sum_{i=0}^{n-1} F \circ \theta^i$ und $\limsup \frac{1}{n} \sum_{i=0}^{n-1} F \circ \theta^i$ für jede \mathcal{A} -messbare Abbildung $F : \Omega \rightarrow \mathbb{R}$ messbar bzgl. \mathcal{J} . Allgemein sind alle θ -invarianten Ereignisse asymptotisch. Das Maß P heißt *ergodisch*, falls $P[A] \in \{0, 1\}$ für alle $A \in \mathcal{J}$ gilt. In dieser allgemeinen Situation kann man zeigen:

Satz 6.14 (Birkhoffs individueller Ergodensatz). Für jede Funktion $F \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$ gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} F(\theta^i(\omega)) = E[F | \mathcal{J}](\omega) \quad \text{für } P\text{-fast alle } \omega \in \Omega.$$

Ist P ergodisch, dann folgt

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} F \circ \theta^i = E[F] \quad P\text{-fast sicher.}$$

Der Beweis findet sich z.B. in den Wahrscheinlichkeitstheorie-Büchern von Breiman oder Durrett. Die \mathcal{L}^2 -Konvergenz lässt sich mit wesentlich einfacheren funktionalanalytischen Methoden zeigen (Ergodensatz von Neumann, siehe z.B. [Varadhan: Probability Theory]).