

R. Grübel
Universität Hannover
Institut für Mathematische Stochastik

STOCHASTISCHE PROZESSE

Sommersemester 2005

0. Einführung

Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, (E, \mathcal{B}) ein meßbarer Raum und T eine nicht-leere Menge. Für jedes $t \in T$ sei $X_t : \Omega \rightarrow E$ eine $(\mathcal{A}, \mathcal{B})$ -meßbare Abbildung (also eine E -wertige Zufallsgröße). Dann nennt man $(X_t)_{t \in T}$ einen *stochastischen Prozeß* mit *Zeitmenge* T und *Zustandsraum* E . Im allgemeinen hat man $T \subset \mathbb{R}$; besonders wichtig sind die Fälle $T = \mathbb{N}, \mathbb{Z}, \mathbb{N}_0, [0, \infty), \mathbb{R}$. Im ersten Teil der Vorlesung ist E eine beliebige abzählbare (also endliche oder abzählbar unendliche) Menge und \mathcal{B} die zugehörige Potenzmenge. In den späteren Teilen werden wir hauptsächlich reellwertige Prozesse betrachten, also $E = \mathbb{R}$, mit gelegentlichen Ausflügen in den \mathbb{R}^d , $d > 1$; \mathcal{B} ist dann die σ -Algebra der Borelsche Teilmengen von E .

Von der Syntax nun zur Semantik: Stochastische Prozesse dienen als Modelle für zeitabhängige zufällige Vorgänge: X_t beschreibt den Zustand eines Systems zur Zeit t . Spielt man beispielsweise nach einer bestimmten Strategie Roulette, so könnte X_n das verbleibende Kapital nach n Durchgängen sein ($E = \mathbb{R}$, $T = \mathbb{N}_0$). In einer Telefonzentrale interessiert man sich für die Anzahl X_t der bis zur Zeit t angekommenen Anrufe ($E = \mathbb{N}_0$, $T = [0, \infty)$). Die Position eines bestimmten Moleküls in einem Gas kann wegen der zahlreichen Zusammenstöße mit anderen Molekülen (abhängig von der Temperatur) als zeitabhängige Zufallsgröße angesehen werden ($E = \mathbb{R}^3$, $T = \mathbb{R}$ oder \mathbb{R}_+). Einige einfache Beispiele für stochastische Prozesse wurden bereits in den hier vorausgesetzten Vorlesungen Stochastik I und II besprochen.

Solche stochastischen Modelle für zeitabhängige zufällige Vorgänge sind in der Physik (statistische Mechanik), den Ingenieurwissenschaften (Kontrolltheorie), Operations Research (Bedienungssysteme), den Wirtschaftswissenschaften (Modellierung von Aktienkursen) und in zahlreichen anderen Gebieten von großer Bedeutung. Stochastische Prozesse haben auch wichtige innermathematische Anwendungen gefunden, beispielsweise im Zusammenhang mit partiellen Differentialgleichungen. Der Übergang zur modernen Analysis ist fließend—in der Tat ist es eines der Ziele der Vorlesung, die Grundlagen für eine Vorlesung über stochastische Analysis zu vermitteln.

Wichtige Anmerkung: Dieses Skript enthält (in geringfügigem Umfang) Material, das in der Vorlesung selbst nicht besprochen wurde; außerdem fehlen natürlich (in größerem Umfang) Illustrationen, Beispiele und Erläuterungen, die in der Vorlesung ad hoc gegeben wurden. Hinweise auf die Vorlesungen Stochastik I und II beziehen sich auf das Sommersemester 2003 und das Wintersemester 2003/2004.

1. Markov-Ketten in diskreter Zeit

1.1 Die Markov-Eigenschaft. Es sei E eine endliche oder abzählbar unendliche Menge wie zum Beispiel $\{0, 1, \dots, n\}$, \mathbb{N}_0 oder \mathbb{Z}^d ($d \geq 1$), $\mathcal{B} = \mathcal{P}(E)$. Weiter sei $T = \mathbb{N}_0$, wir haben also einen Prozeß $X = (X_n)_{n \in \mathbb{N}_0}$. Ein solcher Prozeß heißt *Markov-Kette*, wenn er die *Markov-Eigenschaft* hat, d.h. wenn für alle $i_0, \dots, i_{n+1} \in E$ und alle $t_0, \dots, t_{n+1} \in T$ mit $t_0 < \dots < t_{n+1}$

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \dots, X_{t_n} = i_n) = \mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_n} = i_n)$$

gilt. Hierbei setzen wir natürlich voraus, dass die beteiligten bedingten Wahrscheinlichkeiten definiert sind. Bei Zeitbereich \mathbb{N}_0 ist die Markov-Eigenschaft äquivalent dazu, dass für alle $n \in \mathbb{N}$ und alle $i_0, \dots, i_{n+1} \in E$

$$\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = \mathbb{P}(X_{n+1} = i_{n+1} | X_n = i_n)$$

gilt (Übungsaufgabe). Betrachtet man den Zeitpunkt n als Gegenwart und somit $0, 1, \dots, n-1$ als Vergangenheit und $n+1$ als Zukunft, so lässt sich dies auch wie folgt formulieren: von der gesamten bisherigen Geschichte des Prozesses ist für die zukünftige Entwicklung allein der gegenwärtige Zustand relevant. Kurz: Markov-Prozesse haben kein Gedächtnis. Gilt darüberhinaus

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(X_1 = j | X_0 = i) \quad \text{für alle } i, j \in E, n \in \mathbb{N},$$

so nennt man die Markov-Kette *zeitlich homogen*; wir werden im folgenden nur solche Prozesse behandeln. Die (Ein-Schritt-)Übergangswahrscheinlichkeiten

$$p_{ij} := \mathbb{P}(X_1 = j | X_0 = i), \quad i, j \in E,$$

fassen wir zur *Übergangsmatrix* $P = (p_{ij})_{i, j \in E}$ zusammen (im Falle $\#E = \infty$ hat diese Matrix unendlich viele Zeilen und Spalten). Die Einträge (Elemente) einer Übergangsmatrix sind als Wahrscheinlichkeiten natürlich nicht-negativ, und als Zeilensummen erhält man

$$\sum_{j \in E} p_{ij} = \sum_{j \in E} \mathbb{P}(X_1 = j | X_0 = i) = \mathbb{P}(X_1 \in E | X_0 = i) = 1,$$

also stets den Wert 1. Matrizen mit diesen Eigenschaften nennt man *stochastische Matrizen*.

Die Verteilung von X_0 bezeichnet man als *Anfangs-* oder *Startverteilung* des Prozesses. Verteilungen auf $(E, \mathcal{P}(E))$ sind bekanntlich durch die zugehörige Wahrscheinlichkeitsmassenfunktion festgelegt. Wir betrachten die Startverteilung als (u.U. unendlich langen) Zeilenvektor $q = (q_i)_{i \in E}$ mit Komponenten (Einträgen) $q_i = \mathbb{P}(X_0 = i)$. Gelegentlich schreiben wir \mathbb{P}_q , wenn wir andeuten wollen, dass die Kette die Startverteilung q hat, und kurz \mathbb{P}_i im Falle $\mathbb{P}(X_0 = i) = 1$ (q ist dann die Ein-Punkt-Verteilung in i). Bei gegebenem q und P kann man die Wahrscheinlichkeiten von im Zusammenhang mit X interessierenden Ereignissen ausrechnen. Wir illustrieren dies zunächst an einem nicht ganz ernst gemeinten Beispiel.

BEISPIEL 1.1 Angenommen, das Wetter am Tage n wird durch eine Zufallsgröße X_n mit Werten in $E = \{1, 2, 3\}$ beschrieben, wobei

1: Wetter schlecht, 2: grau in grau, 3: Sonne scheint

bedeute. Wir nehmen an, dass $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette ist mit Übergangsmatrix

$$P = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix}.$$

Dieses Modell geht beispielsweise davon aus, dass morgen mit Wahrscheinlichkeit $1/4$ die Sonne scheint, wenn heute das Wetter grau-in-grau ist, und zwar unabhängig davon, ob gestern die Sonne schien oder das Wetter schlecht war. Mit welcher Wahrscheinlichkeit scheint übermorgen die Sonne, wenn heute das Wetter schlecht ist? Eine Zerlegung nach dem Zustand zur Zeit 1 liefert unter wiederholter Verwendung der Markov-Eigenschaft

$$\begin{aligned} \mathbb{P}(X_2 = 3 | X_0 = 1) &= \sum_{j=1}^3 \mathbb{P}(X_2 = 3, X_1 = j | X_0 = 1) \\ &= \sum_{j=1}^3 \mathbb{P}(X_2 = 3 | X_0 = 1, X_1 = j) \mathbb{P}(X_1 = j | X_0 = 1) \\ &= \sum_{j=1}^3 \mathbb{P}(X_2 = 3 | X_1 = j) \mathbb{P}(X_1 = j | X_0 = 1) \\ &= p_{11}p_{13} + p_{12}p_{23} + p_{13}p_{33} \\ &= \frac{3}{4} \cdot 0 + \frac{1}{4} \cdot \frac{1}{4} + 0 \cdot \frac{1}{4} \\ &= \frac{1}{16}. \end{aligned}$$

Ähnlich erhält man als Wahrscheinlichkeit dafür, dass das Wetter 7 Tage lang schlecht bleibt (wenn es gegenwärtig schlecht ist), den Wert

$$\mathbb{P}_1(X_1 = X_2 = \dots = X_7 = 1) = p_{11}^7 = \left(\frac{3}{4}\right)^7.$$

Die n -Schritt-Übergangswahrscheinlichkeiten werden durch

$$p_{ij}(n) = \mathbb{P}(X_n = j | X_0 = i), \quad i, j \in E$$

definiert und zur Matrix $P(n) := (p_{ij}(n))_{i,j \in E}$ zusammengefasst.

SATZ 1.2 (Die Chapman-Kolmogorov-Gleichungen)

Für alle $i, j \in E$, $n, m \in \mathbb{N}$ gilt

$$p_{ij}(n+m) = \sum_{k \in E} p_{ik}(n)p_{kj}(m),$$

also in Matrixschreibweise $P(n+m) = P(n)P(m)$. Insbesondere gilt $P(n) = P^n$ für alle $n \in \mathbb{N}$.

BEWEIS: Eine Zerlegung nach dem Zustand des Prozesses zur Zeit n liefert, ganz ähnlich wie im obigen Beispiel,

$$\begin{aligned} p_{ij}(n+m) &= \sum_{k \in E} \mathbb{P}(X_{n+m} = j, X_n = k | X_0 = i) \\ &= \sum_{k \in E} \mathbb{P}(X_{n+m} = j | X_0 = i, X_n = k) \mathbb{P}(X_n = k | X_0 = i) \\ &= \sum_{k \in E} \mathbb{P}(X_{n+m} = j | X_n = k) \mathbb{P}(X_n = k | X_0 = i) \\ &= \sum_{k \in E} \mathbb{P}(X_m = j | X_0 = k) \mathbb{P}(X_n = k | X_0 = i) \\ &= \sum_{k \in E} p_{kj}(m) p_{ik}(n), \end{aligned}$$

wobei wir die Homogenität ausgenutzt haben. □

Will man in der Situation von Beispiel 1.1 das Wetter eine Woche im voraus wissen, so ergibt sich mit Satz 1.2

$$\begin{aligned} P(7) = P^7 &= \frac{1}{16384} \begin{pmatrix} 9105 & 5461 & 1818 \\ 9097 & 5462 & 1825 \\ 9104 & 5461 & 1819 \end{pmatrix} \\ &\approx \begin{pmatrix} 0.555725 & 0.333313 & 0.110962 \\ 0.555237 & 0.333374 & 0.111389 \\ 0.555664 & 0.333313 & 0.111023 \end{pmatrix}. \end{aligned}$$

Bemerkenswerterweise unterscheiden sich die Zeiten kaum voneinander—in einer Woche scheint mit Wahrscheinlichkeit ≈ 0.11 die Sonne, egal ob heute das Wetter schlecht, grau-in-grau oder gut ist. Wir werden auf dieses Phänomen später genau eingehen.

Die Verteilung des Prozesses $(X_n)_{n \in \mathbb{N}_0}$ zur Zeit n wird durch den ‘Verteilungsvektor’ $q(n) = (q_i(n))_{i \in E}$ mit $q_i(n) := \mathbb{P}(X_n = i)$ beschrieben; mit $n = 0$ erhält man die bereits oben erwähnte Startverteilung. Wir betrachten auch $q(n)$ als Zeilenvektor und können dann eine weitere Verbindung zu den Operationen der linearen Algebra herstellen. (Man beachte, dass durch die möglicherweise unendliche Länge der Zeilen und Spalten hier keine Probleme entstehen, da nur nicht-negative Werte vorkommen.)

SATZ 1.3 *Es gilt $q(n) = q(0)P^n$ für alle $n \in \mathbb{N}$.*

BEWEIS: Zerlegung nach dem Zustand zur Zeit 0 liefert

$$\begin{aligned} q_i(n) = \mathbb{P}(X_n = i) &= \sum_{k \in E} \mathbb{P}(X_n = i | X_0 = k) \mathbb{P}(X_0 = k) \\ &= \sum_{k \in E} q_k(0) p_{ki}(n). \end{aligned}$$

□

Durch die Startverteilung q und die Übergangsmatrix P sind also die Verteilungen aller X_n festgelegt. Wie wir in Abschnitt 1.6 sehen werden, gilt sogar eine erheblich weiter gehende Aussage.

Markov-Ketten tauchen in vielen Anwendungen auf. Häufig hat man eine Struktur der Form

$$X_0 = i_0 \in E \text{ fest,} \quad X_{n+1} = f(X_n, Y_{n+1}) \text{ für alle } n \in \mathbb{N}_0,$$

wobei $(Y_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten Zufallsgrößen mit Werten in einer Menge S und $f : E \times S \rightarrow E$ eine Funktion ist. Man sieht leicht, dass dieses System die Markov-Eigenschaft hat, und wegen

$$\mathbb{P}(X_{n+1} = j | X_n = i) = \mathbb{P}(Y_{n+1} \in \{s \in S : f(i, s) = j\})$$

hat man auch die zeitliche Homogenität. Einen solchen Prozess $(X_n)_{n \in \mathbb{N}_0}$ kann man als *stochastisches dynamisches System* auffassen.

BEISPIEL 1.4 (i) Ist G eine endliche oder abzählbar unendliche Gruppe mit Verknüpfung ‘+’ und neutralem Element ‘0’, und ist $(Y_n)_{n \in \mathbb{N}}$ eine Folge von

unabhängigen, identisch verteilten Zufallsgrößen mit Werten in G , so bezeichnet man die Folge $(X_n)_{n \in \mathbb{N}_0}$ der Partialsummen

$$X_0 = 0, \quad X_n = \sum_{m=1}^n Y_m,$$

als *Irrfahrt in G mit Start in 0 und Schrittweitenverteilung $\mathcal{L}(Y_1)$* . Von besonderem Interesse ist der Fall $G = \mathbb{Z}$, $\mathbb{P}(Y_n = 1) = 1 - \mathbb{P}(Y_n = -1)$, die *einfache Irrfahrt* auf den ganzen Zahlen ('drunkard's walk'). Im Falle $\mathbb{P}(Y_n = 1) = \mathbb{P}(Y_n = -1)$ spricht man von der einfachen *symmetrischen* Irrfahrt in Dimension $d = 1$. Die höherdimensionalen Varianten ergeben sich bei $E = \mathbb{Z}^d$, $d > 1$, wenn die Zuwächse $Y_n, n \in \mathbb{N}$, auf $\{\pm e_i : 1 \leq i \leq d\}$ gleichverteilt sind. Hierbei seien e_1, \dots, e_d die d kanonischen Einheitsvektoren in \mathbb{R}^d , d.h. e_i ist der Vektor in \mathbb{R}^d , der in der i -ten Komponente eine 1 und sonst lauter Nullen hat.

(ii) Es sei $E = \{0, 1, \dots, N\}$,

$$p_{ij} = \begin{cases} p, & i < N, j = i + 1, \\ 1 - p, & i > 0, j = i - 1, \\ 1, & i = N, j = \mathbb{N} \text{ oder } i = 0, j = 0. \end{cases}$$

Dies modelliert die Situation aus Beispiel 4.27 der Vorlesung Stochastik I: X_n ist das Kapital des Spielers I nach n Runden, $i_0 \in \{1, \dots, N-1\}$ sein Anfangskapital. In jeder Runde wird eine ' p -Münze' (p ist die Wahrscheinlichkeit für 'Kopf') geworfen; bei 'Kopf' gewinnt, bei 'Zahl' verliert Spieler I eine Geldeinheit. Im Falle $X_n = 0$ (Ruin) oder $X_n = N$ (Ziel erreicht) wird das Spiel beendet. Auch dieses Beispiel passt in das obige Schema: $Y_n, n \in \mathbb{N}$, seien unabhängig und identisch verteilt mit $P(Y_n = 1) = 1 - P(Y_n = -1) = p$,

$$f(x, y) := \begin{cases} x + y, & \text{falls } x \in \{1, \dots, N-1\}, \\ x, & \text{falls } x \in \{0, N\}. \end{cases}$$

Man kann $(X_n)_{n \in \mathbb{N}_0}$ auch zu einer einfachen Irrfahrt $(\tilde{X}_n)_{n \in \mathbb{N}_0}$ auf \mathbb{Z} (siehe Teil (i) des Beipiels) in Beziehung setzen:

$$X_0 := \tilde{X}_0, \quad X_n := \tilde{X}_{\tau \wedge n}$$

mit

$$\tau := \inf\{k \in \mathbb{N}_0 : \tilde{X}_k \in \{0, N\}\}.$$

In Worten: die Irrfahrt wird beim Austritt aus $\{1, \dots, N-1\}$ gestoppt.

(iii) Wir betrachten eine zufällige Zeichenkette über dem Alphabet $E_0 := \{A, C, G, T\}$, also eine Folge $(Z_n)_{n \in \mathbb{N}_0}$ von unabhängigen Zufallsgrößen mit Werten in E_0 und

$$\mathbb{P}(Z_n = A) = p_A, \dots, \mathbb{P}(Z_n = T) = p_T,$$

wobei $p_A, \dots, p_T \geq 0$ mit $p_A + \dots + p_T = 1$ vorgegeben sind. Angenommen, man sucht ein bestimmtes Muster ('pattern') der Länge L in der Zeichenkette. Fasst man jeweils L aufeinanderfolgende Buchstaben zu einem Wort zusammen, setzt man also

$$X_n^{(L)} := (Z_n, Z_{n+1}, \dots, Z_{n+L-1}),$$

so erhält man eine Markov-Kette mit Zustandsraum $E = E_0^L$ und das erstmalige Auftreten des Patterns entspricht der Eintrittszeit von $(X_n^{(L)})_{n \in \mathbb{N}_0}$ in einen bestimmten Zustand. Dieser stochastische Prozess hat die bemerkenswerte Eigenschaft, dass Variablen mit Zeitabstand $\geq L$ voneinander unabhängig sind. Fragen dieser Art sind beispielsweise in der Genanalyse von Interesse.

(iv) Ein besonders einfaches Beispiel aus dem (großen) Bereich der Warteschlangen: Es seien $0, 1, 2, \dots$ die Zeitpunkte, zu denen eine Seilbahn abfährt, die maximal K Touristen befördern kann. Im Zeitintervall $(n-1, n)$ (nach Abfahrt der letzten und vor Ankunft der nächsten Bahn) kommen Y_n neue Touristen an; X_n sei die Anzahl der auf die n -te Bahn wartenden Personen, also die Länge der Warteschlange. Offensichtlich gilt dann

$$X_{n+1} = \max\{0, X_n - K\} + Y_{n+1}$$

(der erste Term ist die Anzahl der Touristen, die bei der letzten Abfahrt zurückgelassen wurden, der zweite die Anzahl der neu hinzugekommenen). Setzt man voraus, dass die Y -Variablen unabhängig und identisch verteilt sind, so erhält man wieder eine homogene Markov-Kette, die sich wiederum mit $f(x, y) := \max\{0, x - k\} + y$ als stochastisches dynamisches System schreiben lässt.

(v) Ein weiteres Beispiel aus dem Bereich 'Operations Research', konkret zur Lagerhaltung: X_n bezeichnet die am n -ten Tag unmittelbar nach Ladenschluss vorhandene Anzahl der Artikel eines bestimmten Typs; Y_n ist die Anzahl der Artikel, die am n -ten Tag verkauft werden (können, wenn vorhanden). Der Ladeninhaber implementiert folgende Strategie: im Falle $X_n \leq R$ werden S Artikel dieses Typs über Nacht besorgt. Dann gilt

$$X_{n+1} = \begin{cases} \max\{X_n + S - Y_{n+1}, 0\}, & X_n \leq R, \\ \max\{X_n - Y_{n+1}, 0\}, & X_n > R. \end{cases}$$

Wie in Teil (iv) erhält man eine homogene Markov-Kette, wenn die Y -Variablen unabhängig und identisch verteilt sind. Interessant wird dieses Beispiel durch den Zusammenhang zur Optimierungs- und Kontrolltheorie: wie sollte man, wenn beispielsweise die Verteilung der Y -Variablen und die Lagerhaltungskosten bekannt sind, die Parameter R und S wählen?

(vi) Die in Beispiel 8.19 der Vorlesung Stochastik II eingeführten Verzweigungsprozesse $(Z_n)_{n \in \mathbb{N}_0}$ sind Markov-Ketten mit Zustandsraum \mathbb{N}_0 und absorbierendem Zustand 0: $\mathbb{P}(Z_{n+1} = 0 | Z_n = 0) = 1$.

1.2 Klassifikation der Zustände. Es sei $(X_n)_{n \in \mathbb{N}_0}$ eine homogene Markov-Kette mit Zustandsraum E und Übergangsmatrix P . Wir sagen, dass i auf j führt, $i, j \in E$, Schreibweise: $i \rightsquigarrow j$, wenn es ein $n \in \mathbb{N}$ mit $p_{ij}(n) > 0$ gibt*; man kann dann ‘in endlich vielen Schritten von i nach j kommen’. Ein Zustand $i \in E$ mit der Eigenschaft, dass für alle $j \in E$ die Implikation $i \rightsquigarrow j \Rightarrow j \rightsquigarrow i$ gilt, heißt *wesentlich*, sonst *unwesentlich*. Auf der Menge der wesentlichen Zustände wird durch

$$i \sim j : \iff i \rightsquigarrow j \text{ und } j \rightsquigarrow i$$

eine Äquivalenzrelation definiert, durch die diese Menge in Äquivalenzklassen von *miteinander kommunizierenden Zuständen* zerlegt wird. Besteht der gesamte Zustandsraum aus einer einzigen solchen Äquivalenzklasse, gilt also

$$\forall i, j \in E \quad \exists n \in \mathbb{N} : p_{ij}(n) > 0$$

(in Worten: alle Zustände kommunizieren miteinander), so nennt man die Kette *irreduzibel*. Wir werden uns im folgenden hauptsächlich mit irreduziblen Markov-Ketten beschäftigen. Schließlich bezeichnet man noch mit

$$d(i) := \text{ggT}\{n \in \mathbb{N} : p_{ii}(n) > 0\}$$

die *Periode* von $i \in E$, im Falle $d(i) = 1$ heißt der Zustand i *aperiodisch*. Bei einer einfachen Irrfahrt auf \mathbb{Z} mit $p_{i,i+1} = p$, $p_{i,i-1} = 1 - p$ für alle $i \in \mathbb{Z}$ kann man im Falle $0 < p < 1$ offensichtlich jeden Zustand i von jedem anderen Zustand j in $|i - j|$ Schritten erreichen, solche Irrfahrten sind also irreduzibel. In einer irreduziblen Kette haben alle Zustände dieselbe Periode (Übungsaufgabe). Bei der einfachen Irrfahrt (mit $0 < p < 1$) gilt offensichtlich

$$X_n - X_0 \text{ ungerade für } n \text{ ungerade, gerade für } n \text{ gerade,}$$

d.h. alle Zustände haben die Periode 2. Der Prozess besucht abwechselnd die Teilmengen $2\mathbb{Z}$ und $2\mathbb{Z} + 1$ des Zustandsraums \mathbb{Z} .

* Dies wird in der Literatur nicht einheitlich gehandhabt, häufig wird auch $n = 0$ zugelassen

LEMMA 1.5 *Ist der Zustand $i \in E$ aperiodisch, so existiert ein n_0 derart, dass $p_{ii}(n) > 0$ für alle $n \geq n_0$ gilt.*

BEWEIS: Es sei $H := \{m \in \mathbb{N} : p_{ii}(m) > 0\}$. Wegen

$$\mathbb{N} \ni d_n := \text{ggT}(H \cap \{1, \dots, n\}) \downarrow \text{ggT}(H) = 1 \quad \text{mit } n \rightarrow \infty$$

gibt es Zahlen $m_1, \dots, m_k \in H$ mit $\text{ggT}\{m_1, \dots, m_k\} = 1$. Zu diesen gibt es dann wiederum ganze Zahlen l_1, \dots, l_k mit $\sum_{j=1}^k l_j m_j = 1$. Diesen Sachverhalt aus der elementaren Zahlentheorie kann man leicht wie folgt einsehen:

$$G := \left\{ \sum_{j=1}^k l_j m_j : l_1, \dots, l_k \in \mathbb{Z} \right\}$$

ist eine Untergruppe von \mathbb{Z} , also im Falle $1 \notin G$ von der Form $G = d\mathbb{Z}$ mit einem $d > 1$. Dieses d wäre dann Teiler von m_1, \dots, m_k , die ja bei geeigneter Wahl der l -Koeffizienten als Elemente von G erscheinen, im Widerspruch zu $\text{ggT}\{m_1, \dots, m_k\} = 1$.

Nun ist H additiv abgeschlossen: Hat man nämlich $n, m \in H$, so folgt mit

$$p_{ii}(n+m) = \sum_{j \in I} p_{ij}(n) p_{ji}(m) \geq p_{ii}(n) p_{ii}(m) > 0$$

(ein einfaches Argument, das wir noch mehrmals verwenden werden), dass auch $n+m$ in H liegt. Ein Umschreiben von $\sum_{j=1}^k l_j m_j = 1$ zu

$$\sum_{\substack{j=1 \\ l_j > 0}}^k l_j m_j = \sum_{\substack{j=1 \\ l_j < 0}}^k (-l_j) m_j + 1$$

zeigt, dass H zwei aufeinanderfolgende Zahlen m und $m+1$ enthält. Setzt man nun $n_0 := m(m-1)$, so gilt für beliebiges $n \geq n_0$

$$\begin{aligned} n &= mk + j \quad \text{mit } 0 \leq j \leq m-1, k \geq m-1 \\ &= (k-j)m + j(m+1), \end{aligned}$$

also $n \in H$. □

Dieses Lemma werden wir bei der Analyse des Langzeitverhaltens von Markov-Ketten benötigen. Ein erstes qualitatives Kriterium in diesem Zusammenhang

ergibt sich aus der Frage, ob eine Kette bei Start in i mit Wahrscheinlichkeit 1 irgendwann einmal zu diesem Startpunkt zurückkehrt. Es sei

$$f_{ij}(n) := \mathbb{P}(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j | X_0 = i)$$

die Wahrscheinlichkeit dafür, dass man ausgehend von i nach genau n Schritten erstmalig den Zustand j erreicht. Man beachte, dass im Falle $i = j$ der Zeitpunkt 0 nicht mitzählt. Dann ist

$$f_{ij}^* := \sum_{n=1}^{\infty} f_{ij}(n) \quad \left(:= \lim_{N \rightarrow \infty} \sum_{n=1}^N f_{ij}(n) \right)$$

die Wahrscheinlichkeit dafür, dass j überhaupt irgendwann einmal besucht wird, wenn man in i startet. Im Falle $f_{ii}^* = 1$ nennt man den Zustand i *rekurrent*, sonst *transient*. Für alle $i \in E$ definieren wir die *Eintrittszeit* in i durch

$$T_i := \inf\{n \in \mathbb{N} : X_n = i\},$$

wobei wir $\inf \emptyset = \infty$ vereinbaren; T_i ist also eine Zufallsgröße mit Werten in $\mathbb{N} \cup \{\infty\}$. Klar:

$$\mathbb{P}(T_j = n | X_0 = i) = f_{ij}(n), \quad \mathbb{P}(T_j < \infty | X_0 = i) = f_{ij}^*.$$

Nach Definition ist ein Zustand i genau dann rekurrent, wenn

$$\mathbb{P}(T_i < \infty | X_0 = i) = 1$$

gilt, die Kette also bei Start in i mit Wahrscheinlichkeit 1 zu ihrem Startpunkt zurückkehrt.

SATZ 1.6 *Ein Zustand $i \in E$ ist genau dann rekurrent, wenn $\sum_{n=1}^{\infty} p_{ii}(n) = \infty$ gilt.*

BEWEIS: Zu den Folgen $(p_{ii}(n))_{n \in \mathbb{N}_0}$, $(f_{ii}(n))_{n \in \mathbb{N}}$ gehören die Potenzreihen $P(z) = \sum_{n=0}^{\infty} p_{ii}(n)z^n$, $F(z) := \sum_{n=1}^{\infty} f_{ii}(n)z^n$. Die Folgen sind beschränkt, also ist der jeweilige Konvergenzradius mindestens 1. Eine Zerlegung nach dem Zeitpunkt der ersten Rückkehr ergibt unter Verwendung der Markov-Eigenschaft und unter Beachtung von $p_{ii}(0) = \mathbb{P}(X_0 = i | X_0 = i) = 1$

$$\begin{aligned} p_{ii}(n) &= \mathbb{P}(X_n = i | X_0 = i) \\ &= \sum_{k=1}^n \mathbb{P}(X_1 \neq i, \dots, X_{k-1} \neq i, X_k = i, X_n = i | X_0 = i) \\ &= \sum_{k=1}^n f_{ii}(k) p_{ii}(n-k) \end{aligned}$$

für alle $n \in \mathbb{N}$, also

$$\begin{aligned} P(z) &= p_{ii}(0) + \sum_{n=1}^{\infty} \sum_{k=1}^n f_{ii}(k) z^k p_{ii}(n-k) z^{n-k} \\ &= p_{ii}(0) + \sum_{k=1}^{\infty} f_{ii}(k) z^k \sum_{n=k}^{\infty} p_{ii}(n-k) z^{n-k} \\ &= 1 + F(z)P(z). \end{aligned}$$

Auflösen nach $P(z)$ ergibt

$$P(z) = \frac{1}{1 - F(z)} \quad \text{für alle } z \text{ mit } |z| < 1.$$

Wir verwenden nun ein einfaches Hilfsmittel aus der Analysis, das auch als Satz von Abel bekannt ist und sich beispielsweise recht einfach mit dem Satz von der monotonen Konvergenz beweisen lässt: Im Falle $a_n \geq 0$ für alle $n \in \mathbb{N}_0$ gilt

$$\lim_{z \uparrow 1} \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} a_n.$$

Hierbei ist ausdrücklich auch bestimmte Divergenz zugelassen, d.h. Gleichheit gilt auch dann, wenn eine der beiden Seiten den Wert unendlich hat. Diesen Satz verwenden wir nun zweimal in der obigen Darstellung:

$$\begin{aligned} \sum_{n=1}^{\infty} f_{ii}(n) = 1 &\iff \lim_{z \uparrow 1} F(z) = 1 \\ &\iff \lim_{z \uparrow 1} P(z) = \infty \\ &\iff \sum_{n=1}^{\infty} p_{ii}(n) = \infty. \end{aligned}$$

□

Dieser Satz hat eine naheliegende stochastische Interpretation, die übrigens auch für einen alternativen, mehr ‘stochastischen’ Beweis genutzt werden kann. Hierzu beobachten wir zunächst, dass $p_{ii}(n) = E[1_{\{i\}}(X_n) | X_0 = i]$ gilt. Mit dem Satz von der monotonen Konvergenz folgt hieraus

$$\begin{aligned} \sum_{n=0}^{\infty} p_{ii}(n) &= \sum_{n=0}^{\infty} E[1_{\{i\}}(X_n) | X_0 = i] \\ &= E\left[\sum_{n=0}^{\infty} 1_{\{i\}}(X_n) | X_0 = i\right] \\ &= E[\#\{n \in \mathbb{N}_0 : X_n = i\} | X_0 = i]. \end{aligned}$$

Man kann Satz 1.6 also wie folgt interpretieren: Die Kette kehrt genau dann mit Wahrscheinlichkeit 1 zu ihrem Startpunkt i zurück, wenn die erwartete Anzahl der Besuche in i insgesamt unendlich ist. Die Markov-Eigenschaft impliziert, dass bei jeder Rückkehr nach i 'alles wieder von vorn anfängt', dieses Resultat ist also nicht überraschend.

BEISPIEL 1.7 Wir betrachten die einfache symmetrische Irrfahrt auf \mathbb{Z}^d , zunächst für $d = 1$. Um in $2n$ Schritten ausgehend von 0 wieder in 0 zu landen (die obigen Periodizitätsüberlegungen zeigen, dass nur gerade n -Werte in Frage kommen), müssen genau n Schritte in die eine und n Schritte in die andere Richtung gemacht werden. Auf den Münzwurf übersetzt bedeutet dies, dass bei $2n$ -maligem Wurf exakt n -mal Kopf erscheint. Dies führt auf

$$p_{00}(2n) = \binom{2n}{n} 2^{-2n}.$$

Mit der Stirling-Formel

$$n! \sim \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n} \quad \text{mit } n \rightarrow \infty$$

erhält man hieraus

$$p_{00}(2n) \sim \frac{\sqrt{2\pi} (2n)^{2n+\frac{1}{2}} e^{-2n}}{(\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n})^2 2^{2n}} \sim \frac{1}{\sqrt{\pi n}},$$

also $\sum_{n=0}^{\infty} p_{00}(n) = \infty$. Alle Zustände sind rekurrent, oder kurz: die eindimensionale symmetrische Irrfahrt ist rekurrent. Dies gilt auch noch bei $d = 2$, ab $d = 3$ ist die einfache symmetrische Irrfahrt jedoch transient (wird in den Übungen näher ausgeführt).

Rekurrenz ist wie Aperiodizität eine sogenannte *Klasseneigenschaft*: In einer irreduziblen Kette sind entweder alle Zustände rekurrent oder keiner, d.h. $\mathbb{P}(T_i < \infty | X_0 = i) = 1$ gilt für alle Zustände $i \in E$, wenn es für ein $i \in E$ gilt (Übungsaufgabe). Was passiert bei $i \neq j$?

LEMMA 1.8 *In einer irreduziblen und rekurrenten Markov-Kette gilt $f_{ij}^* = 1$ für alle $i, j \in E$.*

BEWEIS: Es seien $i, j \in E$ mit $i \neq j$. Eine Zerlegung nach dem ersten

(möglicherweise nicht vor n oder sogar nie stattfindenden) Besuch in j ergibt

$$\begin{aligned} p_{ji}(n) &= \sum_{k=1}^{n-1} \mathbb{P}(X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j, X_n = i | X_0 = j) \\ &\quad + \mathbb{P}(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = i | X_0 = j) \\ &= \sum_{k=1}^{n-1} \mathbb{P}(X_1 \neq j, \dots, X_{k-1} \neq j, X_k = j | X_0 = j) p_{ji}(n-k) \\ &\quad + \mathbb{P}(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = i | X_0 = j). \end{aligned}$$

Ist nun n der kleinste Wert mit $p_{ji}(n) > 0$ (beachte: $j \rightsquigarrow i$), so muss

$$\alpha := \mathbb{P}(X_1 \neq j, \dots, X_{n-1} \neq j, X_n = i | X_0 = j) > 0$$

gelten, da alle anderen Summanden einen Faktor $p_{ji}(m)$ mit einem $m < n$ enthalten. In Worten: Es ist es möglich, von j nach i ohne zwischenzeitliche Rückkehr nach j zu gehen. Wegen

$$\begin{aligned} 0 &= 1 - f_{jj}^* \\ &= \mathbb{P}(X_m \neq j \text{ für alle } m \in \mathbb{N} | X_0 = j) \\ &\geq \mathbb{P}(X_m \neq j \text{ für alle } m \in \mathbb{N}, X_n = i | X_0 = j) \\ &= \mathbb{P}(X_m \neq j \text{ für } m = 1, \dots, n-1, X_n = i | X_0 = j) \\ &\quad \cdot \mathbb{P}(X_m \neq j \text{ für alle } m > n | X_n = i) \\ &= \alpha(1 - f_{ij}^*) \end{aligned}$$

impliziert $\alpha > 0$ die gewünschte Aussage $f_{ij}^* = 1$. □

Die Eintrittszeitverteilung bei beliebiger Startverteilung lässt sich wegen

$$\mathbb{P}(T_i = n) = \sum_{j \in I} \mathbb{P}(T_i = n | X_0 = j) \mathbb{P}(X_0 = j)$$

auf die Werte $f_{ij}(n)$, $i, j \in E$, zurückführen. Analog erhält man bei einer irreduziblen und rekurrenten Kette

$$\begin{aligned} \mathbb{P}(T_i < \infty) &= \sum_{j \in E} \mathbb{P}(T_i < \infty | X_0 = j) \mathbb{P}(X_0 = j) \\ &= \sum_{j \in E} f_{ji}^* \mathbb{P}(X_0 = j) \\ &= \sum_{j \in E} 1 \mathbb{P}(X_0 = j) = 1, \end{aligned}$$

d.h. jeder Zustand wird mit Wahrscheinlichkeit 1 irgendwann einmal besucht, egal mit welcher Verteilung man startet.

1.3 Der Hauptgrenzwertsatz für Markov-Ketten. Es sei $(X_n)_{n \in \mathbb{N}_0}$ wieder eine Markov-Kette mit Übergangsmatrix P und Zustandsraum E .

DEFINITION 1.9 Ein Wahrscheinlichkeitsvektor $(\pi_i)_{i \in E}$ (also $\pi_i \geq 0$ für alle $i \in E$, $\sum_{i \in E} \pi_i = 1$) heißt *stationäre Verteilung* zu $(X_n)_{n \in \mathbb{N}_0}$ (eigentlich besser: zur Übergangsmatrix P), wenn $\pi P = \pi$ gilt, also

$$\sum_{j \in E} \pi_j p_{ji} = \pi_i \quad \text{für alle } i \in E.$$

Aus $\pi P = \pi$ folgt natürlich $\pi P^n = \pi$ für alle $n \in \mathbb{N}$. Ist also die Startverteilung einer Markov-Kette eine solche stationäre Verteilung, so haben nach Satz 1.3 alle X_n , $n \in \mathbb{N}_0$, diese Verteilung—die Kette ist *stationär*, ‘im Gleichgewicht’.

BEISPIEL 1.10 (i) Bei endlichem Zustandsraum läuft die Bedingung aus Definition 1.9 auf ein lineares Gleichungssystem hinaus. In Beispiel 1.1 (Wetter) hatten wir

$$P = \begin{pmatrix} 3/4 & 1/4 & 0 \\ 1/4 & 1/2 & 1/4 \\ 1/2 & 1/4 & 1/4 \end{pmatrix},$$

erhalten also

$$\begin{aligned} \pi_1 \cdot \frac{3}{4} + \pi_2 \cdot \frac{1}{4} + \pi_3 \cdot \frac{1}{2} &= \pi_1, \\ \pi_1 \cdot \frac{1}{4} + \pi_2 \cdot \frac{1}{2} + \pi_3 \cdot \frac{1}{4} &= \pi_2, \\ \pi_1 \cdot 0 + \pi_2 \cdot \frac{1}{4} + \pi_3 \cdot \frac{1}{4} &= \pi_3. \end{aligned}$$

Als Lösungsmenge erhält man den linearen Raum

$$L = \{(5\alpha, 3\alpha, \alpha) : \alpha \in \mathbb{R}\},$$

und die Bedingung $\pi_1 + \pi_2 + \pi_3 = 1$ führt auf die (eindeutige) Lösung $\pi = (5/9, 1/3, 1/9)$. Zur ‘Wetterkette’ existiert also genau eine stationäre Verteilung. Interessanterweise liegen die Zeilen der in Anschluss an Satz 1.2 berechneten Matrix P^7 alle sehr nahe bei π .

(ii) Bei der einfachen symmetrischen Irrfahrt auf \mathbb{Z} , mit $p_{i,i+1} = p_{i,i-1} = 1/2$ für alle $i \in \mathbb{Z}$, führt die Bedingung $\pi P = \pi$ auf

$$\pi_{i-1} \cdot \frac{1}{2} + \pi_{i+1} \cdot \frac{1}{2} = \pi_i \quad \text{für alle } i \in \mathbb{Z},$$

also $\pi_{i+1} - \pi_i = \pi_i - \pi_{i-1}$. Die Differenzen einer Lösung hätten somit alle denselben Wert, was sich mit $\sum_{i \in \mathbb{Z}} \pi_i = 1$ nicht verträgt. Zur einfachen symmetrischen Irrfahrt existiert also keine stationäre Verteilung.

Zentrales Resultat dieses Abschnitts ist der nächste Satz, für den wir noch zwei Hilfsmittel benötigen.

LEMMA 1.11 *Ist π stationäre Verteilung einer irreduziblen Markov-Kette, so gilt $\pi_i > 0$ für alle $i \in E$.*

BEWEIS: $\pi P^n = \pi$ impliziert $\sum_{j \in E} \pi_j p_{ji}(n) = \pi_i$, also folgt aus $\pi_i = 0$

$$\exists n \in \mathbb{N} : p_{ji}(n) > 0 \quad \Rightarrow \quad \pi_j = 0.$$

Wegen der vorausgesetzten Irreduzibilität ist die linke Seite dieser Implikation stets erfüllt, d.h. man hätte $\pi \equiv 0$. \square

LEMMA 1.12 *Existiert zu einer irreduziblen Markov-Kette eine stationäre Verteilung, so ist die Kette rekurrent.*

BEWEIS: Rekurrenz ist, wie am Ende von Abschnitt 1.2 erklärt, eine Klasseneigenschaft. Folglich muss, wenn die Kette nicht rekurrent ist,

$$\sum_{n=1}^{\infty} p_{ij}(n) < \infty \quad \text{für alle } i, j \in E \quad (1)$$

gelten. (Detaillierter begründet man dies mit Satz 1.6 in Verbindung mit Aufgabe 2 (iv) und der Beobachtung, dass aus $i \rightsquigarrow j$ die Ungleichung $f_{ij}^* > 0$ folgt.) Sei nun $\pi = (\pi_i)_{i \in E}$ eine stationäre Verteilung, sei $j \in E$. Nach Lemma 1.11 gilt dann $\pi_j > 0$. Wähle nun ein $\epsilon > 0$ mit $2\epsilon < \pi_j$. Wegen $\sum_{i \in E} \pi_i = 1$ existiert zu diesem ϵ eine endliche Teilmenge F des Zustandsraums E mit $\sum_{i \notin F} \pi_i < \epsilon$ (bei endlichem E kann man $F = E$ nehmen). Für jedes $i \in F$ wiederum findet man unter Verwendung von (1) ein $n_0(i)$ derart, dass

$$p_{ij}(n) < \frac{\epsilon}{\#F} \quad \text{für alle } n \geq n_0(i)$$

gilt. Sei $n \geq \max\{n_0(i) : i \in F\}$. Dann folgt insgesamt wegen $\pi P^n = \pi$

$$\begin{aligned} 2\epsilon < \pi_j &= \sum_{i \in E} \pi_i p_{ij}(n) \\ &= \sum_{i \in F} \pi_i p_{ij}(n) + \sum_{i \notin F} \pi_i p_{ij}(n) \\ &\leq \sum_{i \in F} p_{ij}(n) + \sum_{i \notin F} \pi_i \\ &< \sum_{i \in F} \frac{\epsilon}{\#F} + \epsilon = 2\epsilon, \end{aligned}$$

also der gewünschte Widerspruch. \square

SATZ 1.13 (Der Hauptgrenzwertsatz für Markov-Ketten)

Es sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible und aperiodische Markov-Kette mit Übergangsmatrix P . Existiert zu P eine stationäre Verteilung π , so ist diese eindeutig, und es gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(X_n = j | X_0 = i) \left(= \lim_{n \rightarrow \infty} p_{ij}(n) \right) = \pi_j \quad \text{für alle } i, j \in E.$$

Man kann diesen Satz mit analytischen Methoden beweisen (siehe z.B. Krenzel). Wir verwenden stattdessen eine sehr elegante ‘stochastische’ Methode, die mit der sogenannten *Koppelungskonstruktion* arbeitet.

BEWEIS: Es seien $(X_n)_{n \in \mathbb{N}_0}$ und $(X'_n)_{n \in \mathbb{N}_0}$ unabhängige Markov-Ketten, beide mit Übergangsmatrix P . Die Startverteilung von $(X_n)_{n \in \mathbb{N}_0}$ sei in i konzentriert, die Startverteilung von $(X'_n)_{n \in \mathbb{N}_0}$ sei π . Wir definieren $(Y_n)_{n \in \mathbb{N}_0}$ durch $Y_n := (X_n, X'_n)$ für alle $n \in \mathbb{N}_0$. Dies ist wieder eine Markov-Kette, nun mit Zustandsraum $E \times E$ und Übergangsmatrix

$$\tilde{P} = (\tilde{p}_{(i,j),(k,l)})_{(i,j),(k,l) \in E \times E}, \quad \tilde{p}_{(i,j),(k,l)} = p_{ik} \cdot p_{jl}.$$

Wir zeigen zunächst, dass eine Kette mit Übergangsmatrix \tilde{P} (wieder) irreduzibel ist. Seien hierzu $(i, j), (k, l) \in E \times E$. Da in der Originalkette $i \rightsquigarrow k$ und $j \rightsquigarrow l$ gilt, gibt es $n, m \in \mathbb{N}$ mit $p_{ik}(n) > 0$, und $p_{jl}(m) > 0$. Die Aperiodizität der Ausgangskette impliziert nach Lemma 1.5, dass es $n_0, m_0 \in \mathbb{N}$ gibt mit

$$p_{kk}(n') > 0 \quad \text{für alle } n' \geq n_0, \quad p_{ll}(m') > 0 \quad \text{für alle } m' \geq m_0.$$

Unter Verwendung der bereits häufiger gebrauchten Ungleichung

$$p_{ij}(n+m) \geq p_{ik}(n) p_{kj}(m) \quad \text{für alle } i, j, k \in E, m, n \in \mathbb{N}$$

folgt hieraus, dass es ein \tilde{n} gibt mit $p_{ik}(\tilde{n}) > 0$ und $p_{jl}(\tilde{n}) > 0$, man hat also $\tilde{p}_{(i,j),(k,l)}(\tilde{n}) > 0$. Dies bedeutet $(i, j) \rightsquigarrow (k, l)$; da (i, j) und (k, l) beliebige Elemente von $E \times E$ waren, ist damit die Irreduzibilität der bivariaten Kette bewiesen.

Zum Nachweis der Aperiodizität reicht es wegen der bereits bewiesenen Irreduzibilität,

$$\text{ggT}\{n \in \mathbb{N} : \tilde{p}_{(i,j),(i,j)}(n) > 0\} = 1$$

für ein $(i, j) \in E \times E$ zu zeigen. Für die Zustände auf der Diagonalen ergibt sich dies aber unmittelbar aus der vorausgesetzten Aperiodizität der Ausgangskette und der Beziehung $\tilde{p}_{(i,i),(i,i)}(n) = p_{ii}(n)^2$.

Sei nun $i_0 \in E$,

$$T := \inf\{n \in \mathbb{N} : Y_n = (i_0, i_0)\}$$

die Eintrittszeit der Y -Kette in den Zustand (i_0, i_0) , also der erste Zeitpunkt, zu dem sich die Ketten X und X' beide in i_0 befinden. Man sieht leicht, dass durch

$$\tilde{\pi}_{(i,j)} := \pi_i \cdot \pi_j \quad \text{für alle } (i, j) \in E \times E$$

eine stationäre Verteilung zu \tilde{P} definiert wird. Nach Lemma 1.12 ist die Kette dann rekurrent, nach der Schlussbemerkung in Abschnitt 1.2 impliziert dies $\mathbb{P}(T < \infty) = 1$. Nun kommen die entscheidenden Schritte, ausgehend von der intuitiven Vorstellung, dass auf $\{T \leq n\}$, also ab der ‘Koppelung’ in i_0 , die Verteilungen von X_n und X'_n sich nicht (mehr) unterscheiden. Konkret führt eine Zerlegung nach dem Wert von T , zusammen mit der Markov-Eigenschaft, auf

$$\begin{aligned} \mathbb{P}(X_n = j, T \leq n) &= \sum_{m=1}^n \mathbb{P}(X_n = j, T = m) \\ &= \sum_{m=1}^n \mathbb{P}(T = m) \mathbb{P}(X_n = j | X_m = i_0) \\ &= \sum_{m=1}^n \mathbb{P}(T = m) p_{i_0 j}(n - m) \\ &= \sum_{m=1}^n \mathbb{P}(T = m) \mathbb{P}(X'_n = j | X'_m = i_0) \\ &= \dots = \mathbb{P}(X'_n = j, T \leq n), \end{aligned}$$

also auf

$$\begin{aligned} |p_{ij}(n) - \pi_j| &= |\mathbb{P}(X_n = j) - \mathbb{P}(X'_n = j)| \\ &\leq |\mathbb{P}(X_n = j, T \leq n) - \mathbb{P}(X'_n = j, T \leq n)| \\ &\quad + |\mathbb{P}(X_n = j, T > n) - \mathbb{P}(X'_n = j, T > n)| \\ &\leq 2\mathbb{P}(T > n). \end{aligned}$$

(Wir haben hierbei die Dreiecksungleichung gebraucht. Der erste Summand verschwindet, wie direkt davor gezeigt wurde, bei dem zweiten haben wir recht grob abgeschätzt.) Insgesamt erhält man

$$0 \leq \limsup_{n \rightarrow \infty} |p_{ij}(n) - \pi_j| \leq 2 \limsup_{n \rightarrow \infty} \mathbb{P}(T > n) = 0$$

wegen $\mathbb{P}(T < \infty) = 1$, also die Behauptung. Die Aussage

$$\pi_j = \lim_{n \rightarrow \infty} p_{ij}(n) \quad \text{für alle } j \in E$$

impliziert natürlich auch die Eindeutigkeit von π . □

Zu diesem Hauptgrenzwertsatz gehört eine stochastische Interpretation der π -Werte, die wir im nächsten Unterabschnitt erhalten.

1.4 Erneuerungstheorie in diskreter Zeit. Es sei $(Y_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen, identisch verteilten \mathbb{N} -wertigen Zufallsgrößen; $f_i := \mathbb{P}(Y_1 = i)$ für $i \in \mathbb{N}$. Die zugehörigen Partialsummen werden definiert durch

$$S_0 = 0, \quad S_n := \sum_{i=1}^n Y_i \quad \text{für } n \in \mathbb{N}.$$

Die Variable Y_n kann beispielsweise als Lebensdauer der n -ten Glühbirne interpretiert werden, wobei bei Defekt direkt ein Austausch erfolgt. Von besonderem Interesse ist die *Erneuerungsfolge* $(u_n)_{n \in \mathbb{N}_0}$ zur Lebensdauerverteilung $(f_i)_{i \in \mathbb{N}}$,

$$u_0 := 1, \quad u_n := \mathbb{P}(S_m = n \text{ für ein } m \in \mathbb{N}_0).$$

Im Zusammenhang mit der Glühbirnen-Anwendung ist u_n die Wahrscheinlichkeit dafür, dass zur Zeit n eine Erneuerung stattfindet.

Die noch verbleibende Lebensdauer der zur Zeit n aktuellen Komponente, auch *Vorwärtsrekurrenzzeit* genannt, ist

$$X_n := \min\{S_m : S_m \geq n\} - n.$$

Man erhält hierdurch eine Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ mit Zustandsraum \mathbb{N}_0 und Übergangsmatrix

$$P = \begin{pmatrix} f_1 & f_2 & f_3 & f_4 & \cdots \\ 1 & 0 & 0 & 0 & \cdots \\ 0 & 1 & 0 & 0 & \cdots \\ 0 & 0 & 1 & 0 & \cdots \\ \cdots & \cdots & \cdots & \ddots & \cdots \end{pmatrix},$$

also

$$p_{0i} = f_{i+1} \quad \text{für alle } i \in \mathbb{N}_0, \quad p_{i,i-1} = 1 \quad \text{für alle } i \in \mathbb{N}.$$

Offensichtlich gilt

$$u_n = \mathbb{P}(X_n = 0 | X_0 = 0) = p_{00}(n),$$

wodurch der Zusammenhang zur Erneuerungsfolge hergestellt wird. Wir nehmen $\text{ggT}\{i : f_i > 0\} = 1$ an. Die Y -Variablen sind gerade die Rückkehrzeiten nach 0, insbesondere gilt

$$f_i = f_{00}(i) = \mathbb{P}(Y_1 = i) \leq p_{00}(i),$$

der Zustand 0 ist unter dieser Bedingung daher aperiodisch. Wegen

$$f_{00}^* = \sum_{i=1}^{\infty} f_i = \mathbb{P}(Y_1 \in \mathbb{N}) = 1$$

ist 0 auch rekurrent. Im Falle $K := \sup\{i : f_i > 0\} = \infty$ hat man schließlich noch Irreduzibilität; bei $K < \infty$ arbeitet man sinnvoller mit dem Zustandsraum $\{0, 1, \dots, K-1\}$.

Unter welchen Bedingungen an $(f_i)_{i \in \mathbb{N}}$ gibt es zu $(X_n)_{n \in \mathbb{N}}$ (bzw. zu P) eine stationäre Verteilung, und wie sieht diese gegebenenfalls aus? Die Bedingung $\pi P = \pi$ führt auf

$$\pi_i = f_{i+1}\pi_0 + \pi_{i+1} \quad \text{für alle } i \in \mathbb{N}_0.$$

Wenn es eine summierbare Lösung π gibt, so muss

$$\sum_{k=i}^{\infty} \pi_k = \pi_0 \sum_{k=i+1}^{\infty} f_k + \sum_{k=i+1}^{\infty} \pi_k \quad \text{für alle } i \in \mathbb{N}_0$$

also

$$\pi_i = \pi_0 \sum_{k=i+1}^{\infty} f_k \quad \text{für alle } i \in \mathbb{N}_0$$

gelten. Mit

$$\sum_{i=0}^{\infty} \sum_{k=i+1}^{\infty} f_k = \sum_{i=1}^{\infty} \sum_{k=i}^{\infty} f_k = \sum_{k=1}^{\infty} \sum_{i=1}^k f_k = \sum_{k=1}^{\infty} k f_k$$

folgt, dass für die Existenz einer stationären Verteilung die Bedingung

$$\mu := \sum_{k=1}^{\infty} k f_k (= EY_1) < \infty$$

notwendig ist, und dass dann die stationäre Verteilung gegeben wird durch

$$\pi_i := \frac{1}{\mu} \sum_{k=i+1}^{\infty} f_k \quad \left(= \frac{P(Y > i)}{EY} \right) \quad \text{für alle } i \in \mathbb{N}_0.$$

Der Hauptgrenzwertsatz für Markov-Ketten liefert dann wegen $u_n = p_{00}(n)$:

SATZ 1.14 (Diskreter Erneuerungssatz)

Ist $(u_n)_{n \in \mathbb{N}_0}$ die Erneuerungsfolge zu einer aperiodischen Lebensdauerverteilung mit endlichem Erwartungswert μ , so gilt

$$\lim_{n \rightarrow \infty} u_n = \frac{1}{\mu}.$$

In der oben betrachteten Situation führt eine Zerlegung nach dem Wert von Y_1 auf die sogenannte *Erneuerungsgleichung*,

$$u_n = \sum_{m=1}^n f_m u_{n-m} \quad \text{für alle } n \in \mathbb{N},$$

d.h. man kann die u -Werte (beginnend mit $u_0 = 1$) rekursiv aus den f -Werten berechnen. Vergleicht man dies mit der Beziehung

$$p_{ii}(n) = \sum_{m=1}^n f_{ii}(m) p_{ii}(n-m)$$

aus dem Beweis zu Satz 1.6, so sieht man, dass bei einer beliebigen Markov-Kette die Folge $(p_{ii}(n))_{n \in \mathbb{N}}$ der Übergangswahrscheinlichkeiten zu einem festen Zustand als Erneuerungsfolge zur Verteilung $(f_{ii}(n))_{n \in \mathbb{N}}$ der Rückkehrzeit T_i in diesen Zustand aufgefasst werden kann. Der diskrete Erneuerungssatz führt somit auf die folgende Ergänzung zum Hauptgrenzwertsatz für Markov-Ketten:

KOROLLAR 1.15 *Unter den Voraussetzungen von Satz 1.13 gilt*

$$\pi_i = \frac{1}{E(T_i | X_0 = i)} \quad \text{für alle } i \in E.$$

Ein Zustand i mit $\mathbb{P}(T_i < \infty | X_0 = i) = 1$ heißt bekanntlich rekurrent. Gilt darüberhinaus $E(T_i | X_0 = i) < \infty$, so heißt i *positiv rekurrent*, sonst *nullrekurrent*. Das Standardbeispiel für eine nullrekurrente Kette ist die eindimensionale einfache symmetrische Irrfahrt.

Ein interessante Anwendung der obigen Ideen erhält man im Zusammenhang mit Mustern in zufälligen Zeichenketten: Vorgegeben ist eine endliche Folge $A = (a_1, \dots, a_r)$ von Buchstaben aus einem endlichen Alphabet Σ , man interessiert sich für die Wartezeit auf dieses *Muster* in einer zufälligen Zeichenkette.

Konkret sei $(X_n)_{n \in \mathbb{N}}$ eine Folge von unabhängigen und identisch verteilten Zufallsvariablen mit Werten in Σ und

$$N_A := \inf\{n \geq r : X_{n-r+1:n} = A\}$$

die Wartezeit auf das Muster A in $(X_n)_{n \in \mathbb{N}}$, hierbei ist $X_{i:j}$ eine Abkürzung für $(X_i, X_{i+1}, \dots, X_j)$. Es sei

$$q_a := \mathbb{P}(X_i = a) > 0 \quad \text{für alle } a \in \Sigma.$$

Aus früheren Vorlesungen ist bekannt, dass $\mathbb{P}(N_a < \infty) = 1$ gilt ('monkey typing Shakespeare'). Ein klassisches Beispiel sind die sogenannten *success runs*: Wird ein Zufallsexperiment unabhängig wiederholt, wobei das Eintreten eines bestimmten Ereignisses als Erfolg betrachtet wird, so wird man mit

$$0 : \text{Misserfolg (W. } 1 - q), \quad 1 : \text{Erfolg (W. } q),$$

auf die obige Situation mit $\Sigma = \{0, 1\}$, $\mathcal{L}(X_i) = \text{Bin}(1, q)$ geführt. Bei

$$A := \underbrace{(1, 1, \dots, 1)}_{r\text{-mal}}$$

ist N_A der Zeitpunkt, zu dem erstmalig ein solcher Erfolgslauf der Länge r abgeschlossen ist. Bei Anwendungen in der Molekularbiologie hat man beispielsweise $\Sigma = \{A, G, C, T\}$ und das Muster entspricht einem Gen, die Zeichenkette einem DNA-Strang.

Wie erhält man zu vorgegebenen $(q_a)_{a \in \Sigma}$, $A \in \Sigma^r$ die Verteilung (oder die Wahrscheinlichkeitserzeugende Funktion, den Erwartungswert) zu N_A ? Wir setzen

$$p_A(0) := 1, \quad p_A(j) := \prod_{l=r-j+1}^r q_{a_l} \quad \text{für } j = 1, \dots, r, \quad p_A := p_A(r).$$

Insbesondere gilt dann

$$\mathbb{P}(X_{n-r+1:n} = A) = p_A \quad \text{für alle } n \geq r.$$

Das *Überlappungspolynom* ϕ_A zu A (und $(q_a)_{a \in \Sigma}$) definieren wir durch

$$\phi_A(z) := \sum_{j=0}^{r-1} \chi_A(j) p_A(j) z^j,$$

wobei die ‘Überlappungsbits’

$$\chi_A(j) := \begin{cases} 1, & A_{j+1:r} = A_{1:r-j}, \\ 0, & \text{sonst,} \end{cases}$$

anzeigen, ob Präfix und Suffix der Länge $r - j$ des Musters übereinstimmen. Wartet man auf einen success run der Länge r , so gilt $\chi_A(j) = 1$ für $j = 0, \dots, r - 1$, also

$$\phi_A(z) = \sum_{j=0}^{r-1} q^j z^j = \frac{1 - q^r z^r}{1 - qz}.$$

Bei $A = (\text{TOBEORNOTTOBE})$ gilt $\chi_A(j) = 1$ für $j = 0, 9$ und $\chi_A(j) = 0$ sonst. (Schreibt man das Muster und das um 9 Positionen verschobene Muster übereinander, so stimmen die Buchstaben im überlappenden Teil überein.) Es sei

$$g_A(z) := \sum_{k=r}^{\infty} \mathbb{P}(N_A = k) z^k$$

die wahrscheinlichkeitserzeugende Funktion zu N_A .

SATZ 1.16 *Mit den obigen Bezeichnungen gilt*

$$g_A(z) = \frac{p_A z^r}{p_A z^r + (1 - z)\phi_A(z)}.$$

BEWEIS: Wir betrachten die Zeitpunkte des *nicht-überlappenden* Eintretens von A . (Bei success runs der Länge $r = 3$ wären dies in der Zeichenkette

001011011110111001011111100...

(nur) die unterstrichenen Positionen.) Die Differenzen sind offensichtlich unabhängig und haben alle dieselbe Verteilung wie N_A . Es sei $(u_n)_{n \in \mathbb{N}_0}$ die zugehörige Erneuerungsfolge, d.h. u_n ist die Wahrscheinlichkeit dafür, dass zum Zeitpunkt n das Muster A komplett ist, wenn überlappende Vorkommen ignoriert werden. Wie im Beweis zu Satz 1.6 erhält man aus der Erneuerungs-gleichung

$$U(z) := \sum_{n=0}^{\infty} u_n z^n = \frac{1}{1 - g_A(z)}, \quad |z| < 1. \quad (1)$$

Es sei $F_n := \{X_{n-r+1:n} = A\}$ das Ereignis, dass zum Zeitpunkt n ($n \geq r$) das Muster A komplettiert wird. Dieses Ereignis ist entweder eine Erneuerung im

obigen Sinne, oder man hatte eine Erneuerung zum Zeitpunkt $n - k$ für ein $k \in \{1, \dots, r - 1\}$, was nur bei $A_{1:r-k} = A_{k+1:r}$ möglich ist. Dies führt auf

$$\mathbb{P}(F_n) = \sum_{k=0}^{r-1} \chi_A(k) u_{n-k} p_A(k)$$

und damit

$$\sum_{n=r}^{\infty} \mathbb{P}(F_n) z^n = \sum_{k=0}^{r-1} \chi_A(k) p_A(k) z^k \sum_{n=r}^{\infty} u_{n-k} z^{n-k} = \phi_A(z) (U(z) - 1), \quad (2)$$

wobei wir $u_1 = \dots = u_{r-1} = 0$ verwendet haben. Andererseits gilt $\mathbb{P}(F_n) = p_A$ für alle $n \geq r$, also

$$\sum_{n=r}^{\infty} \mathbb{P}(F_n) z^n = \frac{p_A z^r}{1 - z}. \quad (3)$$

Aus (1-3) folgt nun die Behauptung. \square

Ist Z eine Zufallsvariable mit wahrscheinlichkeitserzeugender Funktion g , so gilt bekanntlich $EZ = g'(1)$.

KOROLLAR 1.17 *Es gilt $EN_A = \phi_A(1)/p_A$.*

Bei Erfolgsläufen der Länge r mit Erfolgswahrscheinlichkeit q erhält man

$$g_A(z) = \frac{q^r z^r (1 - qz)}{1 - z + q^r (1 - q) z^{r+1}}, \quad EN_A = \frac{1 - q^r}{q^r (1 - q)}.$$

Wirft man beispielsweise eine faire Münze einmal pro Sekunde, so muss man im Mittel 62 Sekunden warten, bis erstmals fünfmal hintereinander ‘Kopf’ erscheint, bei $d = 10, 15, 20$ ergeben sich als mittlere Wartezeiten 34.1 Minuten, 18.2 Stunden bzw. 24.3 Tage. Wartet man analog beim Würfelwurf auf runs von Sechsen, so ergeben sich bei $d = 5, 10, 15, 20$ die Werte 2.6 Stunden, 28.0 Monate, 18.1 Jahre und schließlich 140.7 Millionen Jahre (fast alle Werte sind gerundet und Band I des Buches von Feller entnommen).

1.5 Markov-Ketten und Martingale. Zu einer Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ betrachten wir die natürliche Filtration

$$(\mathcal{F}_n)_{n \in \mathbb{N}_0}, \quad \mathcal{F}_n := \sigma(\{X_k : k \leq n\}).$$

Es sei $h : E \rightarrow \mathbb{R}$ eine Funktion. Wir werden gelegentlich h mit dem ‘Spaltenvektor’ $(h_i)_{i \in E}$, $h_i := h(i)$ für alle $i \in E$, identifizieren. Die Übergangsmatrix $P = (p_{ij})_{i,j \in E}$ zu X definiert dann durch

$$Ph : E \rightarrow \mathbb{R}, \quad (Ph)(i) := \sum_{j \in E} p_{ij} h_j \quad \text{für alle } i \in E$$

einen linearen Operator auf (einer Teilmenge von) \mathbb{R}^E . Das folgende Lemma gibt diesem Operator eine stochastische Interpretation.

LEMMA 1.18 *Es sei $h : E \rightarrow \mathbb{R}$ eine Funktion mit der Eigenschaft*

$$\sum_{j \in E} p_{ij} |h_j| < \infty \quad \text{für alle } i \in E.$$

Dann gilt

$$(Ph)(X_n) = E[h(X_{n+1}) | \mathcal{F}_n].$$

BEWEIS: Mit der Markov-Eigenschaft erhält man

$$\begin{aligned} & \int_{\{X_0=i_0, \dots, X_n=i_n\}} (Ph)(X_n) d\mathbb{P} \\ &= \sum_{j \in E} p_{i_n j} h_j \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) \\ &= \sum_{j \in E} h_j \mathbb{P}(X_{n+1} = j | X_n = i_n) \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) \\ &= \sum_{j \in E} h_j \mathbb{P}(X_{n+1} = j | X_0 = i_0, \dots, X_n = i_n) \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) \\ &= \sum_{j \in E} h_j \mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = j) \\ &= \sum_{j \in E} \int_{\{X_0=i_0, \dots, X_n=i_n, X_{n+1}=j\}} h(X_{n+1}) d\mathbb{P} \\ &= \int_{\{X_0=i_0, \dots, X_n=i_n\}} h(X_{n+1}) d\mathbb{P}. \end{aligned}$$

Die Mengen der Form

$$\{X_0 = i_0, \dots, X_n = i_n\}, \quad i_0, \dots, i_n \in E,$$

bilden ein durchschnittsstabiles Erzeugendensystem zu \mathcal{F}_n , also folgt hieraus

$$\int_A (Ph)(X_n) d\mathbb{P} = \int_A h(X_{n+1}) d\mathbb{P} \quad \text{für alle } A \in \mathcal{F}_n.$$

Da $(Ph)(X_n)$ \mathcal{F}_n -messbar ist, folgt insgesamt die Behauptung. \square

Die Bedingung in dem Lemma ist offensichtlich für beschränkte Funktionen h erfüllt. Die bewiesene Formel kann wieder so interpretiert werden, dass eine Markov-Kette ‘kein Gedächtnis’ hat: Will man $h(X_{n+1})$ vorhersagen, so ist bei Kenntnis der Werte von X_0, \dots, X_n nur der letzte relevant (hierzu später mehr).

SATZ UND DEFINITION 1.19 *Es seien P eine Übergangsmatrix auf E und $h : E \rightarrow \mathbb{R}$ eine Funktion mit der Eigenschaft*

$$\sum_{j \in E} p_{ij} |h_j| < \infty \quad \text{für alle } i \in E.$$

Gilt dann $Ph = h$, so nennen wir h harmonisch; im Falle $Ph \geq h$ bzw. $Ph \leq h$ heißt h sub- bzw. superharmonisch. Ist $X = (X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Übergangsmatrix P und h harmonisch (subharmonisch, superharmonisch), so ist $(h(X_n), \mathcal{F}_n)_{n \in \mathbb{N}_0}$ ein Martingal (Submartingal, Supermartingal).

BEWEIS: Lemma 1.8 liefert

$$E[h(X_{n+1}) | \mathcal{F}_n] = h(X_n) + (Ph - h)(X_n).$$

Bei subharmonischem h bedeutet dies

$$h(X_n) \leq E[h(X_{n+1}) | \mathcal{F}_n] \quad \text{für alle } n \in \mathbb{N}_0,$$

d.h. $(h(X_n), \mathcal{F}_n)_{n \in \mathbb{N}_0}$ ist ein Submartingal. Der superharmonische Fall kann offensichtlich ganz analog behandelt werden. \square

Ist $h : E \rightarrow \mathbb{R}$ eine beschränkte Funktion, so wird durch

$$\begin{aligned} Z_n^h &:= h(X_n) - \sum_{k=1}^n \left(E[h(X_k) | \mathcal{F}_{k-1}] - h(X_{k-1}) \right) \\ &= h(X_n) - \sum_{k=1}^n ((P - I)(h))(X_{k-1}) \end{aligned}$$

ein Martingal $(Z_n^h, \mathcal{F}_n)_{n \in \mathbb{N}_0}$ definiert:

$$\begin{aligned} E[Z_{n+1}^h | \mathcal{F}_n] &= E[h(X_{n+1}) | \mathcal{F}_n] - \sum_{k=1}^{n+1} E[E[h(X_k) | \mathcal{F}_{k-1}] - h(X_{k-1}) | \mathcal{F}_n] \\ &= h(X_n) - \sum_{k=1}^n \left(E[h(X_k) | \mathcal{F}_{k-1}] - h(X_{k-1}) \right) \\ &= Z_n^h. \end{aligned}$$

Man nennt diese die *Lévy-Martingale* zu $X = (X_n)_{n \in \mathbb{N}_0}$. Man erhält hierdurch eine Zerlegung

$$h(X) = M + A$$

von $h(X) = (h(X_n))_{n \in \mathbb{N}_0}$ in ein Martingal $M = Z^h$ und einen Prozess $A = (A_n)_{n \in \mathbb{N}}$,

$$A_n := \sum_{k=1}^n ((P - I)(h))(X_{k-1}),$$

der in dem Sinne *vorhersehbar* ist, dass für alle $n \in \mathbb{N}$ die Zufallsgröße A_n messbar ist bezüglich \mathcal{F}_{n-1} . Eine zweite wichtige Beobachtung in diesem Zusammenhang ist, dass die Markov-Eigenschaft sich über die Lévy-Martingale charakterisieren lässt:

SATZ 1.20 *Es sei $(X_n)_{n \in \mathbb{N}_0}$ ein stochastischer Prozess mit Werten in E und natürlicher Filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$, P sei eine stochastische Matrix auf E . Ist dann für alle beschränkten $h : E \rightarrow \mathbb{R}$ $(Z_n^h, \mathcal{F}_n)_{n \in \mathbb{N}_0}$ mit*

$$Z_n^h := h(X_n) - \sum_{k=1}^n ((P - I)(h))(X_{k-1})$$

ein Martingal, so ist X eine homogene Markov-Kette mit Übergangsmatrix P .

BEWEIS: Es seien $i_0, \dots, i_n, i_{n+1} \in E$, $A := \{X_0 = i_0, \dots, X_n = i_n\}$ und $h := 1_{\{i_{n+1}\}}$. Aus $E[Z_{n+1}^h | \mathcal{F}_n] = Z_n^h$ erhält man

$$E[h(X_{n+1}) | \mathcal{F}_n] = (Ph)(X_n)$$

(man beachte, dass dies nicht aus Lemma 1.8 folgt, da wir ja (noch) nicht wissen, dass X eine Markov-Kette ist), also

$$\int_A h(X_{n+1}) d\mathbb{P} = \int_A (Ph)(X_n) d\mathbb{P}.$$

Die linke Seite ist

$$\int_A h(X_{n+1}) d\mathbb{P} = \mathbb{P}(X_0 = i_0, \dots, X_n = i_n, X_{n+1} = i_{n+1}),$$

die rechte

$$\begin{aligned} \int_A (Ph)(X_n) d\mathbb{P} &= \int_{\{X_0 = i_0, \dots, X_n = i_n\}} (Ph)(i_n) d\mathbb{P} \\ &= p_{i_n i_{n+1}} \mathbb{P}(X_0 = i_0, \dots, X_n = i_n), \end{aligned}$$

also folgt $\mathbb{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = p_{i_n i_{n+1}}$. □

Durch diese Verbindung können Resultate der Martingalthorie (Ungleichungen, Optional Stopping, Konvergenzsätze) für das Gebiet Markov-Ketten nutzbar gemacht werden.

BEISPIEL 1.21 Es sei $(X_n)_{n \in \mathbb{N}_0}$ die einfache symmetrische Irrfahrt mit Start in $i_0 \in \mathbb{Z}$, also $\mathbb{P}(X_0 = i_0) = 1$ und

$$\mathbb{P}(X_{n+1} = j + 1 | X_n = j) = \mathbb{P}(X_{n+1} = j - 1 | X_n = j) = \frac{1}{2}.$$

Für die identische Abbildung $h : \mathbb{Z} \rightarrow \mathbb{Z}$, $j \mapsto j$, gilt offensichtlich

$$(Ph)(j) = \frac{1}{2}h(j-1) + \frac{1}{2}h(j+1) = h(j),$$

also ist h eine harmonische Funktion. Wegen $|X_n| \leq |i_0| + n$ existiert auch der Erwartungswert zu $h(X_n)$ ($= X_n$) für alle $n \in \mathbb{N}_0$, also ist $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}_0}$ ein Martingal (diese Aussage ergibt sich auch als Spezialfall von Beispiel 8.2 (i) der Vorlesung Stochastik II). Es seien nun $a, b \in \mathbb{Z}$ mit $a < i_0 < b$ und

$$\tau := \inf\{n \in \mathbb{N} : X_n = a \text{ oder } X_n = b\}$$

der erste Zeitpunkt, zu dem der Prozess das Intervall (a, b) verlässt. Nach dem OST ist dann $(X_{\tau \wedge n}, \mathcal{F}_{\tau \wedge n})_{n \in \mathbb{N}_0}$ wieder ein Martingal, das aufgrund seiner Beschränktheit (es gilt $|X_{\tau \wedge n}| \leq |a| \vee |b|$) nach dem Vorwärtskonvergenzsatz von Doob fast sicher gegen eine Zufallsvariable X_∞ konvergiert. Auf $\{\tau = \infty\}$ gilt

$$|X_{\tau \wedge (n+1)} - X_{\tau \wedge n}| = |X_{n+1} - X_n| = 1,$$

also folgt aus dieser Konvergenz $\mathbb{P}(\tau < \infty | X_0 = i_0) = 1$. In Worten: Jedes endliche Intervall wird irgendwann verlassen. Hieraus wiederum folgt, dass X_∞ fast sicher mit X_τ übereinstimmt. Die Beschränktheit des Martingals liefert zusätzlich die L^1 -Konvergenz (Satz von der majorisierten Konvergenz) und somit

$$EX_\tau = \lim_{n \rightarrow \infty} EX_{\tau \wedge n} = EX_0 = i_0.$$

Setzt man

$$\tau_a := \inf\{n \in \mathbb{N} : X_n = a\}, \quad \tau_b := \inf\{n \in \mathbb{N} : X_n = b\},$$

so hat man $\tau = \tau_a \wedge \tau_b$ und $\mathbb{P}(\tau < \infty | X_0 = i_0) = 1$ impliziert $\mathbb{P}(\tau_a = \tau_b | X_0 = i_0) = 0$. Aus $EX_\tau = i_0$ folgt somit

$$a \mathbb{P}(\tau_a < \tau_b | X_0 = i_0) + b (1 - \mathbb{P}(\tau_a < \tau_b | X_0 = i_0)) = i_0,$$

also

$$\mathbb{P}(\tau_a < \tau_b | X_0 = i_0) = \frac{b - i_0}{b - a}, \quad \mathbb{P}(\tau_b < \tau_a | X_0 = i_0) = \frac{i_0 - a}{b - a}.$$

Bei dem in der Vorlesung Stochastik I besprochenen Ruinproblem (Beispiel 4.27) geht es (u.a.) darum, mit welcher Wahrscheinlichkeit ein Spieler bei Anfangskapital n das Kapital N ($> n$) erreicht, bevor der Ruin (Kapital 0) eintritt. Wird in jeder Runde eine faire Münze geworfen, so wird diese Wahrscheinlichkeit zu $\mathbb{P}(\tau_b < \tau_a | X_0 = i_0)$ mit $i_0 = n$, $b = N$ und $a = 0$. Man erhält das dort mit analytischen Hilfsmitteln erhaltenen Resultat (und zusätzlich eine Antwort auf die damals nicht behandelte Frage, ob das Spiel überhaupt aufhört).

Mit der obigen Formel kann auch die Rekurrenz von $(X_n)_{n \in \mathbb{N}_0}$ nachgewiesen werden; auf analytischem Wege ist dies bereits in Beispiel 1.7 geschehen. Wir setzen $\eta := \mathbb{P}(\tau_0 = \infty | X_0 = 1)$. Eine Zerlegung nach dem Wert von X_1 zusammen mit einer Symmetrieüberlegung zeigt, dass

$$\eta = 0 \implies \mathbb{P}(\tau_0 < \infty | X_0 = 0) = 1$$

gilt. Ist also der Zustand 0 nicht rekurrent, so muss $\eta > 0$ gelten. Oben wurde gezeigt, dass man für alle $b > 1$

$$\mathbb{P}(\tau_b < \tau_0 | X_0 = 1) = \frac{1}{b}$$

hat. Wählt man b so groß, dass $\eta > 1/b$ gilt, so folgt hieraus

$$\begin{aligned} \mathbb{P}(\tau_0 \wedge \tau_b = \infty | X_0 = 1) &\geq \mathbb{P}(\tau_0 = \infty | X_0 = 1) - \mathbb{P}(\tau_b < \tau_0 | X_0 = 1) \\ &= \eta - \frac{1}{b} > 0, \end{aligned}$$

im Widerspruch zu $\mathbb{P}(\tau_0 \wedge \tau_b < \infty | X_0 = 1) = 1$.

Unser zweites Beispiel stammt aus den Bereichen Vielteilchensysteme (Physik) und zellulare Automaten (Informatik).

BEISPIEL 1.22 Es seien $L, N \in \mathbb{N}$, $L, N \geq 2$, $I := \{1, \dots, N\} \times \{1, \dots, N\}$ und $E := \{1, \dots, L\}^I$ die Menge aller Abbildungen $f : I \rightarrow \{1, \dots, L\}$. Interpretation: $f(i, j) = l$ heißt, dass der ‘Gitterpunkt’ (i, j) die ‘Farbe’ l hat. Bei $N = 4$, $L = 2$ (mit den Farben Schwarz und Weiß) wären also

$$\begin{array}{cccc} \bullet & \circ & \bullet & \circ \\ \circ & \circ & \bullet & \bullet \\ \bullet & \circ & \circ & \circ \\ \circ & \bullet & \bullet & \circ \end{array} \quad \text{und} \quad \begin{array}{cccc} \circ & \circ & \bullet & \bullet \\ \bullet & \bullet & \circ & \circ \\ \circ & \bullet & \bullet & \circ \\ \bullet & \circ & \circ & \bullet \end{array}$$

mögliche Zustände, hier auch *Konfigurationen* genannt. Auf E betrachten wir den folgenden Übergangsmechanismus: Es wird ein Punkt $A \in I$ zufällig und gleichverteilt ausgewählt, dann ein Nachbar B von A , wieder zufällig und gleichverteilt aus den vier Möglichkeiten. Nachbarn werden hierbei modulo N bestimmt, d.h. beispielsweise bei $N = 4$, dass der Punkt $(1, 1)$ die Nachbarn $(1, 2)$, $(2, 1)$, $(4, 1)$ und $(1, 4)$ hat. Die neue Konfiguration entsteht nun aus der alten, indem man $f(A)$ durch $f(B)$ ersetzt (die Farbe eines zufällig gewählten Punktes wird durch die eines seiner zufällig gewählten Nachbarn ersetzt). Es sei $(X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit diesem Übergangsmechanismus. Klar: Die ‘monochromatischen Zustände’, also die $f \in E$ mit $f \equiv l$ für ein $l \in \{1, \dots, L\}$, sind absorbierend. Wie verhält sich diese Kette mit $n \rightarrow \infty$?

Sei $l \in \{1, \dots, L\}$ fest, Y_n die Anzahl der Punkte mit Farbe l in der zufälligen Konfiguration X_n . Ist (A, B) ein Paar benachbarter Punkte in I , so ist die Wahrscheinlichkeit dafür, dass man im $(n+1)$ -ten Schritt erst A und dann B gleich der Wahrscheinlichkeit dafür, dass man erst B und dann A als farbbestimmenden Nachbarn wählt. Klar:

$$\begin{aligned} X_n(A) = X_n(B) &\implies Y_{n+1} = Y_n, \\ X_n(A) = l, X_n(B) \neq l &\implies Y_{n+1} = Y_n - 1, \\ X_n(A) \neq l, X_n(B) = l &\implies Y_{n+1} = Y_n + 1. \end{aligned}$$

Die obige Symmetrieüberlegung zeigt nun, dass $E[Y_{n+1} | X_0, \dots, X_n] = Y_n$ gilt, d.h. $(Y_n)_{n \in \mathbb{N}_0}$ ist ein nichtnegatives Martingal. Nach dem Vorwärtskonvergenzsatz von Doob (Satz 8.18 der Vorlesung Stochastik II) konvergiert dieses \mathbb{P} -f.s. gegen eine Zufallsvariable Y_∞ . Nun ist $(Y_n(\omega))_{n \in \mathbb{N}_0}$ für jedes $\omega \in \Omega$ eine Folge von ganzen Zahlen, muss bei Konvergenz also ab einem $n_0 = n_0(\omega)$ konstant sein. Als Konstanten kommen nur 0 und N^2 in Frage. Um dies einzusehen, argumentiert man wie folgt: Für alle $k \in \{1, \dots, N^2 - 1\}$ gilt

$$\mathbb{P}(Y_{n+j} = k | Y_n = \dots = Y_{n+j-1} = k) \leq 1 - \frac{1}{4N^2},$$

denn es gibt dann mindestens ein Nachbarschaftspaar (A, B) , dessen Wahl auf $|Y_{n+1} - Y_n| = 1$ führen würde. Mit der Multiplikationsregel für bedingte Wahrscheinlichkeiten folgt hieraus

$$\mathbb{P}(Y_n = Y_{n+1} = \dots = Y_{n+j} = k) \leq \left(1 - \frac{1}{4N^2}\right)^j,$$

also $\mathbb{P}(A_n) = 0$ für das Ereignis A_n , dass Y_m für alle $m \geq n$ den Wert k annimmt. Nun gilt

$$\{\omega \in \Omega : \exists n \forall m \geq n : Y_m(\omega) = k\} = \bigcup_{n=0}^{\infty} A_n,$$

und es folgt

$$\mathbb{P}(\lim_{n \rightarrow \infty} Y_n = k) = \mathbb{P}(\exists n \forall m \geq n : Y_m = k) = 0.$$

Man hat also $\mathbb{P}(Y_\infty \in \{0, N^2\}) = 1$. Das Martingal ist offensichtlich beschränkt und damit auch L^1 -konvergent, d.h. es gilt $EY_\infty = \lim_{n \rightarrow \infty} EY_n = EY_0$. Damit erhalten wir insgesamt:

$$\begin{aligned} \mathbb{P}(\text{Absorption in } f \equiv l) &= \mathbb{P}(Y_\infty = N^2) = \frac{1}{N^2} EY_\infty = \frac{1}{N^2} EY_0 \\ &= \frac{1}{N^2} \#\{A \in I : X_0(A) = l\}. \end{aligned}$$

In Worten: Die Wahrscheinlichkeit dafür, dass eine bestimmte Farbe sich durchsetzt, ist gleich dem Anteil dieser Farbe an der Ausgangskonfiguration (Konzentration von Farben auf zusammenhängende Bereiche etc. spielt also keine Rolle).

Martingalargumente lassen sich häufig auch in solchen Situationen zur Beschreibung des qualitativen Verhaltens von Markov-Ketten verwenden, in denen das explizite Ausrechnen der relevanten Größen nicht möglich ist.

1.6 Die starke Markov-Eigenschaft. Ist $X = (X_t)_{t \in T}$ ein stochastischer Prozess auf $(\Omega, \mathcal{A}, \mathbb{P})$ mit Werten in (E, \mathcal{B}) , so nennt man die Abbildungen

$$T \rightarrow E, \quad t \mapsto X_t(\omega)$$

($\omega \in \Omega$) die *Pfade* des Prozesses. Man kann dann X als Abbildung von Ω in E^T auffassen, jedem $\omega \in \Omega$ wird der zugehörige Pfad zugeordnet. Versieht man E^T (oder eine geeignete Teilmenge hiervon) mit einer geeigneten σ -Algebra, so wird X zu einer messbaren Abbildung, also zu einer Zufallsgröße, deren Werte Funktionen sind. Bei abzählbarem E und $T = \mathbb{N}_0$ ist dies bei $\mathcal{P}(E)^{\otimes \mathbb{N}_0}$ der Fall (es gab hierzu eine Aufgabe in der Vorlesung Stochastik II), einen zugehörigen durchschnittsstabilen Erzeuger liefern die Mengen

$$Z(i_0, i_1, \dots, i_n) := \{i_0\} \times \{i_1\} \times \dots \times \{i_n\} \times E \times E \times \dots,$$

$n \in \mathbb{N}_0, i_0, \dots, i_n \in E$. Ist $X = (X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Startverteilung q und Übergangsmatrix P , so gilt

$$\mathbb{P}(X \in Z(i_0, i_1, \dots, i_n)) = \mathbb{P}(X_0 = i_0, \dots, X_n = i_n) = q_{i_0} \prod_{k=0}^{n-1} p_{i_k, i_{k+1}},$$

die Verteilung von X ist also durch q und P festgelegt. Die Abhängigkeit von q wird oft durch die Notation \mathbb{P}_q (oder \mathbb{P}_i , wenn X_0 auf ein $i \in E$ konzentriert ist) angedeutet. Wir behandeln nun eine Variante der Markov-Eigenschaft, die durch diese Sichtweise von stochastischen Prozessen als Zufallsgrößen mit Werten in Pfadräumen nahegelegt wird.

Es sei $X = (X_n)_{n \in \mathbb{N}_0}$ eine Markov-Kette mit Übergangsmatrix P und natürlicher Filtration $(\mathcal{F}_n)_{n \in \mathbb{N}_0}$, τ sei eine \mathbb{P} -fast sicher endliche Stoppzeit zu dieser Filtration. Wir nennen

$$X^\tau = (X_n^\tau)_{n \in \mathbb{N}_0}, \quad X_n^\tau := X_{\tau \wedge n} \quad \text{für alle } n \in \mathbb{N}_0$$

den bei τ eingefrorenen (gestoppten) Prozess,

$$Y = (Y_n)_{n \in \mathbb{N}_0}, \quad Y_n := X_{\tau+n} \quad \text{für alle } n \in \mathbb{N}_0$$

(auf $\{\tau = \infty\}$ sei Y ‘irgendwie’ definiert) den *Post- τ -Prozess*.

Wir betrachten zunächst den Fall $\tau \equiv N$ für ein $N \in \mathbb{N}_0$. Aus Aufgabe 9 ist bekannt, dass dann \mathcal{F}_N und $\sigma(Y)$ unter X_N bedingt unabhängig sind: Für alle $A \in \mathcal{F}_N$, $B \in \sigma(Y)$ und $i \in E$ gilt

$$\mathbb{P}(A \cap B | X_N = i) = \mathbb{P}(A | X_N = i) \mathbb{P}(B | X_N = i).$$

Dies kann man umschreiben zu $\mathbb{P}(B | X_N = i, A) = \mathbb{P}(B | X_N = i)$, wobei wir wie immer voraussetzen, dass die beteiligten bedingten Wahrscheinlichkeiten definiert sind. Insbesondere gilt also

$$\begin{aligned} \mathbb{P}(Y_{n+1} | Y_0 = i_0, \dots, Y_n = i_n) &= \mathbb{P}(X_{N+n+1} = i_{n+1} | X_N = i_0, \dots, X_{N+n} = i_n) \\ &= \mathbb{P}(X_{N+n+1} = i_{n+1} | X_{N+n} = i_n) \\ &= p_{i_n, i_{n+1}}, \end{aligned}$$

wobei wir in der zweitletzten Gleichung die obige Aussage mit $N + n$ anstelle von n verwendet haben. Dies zeigt, dass Y wieder eine Markov-Kette ist, mit derselben Übergangsmatrix wie die Ausgangskette und natürlich mit Startverteilung $\mathcal{L}(Y_0) = \mathcal{L}(X_N)$.

SATZ 1.23 (Starke Markov-Eigenschaft)

Mit den oben eingeführten Bezeichnungen gilt:

- (i) Y ist eine Markov-Kette mit Übergangsmatrix P und Startverteilung $\mathcal{L}(X_\tau)$,
- (ii) \mathcal{F}_τ und $\sigma(Y)$ sind unter X_τ bedingt unabhängig.

BEWEIS: (i) Wir zerlegen nach dem Wert von τ :

$$\begin{aligned}
& \mathbb{P}(Y_{n+1} = i_{n+1}, Y_0 = i_0, \dots, Y_n = i_n) \\
&= \sum_{l=0}^{\infty} \mathbb{P}(X_{l+n+1} = i_{n+1}, X_l = i_0, \dots, X_{l+n} = i_n, \tau = l) \\
&= \sum_{l=0}^{\infty} \mathbb{P}(X_{l+n+1} = i_{n+1} | X_l = i_0, \dots, X_{l+n} = i_n, \tau = l) \\
&\quad \mathbb{P}(X_l = i_0, \dots, X_{l+n} = i_n, \tau = l) \\
&= p_{i_n, i_{n+1}} \sum_{l=0}^{\infty} \mathbb{P}(X_l = i_0, \dots, X_{l+n} = i_n, \tau = l) \\
&= p_{i_n, i_{n+1}} \mathbb{P}(X_\tau = i_0, \dots, X_{\tau+n} = i_n) \\
&= p_{i_n, i_{n+1}} \mathbb{P}(Y_0 = i_0, \dots, Y_n = i_n),
\end{aligned}$$

also

$$\begin{aligned}
& \mathbb{P}(Y_{n+1} = i_{n+1} | Y_0 = i_0, \dots, Y_n = i_n) \\
&= \frac{\mathbb{P}(Y_{n+1} = i_{n+1}, Y_0 = i_0, \dots, Y_n = i_n)}{\mathbb{P}(Y_0 = i_0, \dots, Y_n = i_n)} \\
&= p_{i_n, i_{n+1}}.
\end{aligned}$$

(ii) Wir zerlegen wieder nach dem Wert von τ : Mit $A := Z(i_0, \dots, i_m)$, $B := Z(j_0, \dots, j_n)$ ergibt sich

$$\begin{aligned}
& \mathbb{P}(X^\tau \in A, Y \in B, X_\tau = k) \\
&= \sum_{l=0}^{\infty} \mathbb{P}(X_{0 \wedge l} = i_0, \dots, X_{m \wedge l} = i_m, X_l = j_0, \dots, X_{l+n} = j_n, X_l = k, \tau = l) \\
&= \sum_{l=0}^{\infty} \mathbb{P}(X_l = j_0, \dots, X_{l+n} = j_n | X_{0 \wedge l} = i_0, \dots, X_{m \wedge l} = i_m, X_l = k, \tau = l) \\
&\quad \mathbb{P}(X_{0 \wedge l} = i_0, \dots, X_{m \wedge l} = i_m, X_l = k, \tau = l) \\
&= \mathbb{P}(X_0 = j_0, \dots, X_n = j_n | X_0 = k) \mathbb{P}(X^\tau \in A, X_\tau = k) \\
&= \mathbb{P}(Y \in B | X_\tau = k) \mathbb{P}(X^\tau \in A | X_\tau = k) \mathbb{P}(X_\tau = k),
\end{aligned}$$

wobei wir in der zweitletzten Gleichung Teil (i) verwendet haben. Dividiere nun durch $\mathbb{P}(X_\tau = k)$. \square

BEMERKUNG 1.24 (Existenz, kanonische Version) Zu Beginn dieses Abschnitts haben wir gesehen, dass durch q und P die Verteilung von X als

Zufallsgröße mit Werten in $(E^{\mathbb{N}_0}, \mathcal{P}(E)^{\otimes \mathbb{N}_0})$ eindeutig festgelegt ist. Zum Abschluss dieses Unterabschnitts besprechen wir die Existenz solcher Prozesse, wobei wir auch kurz auf allgemeine Zustandsräume eingehen. Es sei also (E, \mathcal{B}) ein beliebiger messbarer Raum. Anstelle einer Übergangsmatrix P haben wir nun eine Übergangswahrscheinlichkeit P von und nach (E, \mathcal{B}) , im diskreten Fall ist $P(i, \{j\}) = p_{ij}$. Der Prozess $X = (X_n)_{n \in \mathbb{N}_0}$ heißt homogene Markov-Kette mit Übergangswahrscheinlichkeit P und Startverteilung μ , wenn $\mathcal{L}(X_0) = \mu$ gilt und $P(X_n, \cdot)$ eine Version der bedingten Verteilung von X_{n+1} unter $\mathcal{F}_n = \sigma\{X_0, \dots, X_n\}$ ist, d.h.

$$\mathbb{P}(X_{n+1} \in B, A) = \int_A P(X_n, B) d\mathbb{P} \quad \text{für alle } B \in \mathcal{B}, A \in \mathcal{F}_n.$$

Mit den in Abschnitt 7.1 der Vorlesung Stochastik II eingeführten Bezeichnungen gilt dann

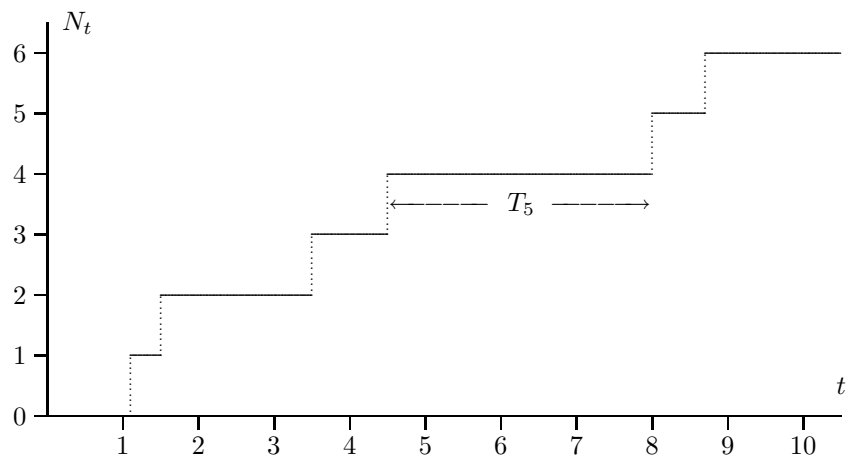
$$\mathcal{L}(X_0, \dots, X_n) = \mu \otimes \underbrace{P \otimes P \otimes \dots \otimes P}_{n\text{-mal}}.$$

Betrachtet man dies als Wahrscheinlichkeitsmaß auf $(E^{n+1}, \mathcal{B}^{\otimes(n+1)})$, so erhält man die Existenz 'für endliche n '. Der Satz von Ionescu-Tulcea (siehe beispielsweise Abschnitt 5.1 in J. Neveu, *Mathematische Grundlagen der Wahrscheinlichkeitstheorie*) besagt, dass es auf $(E^{\mathbb{N}_0}, \mathcal{B}^{\otimes \mathbb{N}_0})$ ein Wahrscheinlichkeitsmaß \mathbb{P} gibt, bei dem als Bild unter der Projektion auf die ersten $n+1$ Koordinaten jeweils dieses Maß erscheint, für alle $n \in \mathbb{N}_0$. Dieses Resultat kann als Verallgemeinerung von Satz 3.9 der Vorlesung Stochastik II betrachtet werden. Auf $(E^{\mathbb{N}_0}, \mathcal{B}^{\otimes \mathbb{N}_0}, \mathbb{P})$ ist nun $X = (X_n)_{n \in \mathbb{N}_0}$, wobei X_n die Projektion auf die n -te Koordinate bezeichnet, eine Markov-Kette mit Startverteilung μ und Übergangswahrscheinlichkeit P . Man bezeichnet diese Konstruktion als den *kanonischen Prozess* zu μ und P .

2. Markov-Ketten in stetiger Zeit

2.1 Ein wichtiger Spezialfall: Der Poisson-Prozess. In diesem Abschnitt geht es um den nach der Brownschen Bewegung wohl wichtigsten stochastischen Prozess. Wir werden dabei auch erstmalig mit einer überabzählbaren Zeitmenge zu tun haben, wodurch sich einige technische Probleme ergeben. Insbesondere werden Aussagen zu den Pfaden der Prozesse wichtig. Aus Gründen der Übersichtlichkeit schreiben wir gelegentlich $X(t)$ oder $X(t, \omega)$ anstelle von X_t oder $X_t(\omega)$.

Wir wollen ein Resultat herleiten, das zeigt, dass Poisson-Prozesse sich als Konsequenz bestimmter allgemeiner und in Anwendungen häufig zumindest näherungsweise erfüllter Annahmen ergeben. Ausgangspunkt ist eine Situation, in der Ereignisse wie beispielsweise die Emission eines Partikels beim radioaktiven Zerfall oder die Ankunft eines Kunden in einem Bedienungssystem zu bestimmten zufälligen Zeitpunkten eintreten. Wir nehmen grundsätzlich an, dass mindestens ein Ereignis eintritt und dass die Anzahl der Ereignisse in jedem kompakten Zeitintervall endlich ist. Schreibt man N_t für die Anzahl der Ereignisse im Intervall $(0, t]$, so erhält man einen sogenannten *Zählprozess* $N = (N_t)_{t \geq 0}$. Das untenstehende Diagramm zeigt den Pfad eines solchen Prozesses. Man sieht, dass hier im betrachteten Zeitintervall die Ereignisse einzeln eingetreten sind; T_k bezeichnet die Zeitspanne zwischen dem $(k - 1)$ -ten und dem k -ten Ereignis.



Der Pfad eines Zählprozesses

Wir betrachten zunächst den Raum der Pfade des Prozesses und definieren hierzu

$$D_0 := \{f : [0, \infty) \rightarrow \mathbb{Z} : f(0) = 0, f \uparrow, f \text{ stetig von rechts}\}.$$

Unsere erste Annahme betrifft die Pfade von N und ergibt sich zwangsläufig aus den obigen Annahmen an das zugrundeliegende Geschehen.

(A1) Alle Pfade von N liegen in D_0 .

Auf D_0 betrachten wir die von den Projektionen

$$\pi_t : D_0 \rightarrow \mathbb{Z}, \quad f \mapsto f(t),$$

erzeugte σ -Algebra $\mathcal{B}(D_0) := \sigma(\{\pi_t : t \geq 0\})$. Eine Abbildung mit Werten in D_0 ist genau dann $\mathcal{B}(D_0)$ -messbar, wenn ihre Verknüpfungen mit π_t , $t \geq 0$, messbar sind. Da wir voraussetzen, dass alle N_t , $t \geq 0$, Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ sind, ist also N als Abbildung von Ω nach D_0 ($\mathcal{A}, \mathcal{B}(D_0)$)-messbar, d.h. der Zählprozess kann als Zufallsgröße mit Werten in einem Funktionenraum oder kurz als zufällige Funktion aufgefasst werden. Insbesondere ist die Verteilung von N ein Wahrscheinlichkeitsmaß auf dem messbaren Raum $(D_0, \mathcal{B}(D_0))$.

Die nächsten beiden Annahmen besagen, dass der Prozess unabhängige und stationäre Zuwächse hat.

(A2) Für beliebige $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$ sind die Zufallsvariablen

$$N_{t_0}, N_{t_1} - N_{t_0}, \dots, N_{t_n} - N_{t_{n-1}}$$

stochastisch unabhängig.

(A3) Bei beliebigem $t > 0$ hängt die Verteilung von $N_{s+t} - N_s$ nicht von s ab.

Bevor wir die letzte Annahme formulieren, wollen wir (A2) und (A3) ein wenig 'abstraktifizieren'. Man überprüft leicht, dass für jedes $u \geq 0$ die Abbildungen

$$S_u : D_0 \rightarrow D_0, \quad f \mapsto f(\cdot \wedge u), \quad Z_u : D_0 \rightarrow D_0, \quad f \mapsto f(u + \cdot) - f(u)$$

$(\mathcal{B}(D_0), \mathcal{B}(D_0))$ -messbar sind. Insbesondere können also auch die stochastischen Prozesse $S_u(N) = (N_{t \wedge u})_{t \geq 0}$ und $Z_u(N) = (N_{u+t} - N_u)_{t \geq 0}$ als Zufallsgrößen mit Werten in $(D_0, \mathcal{B}(D_0))$ aufgefasst werden; $S_u(N)$ ist der zum Zeitpunkt u gestoppte oder 'eingefrorene' Prozess, $Z_u(N)$ entsteht aus N , indem man erst zum Zeitpunkt u mit dem Zählen anfängt. Die Mengen

$$A_{t_1, \dots, t_n}^{i_1, \dots, i_n} := \{f \in D_0 : f(t_j) - f(t_{j-1}) = i_j \text{ für } j = 1, \dots, n\},$$

$0 =: t_0 \leq t_1 \leq \dots \leq t_n < \infty$, $i_1, \dots, i_n \in \mathbb{N}_0$, bilden ein durchschnittstabiles Erzeugendensystem von $\mathcal{B}(D_0)$. Nun lässt sich ein Ereignis der Form

$$\left\{ S_u(N) \in A_{t_1, \dots, t_k}^{i_1, \dots, i_k} \right\} \cap \left\{ Z_u(N) \in A_{s_1, \dots, s_l}^{j_1, \dots, j_l} \right\}$$

auf Bedingungen

$$\begin{aligned} N_{t_1 \wedge u} - N_{t_0 \wedge u} &= i_1, \dots, N_{t_k \wedge u} - N_{t_{k-1} \wedge u} = i_k, \\ N_{u+s_1} - N_u &= j_1, \dots, N_{u+s_l} - N_{u+s_{l-1}} = j_l \end{aligned}$$

an die Zuwächse umformulieren, deren Unabhängigkeit liefert also

$$\begin{aligned} \mathbb{P}(S_u(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, Z_u(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ = \mathbb{P}(S_u(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}) \cdot \mathbb{P}(Z_u(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \end{aligned}$$

und somit die Unabhängigkeit von $S_u(N)$ und $Z_u(N)$. Analog erhält man mit der Annahme (A3)

$$\mathbb{P}(Z_u(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) = \mathbb{P}(N \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}),$$

$Z_u(N)$ und N haben also dieselbe Verteilung. Im folgenden Lemma werden beide Aussagen auf Stoppzeiten verallgemeinert, wobei wir die natürliche Filtration

$$(\mathcal{F}_t)_{t \geq 0}, \quad \mathcal{F}_t := \sigma(\{N_s : s \leq t\})$$

verwenden. Wir nennen

$$S_\tau(N) = (N_{\tau \wedge t})_{t \geq 0}, \quad Z_\tau(N) = (N_{\tau+t} - N_\tau)_{t \geq 0}$$

auch den *Prä- τ* - und den (normalisierten) *Post- τ -Prozess*; die zugehörigen Messbarkeitsfragen werden in den Übungen behandelt.

LEMMA 2.1 *Ist τ eine endliche Stoppzeit, so sind die Prozesse $(N_{\tau \wedge t})_{t \geq 0}$ und $(N_{\tau+t} - N_\tau)_{t \geq 0}$ stochastisch unabhängig. Außerdem hat $(N_{\tau+t} - N_\tau)_{t \geq 0}$ dieselbe Verteilung wie $(N_t)_{t \geq 0}$.*

BEWEIS: Wir nehmen zunächst an, dass τ einen abzählbaren Wertebereich hat. Eine Zerlegung nach dem Wert von τ ergibt dann für beliebige Zeitpunkte $0 \leq s_1 \leq \dots \leq s_k$, $0 \leq t_1 \leq \dots \leq t_l$ und beliebige $i_1, \dots, i_k, j_1, \dots, j_l \in \mathbb{N}_0$

$$\begin{aligned} \mathbb{P}(S_\tau(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, Z_\tau(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ = \sum_u \mathbb{P}(S_u(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, \tau = u, Z_u(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}). \end{aligned}$$

Da τ eine Stoppzeit ist, gilt $\{\tau = u\} \in \mathcal{F}_u$, also erhält man mit der Unabhängigkeit von $\mathcal{F}_u = \sigma(S_u(N))$ und $\sigma(Z_u(N))$ und der Verteilungsgleichheit von N und $Z_u(N)$

$$\begin{aligned} & \mathbb{P}(S_\tau(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, Z_\tau(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ &= \sum_u \mathbb{P}(S_u(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, \tau = u) \cdot \mathbb{P}(Z_u(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ &= \mathbb{P}(N \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \cdot \sum_u \mathbb{P}(S_u(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, \tau = u) \\ &= \mathbb{P}(N \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \cdot \mathbb{P}(S_\tau(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}). \end{aligned}$$

Im Spezialfall $k = 1, s_1 = 0, i_1 = 0$ wird dies zu

$$\mathbb{P}(Z_\tau(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) = \mathbb{P}(N \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}),$$

$Z_\tau(N)$ und N haben also dieselbe Verteilung. Setzt man dies oben ein, so folgt

$$\begin{aligned} & \mathbb{P}(S_\tau(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, Z_\tau(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ &= \mathbb{P}(S_\tau(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}) \cdot \mathbb{P}(Z_\tau(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \end{aligned}$$

und somit die Unabhängigkeit von $S_\tau(N)$ und $Z_\tau(N)$.

Bei beliebigem (endlichen) τ erhält man durch $\tau_n := \lceil n\tau \rceil / n$ eine Folge von Stoppzeiten mit abzählbarem Wertebereich, die ‘von rechts’ gegen τ konvergiert. Mit der Annahme (A1) folgt, dass für alle $t \geq 0$

$$N_{\tau_n \wedge t} \rightarrow N_{\tau \wedge t}, \quad N_{\tau_n + t} \rightarrow N_{\tau + t} \quad \mathbb{P}\text{-fast sicher}$$

mit $n \rightarrow \infty$ gilt. Bei ganzzahligen Zufallsvariablen impliziert dies, dass die Variablen mit gegen 1 konvergierender Wahrscheinlichkeit gleich sind, d.h. man hat

$$\lim_{n \rightarrow \infty} \mathbb{P}(N_{\tau_n \wedge t} = N_{\tau \wedge t}) = 1, \quad \lim_{n \rightarrow \infty} \mathbb{P}(N_{\tau_n + t} = N_{\tau + t}) = 1$$

für alle $t \geq 0$, und damit

$$\begin{aligned} & \mathbb{P}(S_\tau(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, Z_\tau(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(S_{\tau_n}(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}, Z_{\tau_n}(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(S_{\tau_n}(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}) \cdot \mathbb{P}(Z_{\tau_n}(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}) \\ &= \mathbb{P}(S_\tau(N) \in A_{s_1, \dots, s_k}^{i_1, \dots, i_k}) \cdot \mathbb{P}(Z_\tau(N) \in A_{t_1, \dots, t_l}^{j_1, \dots, j_l}). \end{aligned}$$

Damit ist die Unabhängigkeit von $S_\tau(N)$ und $Z_\tau(N)$ für beliebige endliche Stoppzeiten gezeigt; die Gleichheit der Verteilungen von N und $Z_\tau(N)$ kann ganz analog übertragen werden. \square

Dieses Lemma zeigt, dass ein den obigen Bedingungen genügender Prozess auch zu zufälligen Zeiten ‘von vorn anfängt und sein Gedächtnis verliert’, wenn nur diese Zeit nicht auf die zukünftige Entwicklung vorgreift.

Die letzte Annahme schließlich wird sich im Zusammenhang mit dem simultanen Auftreten von mehr als einem Ereignis als wichtig erweisen: Wir werden sehen, dass sie zusammen mit den anderen Annahmen dazu führt, dass mit Wahrscheinlichkeit 1 die zu zählenden Ereignisse *einzel*n auftreten.

$$(A4) \quad \mathbb{P}(N_h \geq 2) = o(h) \text{ mit } h \downarrow 0.$$

Hier nun ist das Hauptresultat dieses Abschnitts.

SATZ 2.2 *Es sei $N = (N_t)_{t \geq 0}$ ein stochastischer Prozess, der den Bedingungen (A1)-(A4) genügt. Dann hat N mit Wahrscheinlichkeit 1 nur Sprünge der Höhe 1 und es existiert ein $\lambda > 0$ derart, dass gilt:*

- (a) *Für alle $s, t \geq 0$ ist $N_{s+t} - N_s$ Poisson-verteilt mit Parameter $\lambda \cdot t$.*
- (b) *Die Zeiten zwischen den aufeinanderfolgenden Sprüngen des Prozesses sind unabhängig und mit Parameter λ exponentialverteilt.*

BEWEIS: Die erste Aussage lässt sich auch in der Form

$$\mathbb{P}(N_s - N_{s-} \geq 2 \text{ für ein } s > 0) = 0$$

schreiben, wobei N_{s-} den (bei einer isotonen Funktion stets existierenden) linksseitigen Limes in s bezeichnet. Für festes $t > 0$ erhält man mit (A3)

$$\begin{aligned} \mathbb{P}(N_s - N_{s-} \geq 2 \text{ für ein } s \in (0, t]) \\ &\leq \mathbb{P}(N_{kt/n} - N_{(k-1)t/n} \geq 2 \text{ für ein } k \in \{1, \dots, n\}) \\ &\leq n \mathbb{P}(N_{t/n} \geq 2). \end{aligned}$$

Annahme (A4) impliziert, dass diese Oberschranke mit $n \rightarrow \infty$ gegen 0 geht. Mit $t \rightarrow \infty$ ergibt sich

$$\begin{aligned} \mathbb{P}(N_s - N_{s-} \geq 2 \text{ für ein } s > 0) \\ &= \lim_{t \rightarrow \infty} \mathbb{P}(N_s - N_{s-} \geq 2 \text{ für ein } s \in (0, t]) = 0. \end{aligned}$$

Zur Herleitung der Verteilungsaussage (a) betrachten wir zunächst die Funktion $\phi, \phi(t) := \mathbb{P}(N_t = 0)$. Mit (A2) und (A3) erhält man für alle $s, t > 0$

$$\phi(s+t) = \mathbb{P}(N_s = 0, N_{s+t} - N_s = 0) = \mathbb{P}(N_s = 0) \mathbb{P}(N_t = 0) = \phi(s)\phi(t),$$

also insbesondere $\phi(k/n) = \phi(1)^{k/n}$ für alle $k, n \in \mathbb{N}$. Da ϕ schwach monoton fallend ist, folgt mit einem einfachen Einschachtelungsargument, dass sogar

$$\phi(t) = \phi(1)^t \quad \text{für alle } t \geq 0$$

gilt. Als Wahrscheinlichkeit muss $\phi(1)$ im Intervall $[0, 1]$ liegen. Im Falle $\phi(1) = 1$ hätte man

$$\mathbb{P}(N_t = 0 \text{ für alle } t > 0) = \lim_{t \rightarrow \infty} \phi(t) = 1,$$

im Widerspruch zu der generellen Annahme, dass mindestens ein Ereignis eintritt. Im Falle $\phi(1) = 0$ hätte man

$$\begin{aligned} \mathbb{P}(N_1 \geq n) &\geq \mathbb{P}(N_{k/n} - N_{(k-1)/n} \geq 1 \text{ für } k = 1, \dots, n) \\ &= (1 - \phi(1/n))^n = 1 \end{aligned}$$

für alle $n \in \mathbb{N}$, also $\mathbb{P}(N_1 = \infty) = 1$ im Widerspruch zu der generellen Annahme, dass in kompakten Intervallen nur endlich viele der zu zählenden Ereignisse eintreten. Sei $\lambda := -\log \phi(1)$. Wir haben $0 < \lambda < \infty$ gezeigt, und dass $\mathbb{P}(N_t = 0) = e^{-\lambda t}$ gilt für alle $t \geq 0$.

Sei nun $t > 0$. Jedes $n \in \mathbb{N}$ definiert eine Zerlegung

$$N_t = \sum_{k=1}^n (N_{kt/n} - N_{(k-1)t/n})$$

von N_t in eine Summe von Zufallsvariablen, die gemäß (A2) unabhängig sind. Die Zufallsvariable

$$X_{nk} := 1_{\mathbb{N}}(N_{kt/n} - N_{(k-1)t/n})$$

zeigt an, ob im Zeitintervall $(kt/n, (k-1)t/n]$ ein Ereignis eintritt. Als Funktion der unabhängigen Zuwächse sind die Variablen X_{n1}, \dots, X_{nn} unabhängig, nach dem obenstehenden Argument sind sie binomialverteilt mit Parametern 1 und $1 - \exp(-\lambda t/n)$. Für $\sum_{k=1}^n X_{nk}$ erhält man somit nach dem Gesetz der seltenen Ereignisse (Satz 4.4 der Vorlesung Stochastik I) im Limes mit $n \rightarrow \infty$ eine Poisson-Verteilung mit Parameter

$$\lim_{n \rightarrow \infty} n(1 - e^{-\lambda t/n}) = \lambda t.$$

Die Annahme (A4) impliziert

$$\begin{aligned} \mathbb{P}\left(\sum_{k=1}^n X_{nk} \neq N_t\right) &= \mathbb{P}\left(\bigcup_{k=1}^n \{N_{kt/n} - N_{(k-1)t/n} \geq 2\}\right) \\ &\leq n \mathbb{P}(N_{t/n} \geq 2) \rightarrow 0 \end{aligned}$$

mit $n \rightarrow \infty$, also ist N_t Poisson-verteilt mit Parameter λt .

Zum Nachweis von (b) verwenden wir eine Induktion. Bezeichnet T_1 den Zeitpunkt des ersten Sprungs, so folgt mit (a)

$$\mathbb{P}(T_1 > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t} \quad \text{für alle } t \geq 0,$$

also ist T_1 in der Tat exponentialverteilt mit Parameter λ . Für $k \geq 2$ sei T_k die Zeit zwischen dem $(k-1)$ -ten und dem k -ten Sprung. Angenommen, wir haben bereits gezeigt, dass T_1, \dots, T_k unabhängig und mit Parameter λ exponentialverteilt sind. Offensichtlich ist $\tau := T_1 + \dots + T_k$ eine Stoppzeit. Wegen

$$\mathbb{P}(\tau > t) = \mathbb{P}(N_t < k) \rightarrow 0 \quad \text{mit } t \rightarrow \infty$$

ist τ mit Wahrscheinlichkeit 1 endlich, wir können also Lemma 2.1 anwenden. Man sieht leicht, dass T_{k+1} die Zeit bis zum ersten Sprung im Post- τ -Prozess ist. Da dieser dieselbe Verteilung wie N hat, ist nach der für den Induktionsanfang ausgeführten Rechnung also auch T_{k+1} mit Parameter λ exponentialverteilt. Der Post- τ -Prozess ist außerdem unabhängig vom Prä- τ -Prozess $S_\tau(N)$. Da die Variablen T_1, \dots, T_k Funktionen von $S_\tau(N)$ sind, erhält man somit auch die Unabhängigkeit von T_{k+1} und $\{T_1, \dots, T_k\}$. \square

Man nennt die durch diesen Satz charakterisierten Prozesse *Poisson-Prozesse*, λ ist die *Rate* oder auch *Intensität* des Prozesses. Bemerkenswert ist hier unter anderem, dass die qualitativen Annahmen (A1)-(A4) eine konkrete parametrische Gestalt für die Verteilungen der Zuwächse und die Verteilungen der Zeiten zwischen den Ereignissen erzwingen.

BEISPIEL 2.3 (Das Inspektions-Paradox)

Der in Abschnitt 1.4 im diskreten Fall eingeführte erneuerungstheoretische Rahmen lässt sich leicht auf die Zeitparametermenge $[0, \infty)$ übertragen: Ist X_k , $k \in \mathbb{N}$, die Lebensdauer der k -ten Glühbirne, $S_n := \sum_{k=1}^n X_k$ der Zeitpunkt der n -ten Erneuerung, so zählt $N_t := \max\{n \in \mathbb{N} : S_n \leq t\}$ die im Zeitintervall $(0, t]$ stattfindenden Erneuerungen. Satz 2.2 zeigt, dass $N = (N_t)_{t \geq 0}$ ein Poisson-Prozess mit Intensität λ ist, wenn die Lebensdauern exponentialverteilt sind mit Parameter λ . In der Erneuerungstheorie interessiert man sich unter anderem für die Variablen

$$V_t := S_{N_t+1} - t, \quad W_t := t - S_{N_t}, \quad L_t := W_t + V_t,$$

die Restlebensdauer, das Alter und die Gesamtlebensdauer der Komponente, die zum Zeitpunkt t im Gebrauch ist. Die Variable V_t ist der Zeitpunkt des ersten Sprunges des Post- t -Prozesses, also wieder exponentialverteilt mit Parameter λ (dass sich für die Restlebensdauer dieselbe Verteilung ergibt wie für die

X -Variablen selbst, ist letztlich eine Konsequenz der bereits aus der Vorlesung Stochastik I bekannten ‘Gedächtnislosigkeit’ der Exponentialverteilungen). Für $0 \leq s \leq t$ gilt

$$\mathbb{P}(W_t \geq s) = \mathbb{P}(N_t - N_{t-s} = 0) = e^{-\lambda s}.$$

Das zum Zeitpunkt t maximal mögliche Alter ist t , also gilt $\mathbb{P}(W_t > s) = 0$ für $s > t$. Als Verteilungsfunktion von W_t ergibt sich somit

$$F_{W_t}(s) = \begin{cases} 1 - e^{-\lambda s} & \text{für } 0 \leq s < t, \\ 1 & \text{für } s \geq t, \end{cases}$$

eine bei t ‘abgeschnittene’ Exponentialverteilung. Als Funktionen der unabhängigen Prozesse $S_t(N)$ und $Z_t(N)$ sind V_t und W_t unabhängig, die Verteilung von L_t ist also als Faltung der bereits bestimmten Verteilungen von V_t und W_t . Als zugehörige Dichtefunktion f_{L_t} erhält man bei $s > t$

$$\begin{aligned} f_{L_t}(s) &= \int f_{V_t}(s-u) F_{W_t}(du) \\ &= \int_0^t \lambda e^{-\lambda(s-u)} \lambda e^{-\lambda u} du + e^{-\lambda t} \lambda e^{-\lambda(s-t)} \\ &= \lambda(1 + \lambda t)e^{-\lambda s}, \end{aligned}$$

und bei $s \leq t$

$$f_{L_t}(s) = \int_0^s \lambda e^{-\lambda(s-u)} \lambda e^{-\lambda u} du = \lambda^2 s e^{-\lambda s}.$$

Aus diesen Rechnungen ergibt sich beispielsweise, dass bei großem t der Erwartungswert der Lebensdauer der der z.Z. t aktuellen (‘inspizierten’) Komponente etwa doppelt so groß ist wie die mittlere Lebensdauer $EX_i = 1/\lambda$ der Komponenten selbst. Ganz allgemein ist bei beliebigen Lebensdauerverteilungen μ die Verteilung der (Gesamt-)Lebensdauer der zu einer vorgegebenen festen Zeit inspizierten Komponente von μ verschieden. Auf den ersten Blick erscheint dies paradox; auf den zweiten Blick leuchtet ein, dass ein vorgegebener Zeitpunkt mit größerer Wahrscheinlichkeit von einem überdurchschnittlich langen zufälligen Intervall überdeckt wird.

BEISPIEL 2.4 Es seien K, X_1, X_2, \dots unabhängige Zufallsvariablen; K sei Poisson-verteilt mit Parameter λT , alle X_i seien gleichverteilt auf dem Intervall $(0, T)$. Schließlich sei $N = (N_t)_{t \geq 0}$ der Zählprozess zu den Variablen X_1, \dots, X_K , einer Stichprobe aus $\text{unif}(0, T)$ vom zufälligen Umfang K , also

$$N_t := \#\{1 \leq i \leq K : X_i \leq t\} \quad \text{für alle } t \geq 0.$$

Es sei $0 =: t_0 \leq t_1 \leq \dots \leq t_n = T$ eine Zerlegung des Intervalles $(0, T]$. Bei beliebigen $i_1, \dots, i_n \in \mathbb{N}_0$ führt das Ereignis

$$\{N_{t_1} - N_{t_0} = i_1, \dots, N_{t_n} - N_{t_{n-1}} = i_n\}$$

auf die Bedingungen, dass K den Wert $k := i_1 + \dots + i_n$ annimmt, und dass für $l = 1, \dots, n$ jeweils i_l von den insgesamt k auf $(0, T)$ gleichverteilten X -Werten in ein ‘Fach’ der relativen Länge $(t_l - t_{l-1})/T$ fallen. Mit der aus der Vorlesung Stochastik I bekannten Multinomialverteilung erhält man

$$\begin{aligned} \mathbb{P}(N_{t_1} - N_{t_0} = i_1, \dots, N_{t_n} - N_{t_{n-1}} = i_n) &= \binom{k}{i_1 \dots i_n} \left(\frac{t_1 - t_0}{T}\right)^{i_1} \cdot \dots \cdot \left(\frac{t_n - t_{n-1}}{T}\right)^{i_n} e^{-\lambda T} \frac{(\lambda T)^k}{k!} \\ &= \prod_{l=1}^n e^{-\lambda(t_l - t_{l-1})} \frac{(\lambda(t_l - t_{l-1}))^{i_l}}{i_l!}, \end{aligned}$$

die Zuwächse des Prozesses $\tilde{N} := (N_t)_{0 \leq t \leq T}$ sind also unabhängig und Poissonverteilt, wobei der Parameter nur von der Länge des zu dem Zuwachs gehörenden Zeitintervalls abhängt. Hieraus folgt, dass \tilde{N} dieselbe Verteilung hat wie ein zur Zeit T gestoppter Poisson-Prozess mit Intensität λ . Diese Beobachtung kann beispielsweise bei der Simulation von Poisson-Prozessen benutzt werden. Umgekehrt besagen diese Rechnungen auch etwas über Poisson-Prozesse: Die Sprungzeitpunkte, also die Zeitpunkte, zu denen in den Anwendungsbeispielen eine Emission eintritt oder ein Kunde ankommt, sind bei gegebener Gesamtzahl der Sprünge bzw. Ereignisse in einem Intervall unabhängig und auf diesem Intervall gleichverteilt.

2.2 Ein Blick auf den allgemeinen Fall. Es sei wieder E eine abzählbare Menge. In diesem Abschnitt betrachten wir Markov-Prozesse $(X_t)_{t \in T}$ mit ‘stetigem’ oder ‘kontinuierlichem’ Zeitparameter, wobei wir in der Regel mit $T = \mathbb{R}_+ = [0, \infty)$ arbeiten. Die bereits im diskreten Fall angedeutete Komplementarität von analytischer und stochastischer Sichtweise wird sich hier fortsetzen und vertiefen.

Bei der Definition der Markov-Eigenschaft und der zeitlichen Homogenität eines Prozesses $(X_t)_{t \geq 0}$ können wir exakt wie in Abschnitt 1.1 verfahren: Der Prozess hat die Markov-Eigenschaft, wenn für alle $i_0, \dots, i_{n+1} \in E$ und alle $t_0, \dots, t_{n+1} \in T$ mit $t_0 < \dots < t_{n+1}$

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_0} = i_0, \dots, X_{t_n} = i_n) = \mathbb{P}(X_{t_{n+1}} = i_{n+1} | X_{t_n} = i_n)$$

gilt; er ist zeitlich homogen, wenn $\mathbb{P}(X_{t+s} = j | X_s = i)$ nicht von s abhängt. Man kann die Übergangswahrscheinlichkeiten

$$p_{ij}(t) := \mathbb{P}(X_t = j | X_0 = i), \quad i, j \in E,$$

wieder zu einer Übergangsmatrix $P(t) = (p_{ij}(t))_{i,j \in E}$ zusammenfassen, und eine Zerlegung nach dem Zustand zur Zeit t liefert wie im Falle diskreter Zeit die Chapman-Kolmogorov-Gleichungen,

$$p_{ij}(t+s) = \sum_{k \in E} p_{ik}(t) p_{kj}(s) \quad \text{für alle } i, j \in E, s, t > 0,$$

also in Matrixschreibweise $P(s+t) = P(s)P(t)$. Schreiben wir $P(0)$ für die Einheitsmatrix, so sehen wir, dass $\{P(t) : t \geq 0\}$ hierdurch zu einer Halbgruppe von stochastischen Matrizen wird. (Die Schreibweise als Menge ist eigentlich unschön, aber wohl-etabliert; gemeint ist natürlich die Funktion $t \mapsto P(t)$.)

An dieser Stelle tritt nun ein fundamentaler Unterschied zur Situation mit $T = \mathbb{N}_0$ zutage—man hat keinen kleinsten Zeitschritt mehr. Bei diskretem Zeitparameter konnte die gesamte Halbgruppe durch die Übergangsmatrix $P = P(1)$ beschrieben werden, alle Übergangsmatrizen ergaben sich als Potenz dieser einen Matrix. Im Falle $T = [0, \infty)$ ist als Konsequenz der Chapman-Kolmogorov-Gleichungen die gesamte Halbgruppe $\{P(t) : t \geq 0\}$ bereits durch das Teilstück $\{P(t) : 0 \leq t \leq \delta\}$ festgelegt, und zwar bei beliebig kleinem $\delta > 0$. Es ist daher nicht überraschend, dass die Ein-Schritt-Übergangsmatrix bei stetiger Zeit durch eine Größe ersetzt wird, die das lokale Verhalten der Halbgruppe mit $t \downarrow 0$ beschreibt.

Bevor wir dies präzisieren, betrachten wir ein konkretes Beispiel.

BEISPIEL 2.5 Ist $N = (N_t)_{t \geq 0}$ ein Poisson-Prozess mit Intensität λ , so gilt

$$\begin{aligned} \mathbb{P}(N_{t_{n+1}} = i_{n+1} \mid N_{t_0} = i_0, \dots, N_{t_n} = i_n) \\ &= \mathbb{P}(N_{t_{n+1}} - N_{t_n} = i_{n+1} - i_n) \\ &= e^{-\lambda(t_{n+1} - t_n)} \frac{(\lambda(t_{n+1} - t_n))^{i_{n+1} - i_n}}{(i_{n+1} - i_n)!} \end{aligned}$$

für $i_{n+1} \geq i_n$, also ist N eine zeitlich homogene Markov-Kette mit den Übergangsfunktionen

$$p_{ij}(t) = \begin{cases} e^{-\lambda t} (\lambda t)^{j-i} / (j-i)! & \text{falls } j \geq i, \\ 0 & \text{sonst.} \end{cases}$$

Mit $p_{ij}(0) = \delta_{ij}$ erhält man für das lokale Verhalten in $t = 0$ die Aussage

$$\lim_{t \downarrow 0} \frac{1}{t} (p_{ij}(t) - p_{ij}(0)) = \begin{cases} \lambda, & \text{falls } j = i + 1, \\ -\lambda, & \text{falls } j = i, \\ 0, & \text{sonst.} \end{cases}$$

Man kann zeigen, dass ganz allgemein unter recht milden Bedingungen bei einer Markov-Kette mit Übergangsfunktionen $p_{ij}(t)$ die rechtsseitigen Ableitungen

$$g_{ij} := \lim_{t \downarrow 0} \frac{1}{t} (p_{ij}(t) - p_{ij}(0))$$

existieren; man nennt die Matrix $G := (g_{ij})_{i,j \in E}$ den *infinitesimalen Erzeuger* oder auch *Generator* zur Übergangshalbgruppe $\{P(t) : t \geq 0\}$. Aus der Definition folgt unmittelbar, dass jede solche Generatormatrix die Eigenschaften

$$g_{ij} \geq 0 \text{ für alle } i, j \in E \text{ mit } i \neq j, \quad g_{ii} \leq 0 \text{ für alle } i \in E \quad (\text{GE1})$$

hat. Darf man Summe und Ableitung vertauschen und existieren die beteiligten Größen, so folgt aus der Tatsache, dass $P(t)$ für jedes $t \geq 0$ Zeilensummen 1 hat, die Beziehung

$$\sum_{j \in E} g_{ij} = 0 \text{ für alle } i \in E, \quad (\text{GE2})$$

die Generatormatrix hat dann also Zeilensummen 0.

Hinter dem Übergang von $\{P(t) : t \geq 0\}$ zu G steckt natürlich die Idee, dass die gesamte Halbgruppe aus dem Generator durch eine Beziehung der Form

$$P(t) = \exp(tG) := \sum_{k=0}^{\infty} \frac{t^k}{k!} G^k \text{ für alle } t \geq 0$$

entsteht, G also die Rolle der Ein-Schritt-Übergangsmatrix aus der diskreten Situation übernimmt. Es ist jedoch zunächst nicht klar, welchen Sinn die Exponentialreihe hier, insbesondere bei unendlichem Zustandsraum, haben soll. Wir betrachten den Fall eines endlichen Zustandsraums etwas näher. Für die Matrixnorm

$$\|G\|_{\infty} := \max_{i,j \in E} |g_{ij}|$$

hat man die Ungleichung

$$\|G^k\|_{\infty} \leq (\#E)^{k-1} \|G\|_{\infty}^k \text{ für alle } k \in \mathbb{N},$$

die Exponentialreihe konvergiert dann also komponentenweise. In der Regel ist die direkte Berechnung von $\exp(tG)$ nicht ohne weiteres möglich. Durch die Analogie zur gewöhnlichen Exponentialfunktion werden die Differentialgleichungen

$$P'(t) = P(t)G, \quad P'(t) = GP(t) \text{ für alle } t \geq 0$$

nahegelegt. In der Tat erhält man mit den Chapman-Kolmogorov-Gleichungen bei endlichem Zustandsraum

$$\begin{aligned} \frac{1}{h}(p_{ij}(t+h) - p_{ij}(t)) &= \frac{1}{h} \left(\sum_{k \in E} p_{ik}(t) p_{kj}(h) - p_{ij}(t) \right) \\ &= p_{ij}(t) \frac{1}{h}(p_{jj}(h) - 1) + \sum_{k \in E, k \neq j} p_{ik}(t) \frac{1}{h} p_{kj}(h) \\ &\rightarrow \sum_{k \in E} p_{ik}(t) g_{kj} \quad \text{mit } t \downarrow 0, \end{aligned}$$

woraus die erste der obigen Differentialgleichungen folgt; die zweite erhält man ganz analog durch eine Zerlegung nach dem Zustand zum Zeitpunkt h . Diese sogenannten *Vorwärts- und Rückwärtsgleichungen von Kolmogorov* gelten bei entsprechenden Regularitätsbedingungen auch bei (abzählbar) unendlichem Zustandsraum; sie erlauben gelegentlich die Herleitung der Übergangsfunktionen zu einer vorgegebenen Generatormatrix G .

BEISPIEL 2.6 Wir betrachten den einfachsten nicht-trivialen Fall, $\#E = 2$, und können $E = \{0, 1\}$ annehmen. Es sei $(X_t)_{t \geq 0}$ eine Markov-Kette mit Start in 0 und Übergangsraten λ für den Übergang von 0 auf 1, μ für den Übergang von 1 auf 0, wobei $\lambda > 0$ und $\mu > 0$ gelte. Der zugehörige Generator ist dann

$$G = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}.$$

Löst man die Vorwärtsgleichungen $P'(t) = P(t)G$ in ihre Komponenten auf, so ergeben sich die folgenden einfachen Differentialgleichungen:

$$\begin{aligned} p'_{00}(t) &= -\lambda p_{00}(t) + \mu p_{01}(t), \\ p'_{10}(t) &= -\lambda p_{10}(t) + \mu p_{11}(t), \\ p'_{01}(t) &= \lambda p_{00}(t) - \mu p_{01}(t), \\ p'_{11}(t) &= \lambda p_{10}(t) - \mu p_{11}(t). \end{aligned}$$

Da es nur zwei mögliche Zustände gibt, muss $p_{01}(t) = 1 - p_{00}(t)$ gelten, also folgt aus der ersten dieser Gleichungen

$$p'_{00}(t) = -(\lambda + \mu) p_{00}(t) + \mu.$$

Unter Beachtung der Bedingung $p_{00}(0) = 1$ erhält man

$$p_{00}(t) = \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} + \frac{\mu}{\lambda + \mu}$$

und analog

$$p_{11}(t) = \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t} + \frac{\lambda}{\lambda + \mu},$$

also insgesamt

$$P(t) = \begin{pmatrix} \frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} + \frac{\mu}{\lambda + \mu} & -\frac{\lambda}{\lambda + \mu} e^{-(\lambda + \mu)t} + \frac{\lambda}{\lambda + \mu} \\ -\frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t} + \frac{\mu}{\lambda + \mu} & \frac{\mu}{\lambda + \mu} e^{-(\lambda + \mu)t} + \frac{\lambda}{\lambda + \mu} \end{pmatrix}.$$

Dieser ‘analytische’ Zugang sagt wenig darüber aus, wie die Markov-Kette $X = (X_t)_{t \geq 0}$ ‘aussieht’. Will man beispielsweise die Verteilung der Zeitspanne untersuchen, die X in einem Zustand bleibt, so sind, wie bereits in Abschnitt 2 erklärt, zusätzliche Annahmen an die Pfade von X nötig. Wir nehmen im folgenden an, dass diese Pfade rechtsstetige Treppenfunktionen sind. Um dies zu präzisieren, betrachten wir auf dem Zustandsraum E die diskrete Topologie (alle Teilmengen sind offen; $i_n \rightarrow i$ ist äquivalent zur Existenz eines n_0 mit $i_n = i$ für alle $n \geq n_0$) und setzen

$$D(E) := \{f : [0, \infty) \rightarrow E : f(t+) = f(t) \text{ für alle } t \geq 0, \\ f(t-) \text{ existiert für alle } t > 0\},$$

wobei wie üblich $f(t+)$ und $f(t-)$ den rechts- bzw. linksseitigen Limes von f im Punkt t bezeichnen. Funktionen aus $D(E)$ sind stückweise konstant und haben in beschränkten Intervallen nur endlich viele Unstetigkeitsstellen (Sprünge). Angenommen, man hat eine Generatormatrix $G = (g_{ij})_{i,j \in E}$ mit den Eigenschaften (GE1) und (GE2). Existiert dann eine Markov-Kette $X = (X_t)_{t \geq 0}$ mit diesen Übergangsraten, also

$$\mathbb{P}(X_{t+h} = j | X_t = i) = \delta_{ij} + g_{ij}h + o(h) \quad \text{mit } h \downarrow 0 \text{ für alle } i, j \in E$$

und Pfaden aus $D(E)$? Der folgende Satz gibt eine Teilantwort.

SATZ 2.7 Für die Matrix G gelte zusätzlich zu (GE1) und (GE2) noch

$$0 < \lambda := \sup_{i \in E} |g_{ii}| < \infty.$$

Es seien $N = (N_t)_{t \geq 0}$ ein Poisson-Prozess mit Intensität λ und $Y = (Y_n)_{n \in \mathbb{N}_0}$ eine von N unabhängige Markov-Kette mit Start in $i_0 \in E$ und Übergangsmatrix $\tilde{P} = (\tilde{p}_{ij})_{i,j \in E}$,

$$\tilde{P} := I + \frac{1}{\lambda} G,$$

also

$$\tilde{p}_{ij} = \delta_{ij} + \frac{g_{ij}}{\lambda} \quad \text{für alle } i, j \in E.$$

Dann ist $X = (X_t)_{t \geq 0}$ mit

$$X_t := Y_{N_t} \quad \text{für alle } t \geq 0$$

eine Markov-Kette mit Start in i_0 , Pfaden aus $D(E)$ und Generatormatrix G .

BEWEIS: Dass \tilde{P} tatsächlich eine stochastische Matrix ist, dürfte ebenso wie die Aussagen zum Startverhalten und zu den Pfaden unmittelbar einsichtig sein. Für beliebige $0 \leq t_0 < t_1 < \dots < t_{n+1}$, $i_1, \dots, i_{n+1} \in E$ gilt

$$\begin{aligned} & \mathbb{P}(X_{t_m} = i_m, 0 \leq m \leq n+1) \\ &= \sum_{k_0, \dots, k_{n+1} \in \mathbb{N}_0} \mathbb{P}(N_{t_m} = k_m, Y_{k_m} = i_m, 0 \leq m \leq n+1) \\ &= \sum_{k_0, \dots, k_{n+1} \in \mathbb{N}_0} \mathbb{P}(N_{t_m} = k_m, 0 \leq m \leq n+1) \mathbb{P}(Y_{k_m} = i_m, 0 \leq m \leq n+1). \end{aligned}$$

Verwendet man bei N die Unabhängigkeit und Stationarität der Zuwächse und bei Y die Markov-Eigenschaft, so erhält man

$$\begin{aligned} & \mathbb{P}(N_{t_m} = k_m, 0 \leq m \leq n+1) \\ &= \mathbb{P}(N_{t_{n+1}-t_n} = k_{n+1} - k_n) \mathbb{P}(N_{t_m} = k_m, 0 \leq m \leq n), \\ & \mathbb{P}(Y_{k_m} = i_m, 0 \leq m \leq n+1) = \tilde{p}_{i_n i_{n+1}}(k_{n+1} - k_n) \mathbb{P}(Y_{k_m} = i_m, 0 \leq m \leq n). \end{aligned}$$

Oben eingesetzt ergibt sich mit $l := k_{n+1} - k_n$

$$\begin{aligned} & \mathbb{P}(X_{t_m} = i_m, 0 \leq m \leq n+1) \\ &= \left(\sum_{l=0}^{\infty} \tilde{p}_{i_n i_{n+1}}(l) \mathbb{P}(N_{t_{n+1}-t_n} = l) \right) \mathbb{P}(X_{t_m} = i_m, 0 \leq m \leq n). \end{aligned}$$

Hieraus folgt, dass X die Markov-Eigenschaft hat und zeitlich homogen ist, und dass die Übergangsfunktionen die Gestalt

$$p_{ij}(t) = \sum_{l=0}^{\infty} \tilde{p}_{ij}(l) e^{-\lambda t} \frac{(\lambda t)^l}{l!}$$

haben. Mit dieser expliziten Form der Übergangsfunktionen (und dem bekannten Verhalten der Exponentialfunktion nahe 0) kann nun auch überprüft werden, dass sich als Übergangsraten gerade die Elemente von G ergeben. \square

Der Satz zeigt, wie man bei vorgegebener Matrix G unter Verwendung einer geeigneten Markov-Kette mit diskretem Zeitparameter und eines Poisson-Prozesses mit einer passenden Intensität eine Markov-Kette in kontinuierlicher Zeit konstruiert, für die G die Generatormatrix ist; der Poisson-Prozess spielt hierbei die Rolle einer ‘stochastischen Uhr’. Allerdings wurde hierbei zusätzlich zu den Bedingungen (GE1-2) vorausgesetzt, dass zu den Diagonalelementen der Generatormatrix eine endliche Schranke existiert.

Was besagt der Satz darüber, wie Markov-Ketten in stetiger Zeit aussehen, und was gibt er her, wenn es um die stochastische Bedeutung der Elemente der Generatormatrix geht? Wie im Beweis zu Satz 2.2 zeigt man, dass Aufenthaltszeiten von Markov-Ketten mit stetigem Zeitparameter exponentialverteilt sind, und zwar nach Definition von G mit Parameter $-g_{ii}$ im Zustand i . Die Diagonalelemente von G bestimmen also die Verteilungen der Aufenthaltszeiten in den verschiedenen Zuständen. Wohin springt die Markov-Kette, wenn ihr Aufenthalt im Zustand i beendet ist? Hierzu betrachten wir zu der Y -Kette aus Satz 2.7 die Folge $Z = (Z_n)_{n \in \mathbb{N}_0}$ der nacheinander besuchten Zustände. Zur Präzisierung definieren wir die Folge $(\tau_n)_{n \in \mathbb{N}_0}$ induktiv durch

$$\tau_0 = 0, \quad \tau_{n+1} := \min\{k > \tau_n : Y_k \neq Y_{\tau_n}\}$$

und setzen $Z_n := Y_{\tau_n}$ für alle $n \in \mathbb{N}_0$. Mit einer Satz 1.23 entsprechenden starken Markov-Eigenschaft kann man zeigen, dass Z wieder eine Markov-Kette ist, und als Übergangswahrscheinlichkeiten erhält man für $i \neq j$ die Werte

$$\hat{p}_{ij} = -\frac{g_{ij}}{g_{ii}},$$

die Off-Diagonalelemente der Generatormatrix regeln also die Verteilung des Sprungziels. Wendet man diese Überlegungen auf die in Beispiel 2.6 besprochene Markov-Kette mit den zwei Zuständen 0 und 1, Start in 0 und Generatormatrix $G = \begin{pmatrix} -\lambda & \lambda \\ \mu & -\mu \end{pmatrix}$ an, so sieht man, dass dieser Prozess die Zustände 0 und 1 abwechselnd besucht, wobei die Aufenthaltszeiten mit Parameter λ und μ (für 0 bzw. 1) exponentialverteilt sind.

Man kann natürlich auch ohne Zuhilfenahme eines Poisson-Prozesses direkt aus den Aufenthaltszeiten und der eingebetteten Sprungkette eine Markov-Kette in stetiger Zeit konstruieren. Ein hierbei auftretendes fundamentales Problem, das in Satz 2.7 durch die Beschränktheitsannahme vermieden wurde, illustrieren wir am nachfolgenden Beispiel.

BEISPIEL 2.8 Es sei $E = \mathbb{N}_0$, $G = (g_{ij})_{i,j \in E}$ mit

$$g_{ij} = \begin{cases} -(i+1)^2, & \text{falls } j = i, \\ (i+1)^2, & \text{falls } j = i+1, \\ 0, & \text{sonst.} \end{cases}$$

Wir starten in 0. Die Folge der nacheinander zu durchlaufenden Zustände wäre dann, wie beim Poisson-Prozess, die Folge der nicht-negativen ganzen Zahlen. Da die Aufenthaltszeiten T_i im Zustand i exponentialverteilt sind mit Parameter $(i+1)^2$, erhält man für die ‘Gesamtlebensdauer’ $T := \sum_{i=0}^{\infty} T_i$ einer Markov-Kette mit diesem Generator

$$ET = \sum_{i=0}^{\infty} ET_i = \sum_{i=1}^{\infty} \frac{1}{i^2} = \frac{\pi^2}{6} < \infty.$$

Die Pfade eines solchen Prozesses liegen nicht in $D(E)$, da in T der linksseitige Grenzwert nicht existiert.

Bei einem unbeschränkten Generator kann also eine Art Explosion eintreten; G bestimmt dann die Kette nur bis zu diesem Zeitpunkt.

Bei der Behandlung des Langzeitverhaltens von Markov-Ketten mit stetigem Zeitparameter kann man weitgehend auf die diskrete Situation zurückgreifen. So nennt man einen Wahrscheinlichkeitsvektor π stationär für $\{P(t) : t \geq 0\}$, wenn $\pi P(t) = \pi$ für alle $t > 0$ gilt. Auch hier ist wieder die Verwendung der Halbgruppe selbst sehr unbequem; praktischer ist ein Kriterium, das stattdessen den Generator verwendet.

SATZ 2.9 *Ist $\{P(t) : t \geq 0\}$ eine Übergangshalbgruppe mit einem Generator G , der den Bedingungen von Satz 2.7 genügt, und ist π ein Wahrscheinlichkeitsvektor mit der Eigenschaft $\pi G = 0$, so ist π stationär für $\{P(t) : t \geq 0\}$.*

BEWEIS: Wir verwenden die Konstruktion aus Satz 2.7. Aus $\pi G = 0$ folgt, dass $\pi \tilde{P} = \pi$ gilt, wobei $\tilde{P} = I + \lambda^{-1}G$ die Ein-Schritt-Übergangsmatrix zur Markov-Kette $Y = (Y_n)_{n \in \mathbb{N}_0}$ bezeichnet. Die Verteilung π ist also stationär für Y , insbesondere gilt $\pi \tilde{P}^k = \pi$ für alle $k \in \mathbb{N}_0$. Aus dem Beweis zu Satz 2.7 ist

$$p_{ij}(t) = \sum_{k=0}^{\infty} \tilde{p}_{ij}(k) e^{-\lambda t} \frac{(\lambda t)^k}{k!}$$

bekannt, also folgt mit Fubini

$$\begin{aligned} (\pi P(t))_j &= \sum_{i \in E} \pi_i p_{ij}(t) \\ &= \sum_{k=0}^{\infty} \left(\sum_{i \in E} \pi_i \tilde{p}_{ij}(k) \right) e^{-\lambda t} \frac{(\lambda t)^k}{k!}, \\ &= \pi_j \sum_{k=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^k}{k!} = \pi_j. \end{aligned}$$

□

Auch Irreduzibilität kann wie im Fall diskreter Zeit behandelt werden, Periodizitäten können nicht auftreten. Bei der Übertragung des Hauptgrenzwertsatzes für Markov-Ketten in diskreter Zeit, Satz 1.13, ist die sogenannte *Skelett-Methode* sehr nützlich: Ist $X = (X_t)_{t \geq 0}$ eine Markov-Kette in stetiger Zeit, so ist offensichtlich für jedes $h > 0$

$$X^{(h)} = (X_n^{(h)})_{n \in \mathbb{N}_0} \quad \text{mit} \quad X_n^{(h)} := X_{nh} \quad \text{für alle } n \in \mathbb{N}_0$$

eine Markov-Kette in diskreter Zeit, und zwischen den Übergangswahrscheinlichkeiten $p_{ij}(t)$ zu X und $p_{ij}^{(h)}(n)$ zu $X^{(h)}$ hat man die einfache Beziehung

$$p_{ij}(nh) = p_{ij}^{(h)}(n) \quad \text{für alle } n \in \mathbb{N}_0, i, j \in E.$$

Satz 1.13 liefert, unter den entsprechenden Voraussetzungen, die Konvergenz der rechten Seiten; mit einem einfachen analytischen Argument folgt dann die Konvergenz der Übergangsfunktionen. Ein Teil dieser Überlegungen wird in den Aufgaben ausführlicher behandelt.

3. Die Brownsche Bewegung

3.1 Definition und erste Eigenschaften. Ein reellwertiger stochastischer Prozess $(X_t)_{t \in T}$ auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ kann als Abbildung X von Ω in die Menge \mathbb{R}^T aller Abbildungen f von T nach \mathbb{R} angesehen werden: $(X(\omega))(t) = X_t(\omega)$. Wir schreiben den Zeitparameter t gelegentlich in eine Klammer statt als Index, also beispielsweise $X(t, \omega)$ anstelle von $X_t(\omega)$, wenn dies die Formeln übersichtlicher macht. Versieht man \mathbb{R}^T mit \mathcal{B}^T , der von den Projektionen

$$\pi_t : \mathbb{R}^T \rightarrow \mathbb{R}, \quad f \mapsto f(t)$$

erzeugten σ -Algebra, so ist X $(\mathcal{A}, \mathcal{B}^T)$ -messbar und \mathbb{P}^X ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^T, \mathcal{B}^T)$. Für unser weiteres Vorgehen benötigen wir eine allgemeinere Familie von Projektionen: Ist $T_0 = \{t_1, \dots, t_n\}$ mit $t_1 < t_2 < \dots < t_n$ eine endliche Teilmenge von T , so nennen wir

$$\pi_{T_0} : \mathbb{R}^T \rightarrow \mathbb{R}^{T_0}, \quad f \mapsto (f(t_1), \dots, f(t_n))$$

die Projektion auf die T_0 -Koordinaten; π_{T_0} ordnet also einer Funktion f ihre Einschränkung $f|_{T_0}$ auf T_0 zu (wir identifizieren gelegentlich \mathbb{R}^{T_0} mit \mathbb{R}^n bei $\#T = n$). Zu jeder solchen Teilmenge von T_0 ist die Verteilung $\mathcal{L}(X_{t_1}, \dots, X_{t_n})$ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}^{T_0}, \mathcal{B}^{T_0})$; sie ergibt sich als Bild von \mathbb{P}^X unter der endlich-dimensionalen Projektion π_{T_0} . Die Gesamtheit dieser Verteilungen nennt man die *Familie der endlich-dimensionalen Verteilungen* des Prozesses (im Englischen *finite dimensional distributions*, oder kurz *Fidi's*).

DEFINITION 3.1 Ein Prozess, dessen sämtliche endlich-dimensionale Verteilungen Normalverteilungen sind, heißt *Gauß-Prozess*. Ist $(X_t)_{t \in T}$ ein Gauß-Prozess, so nennen wir $t \mapsto EX_t$ die *Erwartungswertfunktion* und $(s, t) \mapsto \text{cov}(X_s, X_t)$ die *Kovarianzfunktion* zu X .

Bekanntlich sind mehrdimensionale Normalverteilungen durch den zugehörigen Erwartungswertvektor und die zugehörige Kovarianzmatrix bereits vollständig festgelegt, also sind bei einem Gauß-Prozess durch Erwartungswertfunktion und Kovarianzfunktion bereits alle endlich-dimensionalen Verteilungen fixiert. Damit aber ist die Verteilung des Prozesses als Wahrscheinlichkeitsmaß auf $(\mathbb{R}^T, \mathcal{B}^T)$ festgelegt, denn die Mengen

$$\pi_{T_0}^{-1}(B_1 \times \dots \times B_n) = \{f \in \mathbb{R}^T : f(t_1) \in B_1, \dots, f(t_n) \in B_n\},$$

wobei B_1, \dots, B_n die Borelschen Teilmengen von \mathbb{R} und $T_0 = \{t_1, \dots, t_n\}$ die endlichen Teilmengen von T durchlaufen, bilden ein durchschnittsstabiles Erzeugendensystem von \mathcal{B}^T .

Ein stochastischer Prozess kann also als Abbildung in die Menge \mathbb{R}^T aller möglichen Pfade $t \mapsto X_t(\omega)$ angesehen werden. Bereits in den vorangegangenen beiden Abschnitten wurde deutlich, dass diese Konstruktion zu grob ist, und dass es sinnvoller ist, Prozesse mit vorgegebenen Pfadeigenschaften zu betrachten. Bei der Behandlung der Brownschen Bewegung werden wir Stetigkeit der Pfade voraussetzen, also einen stochastischen Prozess $X = (X_t)_{t \in T}$ als Zufallsgröße mit Werten im Raum $C(T)$ der stetigen Funktionen von T nach \mathbb{R} auffassen; in der Regel geht es dabei um $T = [0, 1]$ oder $T = [0, \infty)$. Als messbare Struktur $\mathcal{B}(C(T))$ auf $C(T)$ betrachten wir wie im Falle \mathbb{R}^T die von den Projektionen erzeugte σ -Algebra. Die Urbilder Borelscher Rechteckmengen unter den endlich-dimensionalen Projektionen bilden weiterhin ein durchschnittsstabiles Erzeugendensystem von $\mathcal{B}(C(T))$; insbesondere ist also auch bei Prozessen mit stetigen Pfaden die Verteilung eines Prozesses durch seine endlich-dimensionalen Verteilungen festgelegt.

DEFINITION 3.2 Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $T = \mathbb{R}_+$ oder $T = [0, R]$ mit einem $R \in (0, \infty)$, $(\mathcal{F}_t)_{t \in T}$ eine Filtration und $(B_t)_{t \in T}$ ein zu $(\mathcal{F}_t)_{t \in T}$ adaptierter Prozess. Gilt dann:

- (a) $\mathbb{P}(B_0 = 0) = 1$,
- (b) \mathbb{P} -fast alle Pfade von $(B_t)_{t \in T}$ sind stetig,
- (c) für alle $s, t \in T$ mit $s < t$ ist $B_t - B_s$ unabhängig von \mathcal{F}_s und $N(0, t-s)$ -verteilt,

so heißt $(B_t, \mathcal{F}_t)_{t \in T}$ eine (eindimensionale) *Brownsche Bewegung* oder auch *Wiener-Prozess* mit Start in 0.

Im Falle $\mathcal{F}_t = \sigma(\{B_s : s \leq t\})$, wenn also die natürliche Filtration vorliegt, sprechen wir kurz von der Brownschen Bewegung $(B_t)_{t \in T}$. Gelegentlich wird gefordert, dass *alle* Pfade der Prozesses stetig sind oder in 0 den Wert 0 haben. Für Fragen, die auf das Ausrechnen irgendwelcher Wahrscheinlichkeiten hinauslaufen, ist diese Unterscheidung bedeutungslos. Sind $X = (X_t)_{t \in T}$ und $Y = (Y_t)_{t \in T}$ stochastische Prozesse auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, so nennen wir X und Y *ununterscheidbar*, wenn

$$X(\cdot, \omega) = Y(\cdot, \omega) \quad \text{für } \mathbb{P}\text{-fast alle } \omega \in \Omega$$

gilt, wenn also außerhalb einer \mathbb{P} -Nullmenge die Pfade (als Abbildungen von T in \mathbb{R}) übereinstimmen. Ist B ein Prozess, der den Bedingungen der obigen Definition genügt, so erhält man durch Abändern auf einer Nullmenge leicht

einen von B ununterscheidbaren Prozess \tilde{B} , der eine Brownsche Bewegung im Sinne von Definition 3.2 bleibt und bei dem alle Pfade stetig sind und in 0 starten; man kann beispielsweise $B(\cdot, \omega)$ durch die Abbildung $\equiv 0$ für alle ω aus der Ausnahmemenge ersetzen oder den Ausgangswahrscheinlichkeitsraum entsprechend verändern. Diese Manipulationen können sich natürlich auf die Filtration auswirken; siehe hierzu Aufgabe 27. Schließlich bemerken wir noch, dass aus Bedingung (c) folgt, dass eine Brownsche Bewegung, wie der in Abschnitt 2.1 besprochene Poisson-Prozess, unabhängige Zuwächse hat. Das Interesse an einem stochastischen Prozess mit diesen Eigenschaften kam zunächst nicht aus der Mathematik. Bereits 1827 beobachtete der Botaniker Brown, dass Pollen im Wasser eine irreguläre Bewegung ausführen, die heute durch Zusammenstöße mit Wassermolekülen (deren mittlere Bewegungsenergie die Wassertemperatur definiert) erklärt wird. Physiker halten das Jahr 1905 für bemerkenswert—neben einer Arbeit über die spezielle Relativitätstheorie und einer Arbeit zum photoelektrischen Effekt erschien damals auch eine Arbeit über das von Brown beobachtete Phänomen (alle von demselben Autor). Der nun unter dem Namen ‘Brownsche Bewegung’ bekannte stochastische Prozess wurde auch bereits sehr früh im Zusammenhang mit der Modellierung der Fluktuationen von Aktienkursen eingesetzt, eine Anwendung, die in den letzten Jahren enormen Zulauf erhalten hat.

Mathematisch ist die Brownsche Bewegung unter anderem deswegen interessant, weil sie zu den drei wichtigsten Klassen stochastischer Prozesse gehört: Sie ist gleichzeitig ein Martingal, ein Gauß-Prozess und ein Markov-Prozess. Darüberhinaus ist dieser Prozess der Startpunkt für die Stochastische Analysis, eines der interessantesten Gebiete der modernen Mathematik. Die Martingaleigenschaft ergibt sich unmittelbar aus Forderung (c) in der Definition der Brownschen Bewegung: Für alle $s \leq t$ gilt:

$$E[B_t | \mathcal{F}_s] = E[B_t - B_s | \mathcal{F}_s] + E[B_s | \mathcal{F}_s] = E(B_t - B_s) + B_s = 0 + B_s = B_s.$$

Markov-Prozesse haben wir bisher nur bei diskretem Zustandsraum behandelt; die hier nötige Erweiterung besprechen wir in einem späteren Unterabschnitt. Die Verbindung zu Gauß-Prozessen ist Gegenstand des folgenden Resultats.

SATZ 3.3 *Eine Brownsche Bewegung $(B_t)_{t \in T}$ ist ein Gauß-Prozess mit Mittelwertfunktion $EB_t \equiv 0$ und Kovarianzfunktion $\text{cov}(B_s, B_t) = s \wedge t := \min\{s, t\}$. Ist umgekehrt $(X_t)_{t \in T}$ (mit T wie in Definition 3.2) ein Gauß-Prozess mit $EX_t \equiv 0$, $\text{cov}(X_s, X_t) = s \wedge t$ und fast sicher stetigen Pfaden, so ist $(X_t)_{t \in T}$ eine Brownsche Bewegung.*

BEWEIS: Ist $(B_t)_{t \in T}$ eine Brownsche Bewegung, so gilt $B_0 = 0$ und $B_t - B_0 \sim N(0, t)$, also $B_t \sim N(0, t)$ und damit $EB_t = 0$. Zusammen mit der

Unabhängigkeit der Zuwächse ergibt dies außerdem

$$\text{cov}(B_s, B_t) = EB_s(B_t - B_s) + EB_s^2 = (EB_s) \cdot (E(B_t - B_s)) + s = s = s \wedge t,$$

für alle s, t mit $s \leq t$. Die Unabhängigkeit der Zuwächse liefert auch eine gemeinsame Normalverteilung für

$$B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}, \quad 0 \leq t_1 \leq \dots \leq t_n,$$

denn aus der Vorlesung Stochastik II ist bekannt, dass ein Zufallsvektor mit normalverteilten und unabhängigen Komponenten eine mehrdimensionale Normalverteilung hat. Als lineare Funktion dieser Zuwächse ist dann auch

$$(B_{t_1}, B_{t_2}, \dots, B_{t_n})$$

normalverteilt.

Es sei nun umgekehrt $(X_t)_{t \in T}$ ein Gauß-Prozess mit stetigen Pfaden und der angegebenen Mittelwert- und Kovarianzfunktion. Wegen $\mathcal{L}(X_0) = N(0, 0)$ bleibt lediglich der Nachweis von Bedingung (c) aus Definition 3.2 mit der natürlichen Filtration. Es seien also $0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq s \leq t$. Mit den Rechenregeln für mehrdimensionale Normalverteilungen erhält man für den Vektor der Zuwächse

$$(X_{s_1}, X_{s_2} - X_{s_1}, \dots, X_{s_n} - X_{s_{n-1}}, X_t - X_s)$$

eine Normalverteilung mit einer Kovarianzmatrix von Diagonalforn, d.h. der Zuwachs $X_t - X_s$ ist unabhängig von den vorangegangenen Zuwächsen und damit von X_{s_1}, \dots, X_{s_n} . Da die Mengen

$$\{X_{s_1} \in B_1, \dots, X_{s_n} \in B_n\}, \quad n \in \mathbb{N}, 0 \leq s_1 \leq \dots \leq s_n \leq s, B_1, \dots, B_n \in \mathcal{B},$$

ein durchschnittsstabiles Erzeugendensystem von \mathcal{F}_s bilden, folgt die Unabhängigkeit der Variablen $X_t - X_s$ von \mathcal{F}_s .

Die Aussage $X_t - X_s \sim N(0, t - s)$ läuft auf ein einfaches Nachrechnen der ersten beiden Momente hinaus. \square

Mit diesem Satz kann man leicht beweisen, dass bestimmte Transformationen aus einer Brownschen Bewegung wieder eine Brownsche Bewegung machen.

SATZ 3.4 *Mit $(B_t)_{t \geq 0}$ sind auch die folgenden Prozesse Brownsche Bewegungen:*

- (a) $(-B_t)_{t \geq 0}$,
- (b) $(B_{a+t} - B_a)_{t \geq 0}$ bei festem $a \geq 0$,
- (c) $(c \cdot B_{t/c^2})_{t \geq 0}$ bei festem $c \neq 0$,
- (d) $(\tilde{B}_t)_{t \geq 0}$ mit $\tilde{B}_t := tB_{1/t}$ für $t > 0$, $\tilde{B}_0 := 0$.

BEWEIS: Wir beweisen nur Teil (d) und stellen den Rest als Übungsaufgabe, (a) ist natürlich ein Spezialfall von (c). Wir verwenden Satz 3.3. Es seien $t_1, \dots, t_n > 0$. Dann ist $(B_{1/t_1}, \dots, B_{1/t_n})$ ein normalverteilter Zufallsvektor, also auch

$$(\tilde{B}_{t_1}, \dots, \tilde{B}_{t_n}) = (t_1 B_{1/t_1}, \dots, t_n B_{1/t_n}),$$

denn normalverteilte Zufallsvektoren bleiben unter affinen Transformationen normalverteilt. Die endlich-dimensionalen Verteilungen von $(\tilde{B}_t)_{t \geq 0}$ sind also normal, d.h. \tilde{B} ist ein Gauß-Prozess. Für die Erwartungswerte erhalten wir bei $t > 0$

$$E\tilde{B}_t = t E B_{1/t} = t \cdot 0 = 0.$$

Zusammen mit $E\tilde{B}_0 = 0$ zeigt dies, dass sich die korrekte Erwartungswertfunktion ergibt. Weiter gilt für alle $s, t > 0$

$$\text{cov}(\tilde{B}_s, \tilde{B}_t) = \text{cov}(s B_{1/s}, t B_{1/t}) = st \left(\frac{1}{s} \wedge \frac{1}{t} \right) = s \wedge t.$$

Zusammen mit einem separaten Argument für den Fall $st = 0$ zeigt dies, dass auch die Kovarianzfunktion passt. Es bleibt also der Nachweis, dass \mathbb{P} -fast alle Pfade des Prozesses \tilde{B} stetig sind. Bekannt ist, dass ein $A \in \mathcal{A}$ mit $\mathbb{P}(A) = 1$ existiert derart, dass für alle $\omega \in A$ die Abbildung $t \mapsto B_t(\omega)$ stetig ist auf $[0, \infty)$. Für alle solchen ω ist dann offensichtlich $t \mapsto \tilde{B}_t(\omega)$ stetig auf $(0, \infty)$; insbesondere gilt

$$\sup_{0 < t \leq 1/m} |\tilde{B}_t(\omega)| = \sup_{q \in \mathbb{Q} \cap (0, 1/m]} |\tilde{B}_q(\omega)| \quad \text{für alle } \omega \in A, m \in \mathbb{N}.$$

Dies wiederum bedeutet, dass $t \rightarrow \tilde{B}_t(\omega)$ in $t = 0$ stetig ist für alle $\omega \in \tilde{F}$, wobei

$$\tilde{F} := \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{q \in \mathbb{Q} \cap (0, 1/m]} \left\{ |\tilde{B}_q| \leq \frac{1}{n} \right\}.$$

Es wurde bereits gezeigt, dass (B_t) und (\tilde{B}_t) die gleichen endlich-dimensionalen Verteilungen haben. Verwendet man die Stetigkeit des Wahrscheinlichkeitsmaßes \mathbb{P} , so impliziert dies $\mathbb{P}(\tilde{F}) = \mathbb{P}(F)$, wobei

$$F := \bigcap_{n=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{q \in \mathbb{Q} \cap (0, 1/m]} \left\{ |B_q| \leq \frac{1}{n} \right\}.$$

Wegen $A \subset F$ gilt also $\mathbb{P}(\tilde{F}) = 1$, damit $\mathbb{P}(\tilde{F} \cap A) = 1$, also sind \mathbb{P} -fast alle Pfade von \tilde{B} auch in $t = 0$ stetig. \square

3.2 Existenz. Zu Einsteins Zeiten war keineswegs klar, ob die Forderungen, die an das die Brownsche Bewegung modellierende mathematische Objekt gestellt werden, nicht vielleicht widersprüchlich sind. Erst Wiener hat in den zwanziger Jahren des letzten Jahrhunderts dieses Problem zur Zufriedenheit der Mathematiker gelöst. Heute sind (mindestens) drei Standardverfahren zum Nachweis der Existenz der Brownschen Bewegung im Sinne von Definition 3.2 bekannt: Man kann dies auf rein maßtheoretischem Wege unter Zuhilfenahme eines Satzes von Kolmogorov über projektive Familien von Maßen und eines Satzes von Kolmogorov und Chentsov über die Existenz stetiger Modifikationen beweisen, man kann funktionalanalytisch vorgehen und die Brownsche Bewegung als eine Art Fourier-Reihe mit zufälligen Koeffizienten konstruieren, und schließlich geht es auch ‘rein stochastisch’, indem man die Brownsche Bewegung als Limes von Irrfahrten darstellt. Wir besprechen die zweite Methode, die im wesentlichen auf Wiener zurückgeht.

SATZ 3.5 *Es gibt einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, auf dem ein stochastischer Prozess $(B_t)_{0 \leq t \leq 1}$ definiert werden kann mit den Eigenschaften*

- (a) $B_0(\omega) = 0$ für alle $\omega \in \Omega$,
- (b) $t \mapsto B_t(\omega)$ ist stetig für alle $\omega \in \Omega$,
- (c) für alle $0 \leq s \leq t \leq 1$ ist $B_t - B_s$ unabhängig von $\sigma(\{B_u : u \leq s\})$ und $N(0, t - s)$ -verteilt.

BEWEIS: Es sei $L^2 = L^2([0, 1], \mathcal{B}_{[0,1]}, \text{unif}(0, 1))$ der Raum der Äquivalenzklassen (modulo Lebesgue-Nullmengen) von Funktionen $f : [0, 1] \rightarrow \mathbb{R}$ mit $\int_0^1 f^2(x) dx < \infty$. Versehen mit dem inneren Produkt

$$\langle f, g \rangle := \int_0^1 f(x)g(x) dx$$

ist dies ein Hilbertraum. Weiter seien

$$S_n := \{(n, k) : 1 \leq k \leq 2^n, k \text{ ungerade}\}$$

für alle $n \in \mathbb{N}_0$ und $S := \bigcup_{n=0}^{\infty} S_n$. Die Funktionen g_{nk} , $(n, k) \in S$, mit $g_{01} \equiv 1$ und

$$g_{nk}(t) := 2^{(n-1)/2} \left(1_{[(k-1)2^{-n}, k2^{-n})} - 1_{[k2^{-n}, (k+1)2^{-n})} \right)$$

für $n \neq 0$ bilden eine Orthonormalbasis (die *Haar-Basis*) von L^2 , d.h. es gilt $\|g_{nk}\| = 1$, $\langle g_{nk}, g_{ml} \rangle = 0$ für $(n, k) \neq (m, l)$ und

$$f = \sum_{(n,k) \in S} \langle f, g_{nk} \rangle g_{nk} \quad \text{für alle } f \in L^2 \quad (1)$$

im Sinne der L^2 -Konvergenz.

Es sei nun $\{Z_{nk} : (n, k) \in S\}$ eine Familie von unabhängigen, $N(0, 1)$ -verteilten Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$. Wir definieren eine Folge $(X_n(t))_{0 \leq t \leq 1}$, $n \in \mathbb{N}_0$, von stochastischen Prozessen auf $(\Omega, \mathcal{A}, \mathbb{P})$ durch

$$X_n(t) := \sum_{m=0}^n \sum_{(m,k) \in S_m} f_{mk}(t) Z_{mk},$$

wobei

$$f_{mk}(t) := \int_0^t g_{mk}(s) ds \quad \text{für alle } t \in [0, 1], (m, k) \in S.$$

Für das Verständnis dieser Konstruktion ist vielleicht eine Illustration hilfreich: Es werden kleiner und schmaler werdende ‘Zelte’ aufeinandergetürmt. Für festes n haben die f_{nk} , $(n, k) \in S_n$, disjunkte Träger und es gilt $0 \leq f_{nk} \leq 2^{-(n+1)/2}$. Da die f_{nk} stetig sind, kann jedes X_n als Zufallsgröße mit Werten in $C[0, 1]$ aufgefasst werden.

Wir behaupten, dass für \mathbb{P} -fast-alles $\omega \in \Omega$ die Folge der Funktionen $t \mapsto X_n(t, \omega)$ mit $n \rightarrow \infty$ gleichmäßig konvergiert; die Grenzfunktion definiert dann $X(\omega)$ und ist automatisch stetig. Aus der Vorlesung Stochastik II ist bekannt, dass für eine standardnormalverteilte Zufallsvariable Z

$$\mathbb{P}(|Z| \geq a) \leq \sqrt{\frac{2}{\pi}} \frac{1}{a} e^{-a^2/2} \quad \text{für alle } a > 0$$

gilt. Es sei $\|f\|_\infty := \sup_{0 \leq t \leq 1} |f(t)|$ für alle $f \in C[0, 1]$. Dann gilt für alle $n \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}(\|X_n - X_{n-1}\|_\infty > a_n) &= \mathbb{P}\left(\sup_{(n,k) \in S_n} |Z_{nk}| > 2^{(n+1)/2} a_n\right) \\ &\leq 2^n \mathbb{P}(|Z_{01}| > 2^{(n+1)/2} a_n) \quad (\text{denn } \#S_n \leq 2^n) \\ &\leq \sqrt{\frac{2}{\pi}} 2^n 2^{-(n+1)/2} \frac{1}{a_n} \exp\left(-\frac{1}{2} 2^{n+1} a_n^2\right). \end{aligned}$$

Setzt man $a_n := (n2^{-n})^{1/2}$, so lässt sich der letzte Ausdruck weiter abschätzen durch

$$\sqrt{\frac{2}{\pi}} 2^n 2^{-(n+1)/2} 2^{n/2} n^{-1/2} \exp\left(-\frac{1}{2} 2^{n+1} 2^{-n} n\right) \leq \frac{1}{\sqrt{\pi}} n^{-1/2} 2^n e^{-n},$$

d.h. mit $A_n := \{\|X_n - X_{n-1}\|_\infty > (n2^{-n})^{1/2}\}$ gilt $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$, und somit nach dem Lemma von Borel-Cantelli $\mathbb{P}(N) = 0$ für $N := \limsup_{n \rightarrow \infty} A_n$. Zu jedem $\omega \notin N$ existiert also ein $n_0 = n_0(\omega)$ derart, dass für alle $m, n \geq n_0$

$$\|X_m(\omega) - X_n(\omega)\|_\infty \leq \sum_{k=(m \wedge n)+1}^{m \vee n} \|X_k(\omega) - X_{k-1}(\omega)\|_\infty \leq \sum_{k=m \wedge n}^{\infty} (k2^{-k})^{1/2}$$

gilt. Mit $m, n \rightarrow \infty$ geht dies gegen 0, also ist $(X_n(\omega))_{n \in \mathbb{N}}$ eine Cauchy-Folge in $(C[0, 1], \|\cdot\|_\infty)$ und hat damit einen Limes in $C[0, 1]$, den wir $X(\omega)$ nennen. Setze $X(\omega) \equiv 0$ für $\omega \in N$. Dies liefert einen Prozess $(X_t)_{t \in T}$ mit stetigen Pfaden.

Als nächstes betrachten wir Erwartungswert und Varianz von X_n . Offensichtlich gilt $EX_n(t) = 0$ für alle $n \in \mathbb{N}$, $0 \leq t \leq 1$. Weiter gilt

$$EX_n(t)X_n(s) = \sum_{m=0}^n \sum_{(m,k) \in S_m} f_{mk}(s)f_{mk}(t),$$

denn die ‘gemischten Terme’ haben wegen der Unabhängigkeit der Z -Variablen den Erwartungswert 0. Mit $f_{mk}(s) = \langle 1_{[0,s]}, g_{mk} \rangle$ und (1) ergibt sich für $n \rightarrow \infty$

$$\begin{aligned} EX_n(t)X_n(s) &\rightarrow \sum_{m=0}^{\infty} \sum_{(m,k) \in S_m} \langle 1_{[0,s]}, g_{mk} \rangle \langle 1_{[0,t]}, g_{mk} \rangle \\ &= \langle 1_{[0,s]}, 1_{[0,t]} \rangle = s \wedge t, \end{aligned} \quad (2)$$

wobei wir die Parsevalsche Gleichung verwendet haben (d.h. man benutzt (1) mit $f = 1_{[0,s]}$ und $f = 1_{[0,t]}$). Hiermit wollen wir nun zeigen, dass die endlich-dimensionalen Verteilungen zu X Normalverteilungen sind und dass die ersten beiden Momente passen. Mit Satz 3.3 folgt dann insgesamt, dass $(X_t)_{0 \leq t \leq 1}$ eine Brownsche Bewegung ist.

Es seien $t_1, \dots, t_k \in [0, 1]$. Aus dem bereits Gezeigten folgt

$$(X_n(t_1), \dots, X_n(t_k)) \rightarrow (X(t_1), \dots, X(t_k)) \quad \mathbb{P}\text{-fast sicher,}$$

also auch in Verteilung. Dies impliziert punktweise Konvergenz der charakteristischen Funktionen

$$\phi_n : \mathbb{R}^k \rightarrow \mathbb{C}, \quad \theta \mapsto E \exp\left(i \sum_{l=1}^k \theta_l X_n(t_l)\right)$$

mit $n \rightarrow \infty$ gegen

$$\phi : \mathbb{R}^k \rightarrow \mathbb{C}, \quad \theta \mapsto E \exp\left(i \sum_{l=1}^k \theta_l X(t_l)\right).$$

Es ist aber klar, wie ϕ_n aussieht:

$$\phi_n(\theta) = \exp\left(-\frac{1}{2}\theta^t \Sigma_n \theta\right) \quad \text{mit } (\Sigma_n)_{ij} = EX_n(t_i)X_n(t_j).$$

Mit (2) folgt nun

$$(\Sigma_n)_{ij} = EX_n(t_i)X_n(t_j) \rightarrow (\Sigma)_{ij} = t_i \wedge t_j$$

also insgesamt

$$\phi(\theta) = \exp\left(-\frac{1}{2}\theta^t \Sigma \theta\right), \quad \Sigma_{ij} = t_i \wedge t_j,$$

d.h. der Zufallsvektor $(X_{t_1}, \dots, X_{t_k})$ ist normalverteilt und es gilt $EX_t = 0$ für alle $t \in [0, 1]$ sowie $\text{cov}(X_s, X_t) = s \wedge t$ für alle $s, t \in [0, 1]$.

Der Prozess $B := X$ hat also die gewünschten Eigenschaften. \square

Der Satz zeigt, dass eine Brownsche Bewegung im Sinne von Definition 3.2 im Falle $T = [0, 1]$ existiert; andere Zeitbereiche werden in den Übungsaufgaben behandelt. Man nennt die Verteilung $W := \mathcal{L}(B)$ des Prozesses $(B_t)_{0 \leq t \leq 1}$, ein Wahrscheinlichkeitsmaß auf $C[0, 1]$ versehen mit der von den Projektionen π_t , $0 \leq t \leq 1$, erzeugten σ -Algebra $\mathcal{B}(C[0, 1])$, das *Wiener-Maß*. Mit $(\Omega, \mathcal{A}, \mathbb{P}) := (C[0, 1], \mathcal{B}(C[0, 1]), W)$ und $B_t := \pi_t$ hat man ein explizites Modell für eine Brownsche Bewegung $(B_t)_{0 \leq t \leq 1}$, die sogenannte *kanonische Konstruktion*.

3.3 Pfadeigenschaften. Wir beginnen mit einer einfachen globalen Eigenschaft.

SATZ 3.6 *Ist $(B_t)_{t \geq 0}$ eine Brownsche Bewegung auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, so gilt für \mathbb{P} -fast alle $\omega \in \Omega$*

$$\sup_{t \geq 0} B_t(\omega) = \infty, \quad \inf_{t \geq 0} B_t(\omega) = -\infty.$$

BEWEIS: Sei $Z := \sup_{t \geq 0} B_t$. Aus Satz 3.4 (c) folgt, dass für alle $c > 0$ Z und cZ dieselbe Verteilung haben (cZ ist die Z -Variable zu $(cB_{t/c^2})_{t \geq 0}$), also gilt wegen $Z \geq 0$

$$\mathbb{P}(Z \in \{0, \infty\}) = 1.$$

Sei $p := \mathbb{P}(Z = 0)$. Nach Satz 3.4 (b) ist $(B_{1+t} - B_1)_{t \geq 0}$ wieder eine Brownsche Bewegung, das zugehörige Supremum kann also ebenfalls mit Wahrscheinlichkeit 1 nur die Werte 0 und ∞ annehmen. Es folgt

$$\begin{aligned} \mathbb{P}(Z = 0) &\leq \mathbb{P}(B_1 \leq 0 \text{ und } B_t \leq 0 \text{ für alle } t \geq 1) \\ &= \mathbb{P}(B_1 \leq 0 \text{ und } \sup_{t \geq 0} (B_{1+t} - B_1) = 0) \\ &= \mathbb{P}(B_1 \leq 0) \mathbb{P}(\sup_{t \geq 0} (B_{1+t} - B_1) = 0), \\ &= \mathbb{P}(B_1 \leq 0) \mathbb{P}(Z = 0) \\ &= \frac{1}{2} p, \end{aligned}$$

wobei wir die Unabhängigkeit von $(B_{1+k} - B_1)_{t \geq 0}$ und B_1 verwendet haben. Aus $p \leq p/2$ und $0 \leq p \leq 1$ folgt natürlich $p = 0$, also $\mathbb{P}(\sup_{t \geq 0} B_t = \infty) = 1$. Mit Satz 3.4 (a) erhält man den zweiten Teil der Behauptung. \square

Da stetige Funktionen auf kompakten Intervallen beschränkt sind, bedeutet dies, dass für jedes $a \in \mathbb{R}$ die Menge $\{t : B_t = a\}$ für \mathbb{P} -fast-alle Pfade unbeschränkt ist; jeder Punkt wird immer wieder besucht. Der Prozess kehrt mit Wahrscheinlichkeit 1 an den Ausgangspunkt 0 zurück (sogar unendlich oft), ist also in diesem Sinne rekurrent. Insbesondere existiert für \mathbb{P} -fast-alle $\omega \in \Omega$ eine (von ω abhängige) Folge $(t_n)_{n \in \mathbb{N}}$ mit $t_n \uparrow \infty$ und $B_{t_n}(\omega) = 0$ für alle $n \in \mathbb{N}$. Verwendet man die Zeitumkehr $(B_t)_{t \geq 0} \rightsquigarrow (\tilde{B}_t)_{t \geq 0}$ aus Satz 3.4 (d), $\tilde{B}_t := tB_{1/t}$, so sieht man, dass für \mathbb{P} -fast alle $\omega \in \Omega$ die Nullstellen des Pfades $t \rightarrow B_t(\omega)$ einen Häufungspunkt in 0 haben. Insbesondere käme als Ableitung in 0 nur der Wert 0 in Frage ... zusammengenommen mit der räumlichen Homogenität der Brownschen Bewegung sieht dies ganz nach einem Beweis dafür aus, dass die Pfade der Brownschen Bewegung konstant sind!

Eine Funktion $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ heißt *Lipschitz-stetig* in $x \in \mathbb{R}_+$, wenn es ein $\delta > 0$ und ein $K < \infty$ gibt mit

$$|f(x) - f(y)| \leq K|x - y| \quad \text{für alle } y \in [x - \delta, x + \delta] \cap \mathbb{R}_+.$$

SATZ 3.7 *\mathbb{P} -fast alle Pfade einer Brownschen Bewegung sind in keinem Punkt Lipschitz-stetig.*

BEWEIS: Es seien zunächst $a, b \in \mathbb{R}$ fest mit $0 < a < b < \infty$. Weiter sei $K > 0$. Für $n \in \mathbb{N}$ mit $a - 2/n > 0$, $k = 1, \dots, n$ setzen wir

$$A_n := \left\{ \omega \in \Omega : \exists s \in (a, b) \forall t \in \left[s - \frac{2}{n}, s + \frac{2}{n} \right] : \right. \\ \left. |B_t(\omega) - B_s(\omega)| \leq K|t - s| \right\}$$

$$\Delta_{nk} := \left| B\left(\frac{k}{n}\right) - B\left(\frac{k-1}{n}\right) \right|.$$

Zu jedem Intervall $I = [s - 2/n, s + 2/n]$ existiert ein $k \in \{\lceil na \rceil, \dots, \lceil nb + 1 \rceil\}$ mit $(k-2)/n, (k-1)/n, k/n, (k+1)/n \in I$, denn das Intervall $[ns - 2, ns + 2]$ enthält vier aufeinanderfolgende ganze Zahlen. Auf A_n erhält man mit der Dreiecksungleichung für $j = k-1, k, k+1$

$$\left| B\left(\frac{j}{n}\right) - B\left(\frac{j-1}{n}\right) \right| \leq \left| B\left(\frac{j}{n}\right) - B(s) \right| + \left| B\left(\frac{j-1}{n}\right) - B(s) \right| \leq \frac{4K}{n},$$

also

$$A_n \subset \bigcup_{k=\lfloor na \rfloor}^{\lceil nb+1 \rceil} \left\{ \omega \in \Omega : \Delta_{nj}(\omega) \leq \frac{4K}{n} \text{ für } j = k-1, k, k+1 \right\}.$$

Da die Zuwächse Δ_{nj} , $j \in \mathbb{N}$, unabhängig und identisch verteilt sind, folgt

$$\begin{aligned} \mathbb{P}(A_n) &\leq \sum_{k=\lfloor na \rfloor}^{\lceil nb+1 \rceil} \mathbb{P}\left(\Delta_{n,k-1} \leq \frac{4K}{n}\right) \mathbb{P}\left(\Delta_{nk} \leq \frac{4K}{n}\right) \mathbb{P}\left(\Delta_{n,k+1} \leq \frac{4K}{n}\right) \\ &\leq \lceil nb+1 \rceil \mathbb{P}\left(\Delta_{n,1} \leq \frac{4K}{n}\right)^3. \end{aligned}$$

Es gilt

$$\sqrt{n} \left(B\left(\frac{1}{n}\right) - B\left(\frac{0}{n}\right) \right) \sim N(0, 1)$$

und daher

$$\mathbb{P}\left(\Delta_{n,1} \leq \frac{4K}{n}\right) = \int_{-4K/\sqrt{n}}^{4K/\sqrt{n}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \leq \frac{8K}{\sqrt{2\pi n}},$$

also insgesamt $\mathbb{P}(A_n) \rightarrow 0$ mit $n \rightarrow \infty$. Offensichtlich gilt $A_n \subset A_{n+1}$, also folgt $\mathbb{P}(A) = 0$ für $A := \bigcup_{n=n_0(a)}^{\infty} A_n$; hierbei ist $n_0(a)$ das kleinste $n \in \mathbb{N}$ mit $a - 2/n > 0$. Lässt man nun K die natürlichen Zahlen durchlaufen, so erhält man als Vereinigung der A -Mengen eine \mathbb{P} -Nullmenge N mit der Eigenschaft, dass für alle $\omega \notin N$ gilt: Zu keinem $s \in (a, b)$ und keinem $\delta > 0$ existiert ein $K < \infty$ mit

$$|B_t(\omega) - B_s(\omega)| \leq K |s - t| \text{ auf } t \in [s - \delta, s + \delta].$$

Die Pfade der Brownschen Bewegung, die in irgendeinem $s \in (a, b)$ Lipschitz-stetig sind, sind also in einer \mathbb{P} -Nullmenge enthalten. Als Vereinigung dieser Mengen mit $a = 1/l$, $b = l$ über $l \in \mathbb{N}$ erhält man also wieder eine Nullmenge— \mathbb{P} -fast alle Pfade sind in keinem $s > 0$ Lipschitz-stetig. Es bleibt die Behandlung des Zeitpunktes $s = 0$; dies geschieht in Aufgabe 32. \square

Damit kann nun das oben angedeutete Dilemma aufgelöst werden: Ist die Funktion $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ in $s \in \mathbb{R}_+$ differenzierbar, so existiert ein (endlicher) Limes zum Differenzenquotienten

$$\frac{f(t) - f(s)}{t - s}, \quad t \rightarrow s,$$

und f wäre dann in s Lipschitz-stetig. Als Konsequenz von Satz 3.7 erhält man also, dass \mathbb{P} -fast alle Pfade der Brownschen Bewegung in keinem Punkt differenzierbar sind. Es erscheint bemerkenswert, dass das Wiener-Maß auf eine Klasse von Funktionen konzentriert ist (den stetigen, nirgends differenzierbaren), die in der klassischen Analysis als uninteressante Perversionen betrachtet wurden.

Als *Totalvariation* $V_a^b f$ einer Funktion $f : \mathbb{R}_+ \rightarrow \mathbb{R}$ auf dem Intervall $[a, b] \subset \mathbb{R}_+$ bezeichnet man das Supremum der Werte

$$\sum_{i=1}^n |f(x_i) - f(x_{i-1})|,$$

wobei das Supremum über alle endlichen Unterteilungen $a = x_0 \leq x_1 \leq \dots \leq x_n = b$ von $[a, b]$ gebildet wird. Nach einem klassischen Satz von Lebesgue sind Funktionen mit endlicher Totalvariation ℓ -fast-überall differenzierbar, also implizieren die obigen Bemerkungen, dass fast alle Pfade der Brownschen Bewegung auf jedem Intervall von unbeschränkter Totalvariation sind—selbst kleinste Pfadstücke sind wegen

$$\left\| \begin{pmatrix} y \\ f(y) \end{pmatrix} - \begin{pmatrix} x \\ f(x) \end{pmatrix} \right\| \geq |f(y) - f(x)|$$

‘unendlich lang’. Man kann dies auch aus dem folgenden, für den Itô-Kalkül sehr wichtigen Satz herleiten. Es sei wieder $(B_t)_{t \geq 0}$ eine Brownsche Bewegung auf $(\Omega, \mathcal{A}, \mathbb{P})$.

SATZ 3.8 *Es sei $(\{t_{nj} : j = 0, \dots, k_n\})_{n \in \mathbb{N}}$ eine Folge von Unterteilungen von $[0, t]$, also*

$$0 = t_{n0} \leq t_{n1} \leq \dots \leq t_{nk_n} = t \quad \text{für alle } n \in \mathbb{N},$$

mit gegen 0 konvergierender Weite, also

$$\lim_{n \rightarrow \infty} \max_{j=1, \dots, k_n} |t_{nj} - t_{n,j-1}| = 0.$$

Dann gilt mit $n \rightarrow \infty$

$$\sum_{j=1}^{k_n} (B(t_{nj}) - B(t_{n,j-1}))^2 \rightarrow t \quad \text{in Wahrscheinlichkeit.}$$

BEWEIS: Es sei Z_n die Summe der quadrierten Zuwächse über die n -te Partition; wir betrachten $E(Z_n - t)^2$. Für $i \neq j$ gilt

$$E\left(\left(B(t_{nj}) - B(t_{n,j-1})\right)^2 - (t_{nj} - t_{n,j-1})\right)\left(\left(B(t_{ni}) - B(t_{n,i-1})\right)^2 - (t_{ni} - t_{n,i-1})\right) = 0,$$

denn die Zufallsvariablen $\left(B(t_{nj}) - B(t_{n,j-1})\right)^2 - (t_{nj} - t_{n,j-1})$ und $\left(B(t_{ni}) - B(t_{n,i-1})\right)^2 - (t_{ni} - t_{n,i-1})$ sind unabhängig und haben Erwartungswert 0. Also folgt mit $t = \sum_{i=1}^n (t_{ni} - t_{n,i-1})$

$$E(Z_n - t)^2 = \sum_{j=1}^{k_n} E\left(\left(B(t_{nj}) - B(t_{n,j-1})\right)^2 - (t_{nj} - t_{n,j-1})\right)^2,$$

denn die gemischten Terme verschwinden. Ist $X \sim N(0, 1)$, so gilt $EX^4 = 3$, also hat man bei $Y \sim N(0, \sigma^2)$

$$E(Y^2 - \sigma^2)^2 = EY^4 - 2\sigma^2 EY^2 + \sigma^4 = 2\sigma^4$$

und damit

$$E\left(\left(B(t_{nj}) - B(t_{n,j-1})\right)^2 - (t_{nj} - t_{n,j-1})\right)^2 = 2(t_{nj} - t_{n,j-1})^2,$$

also

$$\begin{aligned} E(Z_n - t)^2 &= 2 \sum_{j=1}^{k_n} (t_{nj} - t_{n,j-1})^2 \\ &\leq 2 \left(\max_{1 \leq j \leq k_n} |t_{nj} - t_{n,j-1}| \right) \sum_{j=1}^{k_n} |t_{nj} - t_{n,j-1}| \\ &\rightarrow 0 \quad \text{mit } n \rightarrow \infty, \end{aligned}$$

da nach Voraussetzung die Weite der Partitionen gegen 0 geht; die Summe ist stets t . Damit ist gezeigt, dass Z_n in L^2 gegen t konvergiert, woraus bekanntlich die Konvergenz in Wahrscheinlichkeit folgt. \square

Wesentlich bei diesem Resultat ist, dass die Zerlegungsfolge nicht von ω abhängt. Sind die Unterteilungen Verfeinerungen voneinander, so gilt sogar fast sichere Konvergenz.

Man nennt den Grenzwert der Summe der quadrierten Zuwächse über feiner werdende Zerlegungen die *quadratische Variation*; fast alle Pfade einer Brownschen Bewegung haben also auf jedem Intervall $[0, t]$ die quadratische Variation t . Es ist bemerkenswert, dass diese Größe nicht von ω abhängt; in Bezug

auf die quadratische Variation haben die Brownschen Pfade eine deterministische Struktur. Mit

$$\sum_{i=1}^n (f(x_i) - f(x_{i-1}))^2 \leq \max_{1 \leq i \leq n} |f(x_i) - f(x_{i-1})| \sum_{i=1}^n |f(x_i) - f(x_{i-1})|$$

folgt, dass eine stetige Funktion f auf einem kompakten Intervall $[a, b]$ die quadratische Variation 0 hat, wenn $V_a^b f < \infty$ gilt. Insbesondere gilt also für \mathbb{P} -fast alle $\omega \in \Omega$, wie bereits oben bemerkt, $V_a^b B(\cdot, \omega) = \infty$ für alle a, b mit $0 \leq a < b < \infty$.

3.4 Die Brownsche Bewegung als Markov-Prozess. Bereits in Bemerkung 1.24 haben wir Markov-Prozesse in diskreter Zeit $T = \mathbb{N}_0$ und mit allgemeinem Zustandsraum (E, \mathcal{B}) mit dem aus der Vorlesung Stochastik II bekannten Begriff der Übergangswahrscheinlichkeit in Verbindung gebracht, diese erscheint als Version der bedingten Verteilung von X_{n+1} unter $\sigma\{X_0, \dots, X_n\}$. Wollen wir eine vergleichbare Konstruktion bei $T = \mathbb{R}_+$ ausführen, so benötigen wir anstelle einer einzelnen Übergangswahrscheinlichkeit eine Halbgruppe von solchen Objekten, in Analogie zum Übergang von Abschnitt 1 auf Abschnitt 2, und können wieder versuchen, diese Halbgruppe durch ihren Generator zu beschreiben. Wir betrachten den Zustandsraum $(\mathbb{R}, \mathcal{B})$.

DEFINITION 3.9 Eine Familie $(P_t)_{t \geq 0}$ von Übergangswahrscheinlichkeiten von $(\mathbb{R}, \mathcal{B})$ nach $(\mathbb{R}, \mathcal{B})$ heißt *Markov-Halbgruppe* oder auch *Übergangshalbgruppe*, wenn für alle $s, t \geq 0$ und alle $A \in \mathcal{B}$ gilt:

$$P_{s+t}(x, A) = \int P_t(y, A) P_s(x, dy). \quad (\text{HG})$$

Schreiben wir kurz $P_s \otimes P_t$ für die rechte Seite, so wird (HG) zu $P_{s+t} = P_s \otimes P_t$.

BEISPIEL 3.10 Für alle $t \geq 0$ sei $P_t(x, \cdot) = N(x, t)$, wobei eine Normalverteilung mit Varianz 0 wie üblich als im Erwartungswert konzentriertes Einpunktmaß aufgefasst wird. Die Messbarkeit von Abbildungen der Form

$$x \mapsto P_t(x, A) = \int_A \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t}(y-x)^2\right) dy$$

mit einer Borel-Menge A ergibt sich aus Überlegungen, die in der Vorlesung Stochastik II im Zusammenhang mit dem Satz von Fubini angestellt wurden. Für das Integral auf der rechten Seite von (HG) erhält man nach Vertauschung der Integrationsreihenfolge den Ausdruck

$$\int_A \left(\int \frac{1}{\sqrt{2\pi t}} \exp\left(-\frac{1}{2t}(z-y)^2\right) \frac{1}{\sqrt{2\pi s}} \exp\left(-\frac{1}{2s}(y-x)^2\right) dy \right) dz.$$

Das innere Integral ist das Faltungsprodukt der Dichten zu $N(0, t)$ und $N(x, s)$ im Punkt z , also der Wert der Wahrscheinlichkeitsdichte in z zur Verteilung $N(x, s + t)$. Man erhält also insgesamt die Wahrscheinlichkeit von A unter $N(x, s + t)$, d.h. die linke Seite von (HG).

Wir schreiben abkürzend $\mathbb{P}[X_t \in A | \mathcal{F}_s]$ für $E[1_A(X_t) | \mathcal{F}_s]$ und erinnern daran, dass Gleichheiten von Zufallsvariablen und bedingten Erwartungswerten mit dem Zusatz ‘ist Version von’ gelesen werden müssen.

DEFINITION 3.11 Es seien $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum mit Filtration $(\mathcal{F}_t)_{t \geq 0}$, $(X_t)_{t \geq 0}$ ein hierzu adaptierter reellwertiger Prozess, μ ein Wahrscheinlichkeitsmaß auf $(\mathbb{R}, \mathcal{B})$ und $(P_t)_{t \geq 0}$ eine Markov-Halbgruppe. Gilt dann $\mathcal{L}(X_0) = \mu$ sowie

$$\mathbb{P}[X_t \in A | \mathcal{F}_s] = P_{t-s}(X_s, A) \quad \text{für alle } A \in \mathcal{B}, 0 \leq s \leq t, \quad (\text{M})$$

so heißt $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein *zeitlich homogener Markov-Prozess mit Übergangshalbgruppe $(P_t)_{t \geq 0}$ und Startverteilung μ* .

In der Regel ist $(\mathcal{F}_t)_{t \geq 0}$ die natürliche Filtration zu $(X_t)_{t \geq 0}$. Zeitliche Homogenität ergibt sich in (M) daraus, dass die rechte Seite von den Zeitpunkten s und t nur über deren Differenz $t - s$ abhängt. Der Zusammenhang zu unserer früheren Interpretation, dass Markov-Ketten ‘kein Gedächtnis’ haben, ergibt sich daraus, dass die Abhängigkeit von \mathcal{F}_s auf der linken Seite sich auf der rechten Seite auf eine Abhängigkeit von X_s reduziert. Schließlich dürfte klar sein, dass aus (M) ein Analogon zu den Chapman-Kolmogorov-Gleichungen aus den Abschnitten 1 und 2 hergeleitet werden kann, was wiederum auf die Halbgruppeneigenschaft (HG) der Übergangswahrscheinlichkeiten führt.

Zur Überprüfung von (M) reicht es,

$$\mathbb{P}(F \cap \{X_t \in A\}) = \int_F P_{t-s}(X_s, A) d\mathbb{P} \quad \text{für alle } F \in \mathcal{F}_s, A \in \mathcal{B}$$

zu zeigen, denn die \mathcal{F}_s -Messbarkeit ergibt sich daraus, dass die rechte Seite in (M) eine Funktion der \mathcal{F}_s -messbaren Zufallsgröße X_s ist; bei den Mengen F kann man sich weiter auf die Elemente eines durchschnittsstabilen Erzeugendensystems von \mathcal{F}_s zurückziehen.

SATZ 3.12 *Es sei $(B_t)_{t \geq 0}$ eine Brownsche Bewegung, $(\mathcal{F}_t)_{t \geq 0}$ die zugehörige natürliche Filtration. Dann ist $(B_t, \mathcal{F}_t)_{t \geq 0}$ ein Markov-Prozess mit Startverteilung δ_0 und Übergangswahrscheinlichkeiten $P_t(x, \cdot) = N(x, t)$.*

BEWEIS: Wir leiten zunächst eine Hilfsaussage her. Es seien $0 = t_0 \leq t_1 \leq \dots \leq t_n$ und $g : \mathbb{R}^n \rightarrow \mathbb{R}$ eine beschränkte, $(\mathcal{B}^n, \mathcal{B})$ -messbare Funktion. Wir setzen $Y_i := B_{t_i} - B_{t_{i-1}}$ für $i = 1, \dots, n$; aus Abschnitt 4 ist bekannt, dass Y_1, \dots, Y_n unabhängig sind und dass $Y_i \sim N(0, t_i - t_{i-1})$ gilt. Damit erhält man

$$\begin{aligned} & Eg(B_{t_1}, \dots, B_{t_n}) \\ &= Eg(Y_1, Y_1 + Y_2, \dots, Y_1 + \dots + Y_n) \\ &= \int \dots \int g(y_1, y_1 + y_2, \dots, y_1 + \dots + y_n) N(0, t_n - t_{n-1})(dy_n) \\ &\quad N(0, t_{n-1} - t_{n-2})(dy_{n-1}) \dots N(0, t_1)(dy_1) \\ &= \int \dots \int g(x_1, \dots, x_n) N(x_{n-1}, t_n - t_{n-1})(dx_n) \dots N(0, t_1)(dx_1). \end{aligned}$$

Es seien nun $A \in \mathcal{B}$, $0 \leq s \leq t$. Zu zeigen ist, dass die Zufallsvariable $N(B_s, t - s)(A)$ eine Version von $\mathbb{P}[B_t \in A | \mathcal{F}_s]$ ist, also im Lichte der obigen Bemerkungen

$$\int_C 1_A(B_t) d\mathbb{P} = \int_C N(B_s, t - s)(A) d\mathbb{P}$$

für alle C aus einem durchschnittstabilen Erzeugendensystem von \mathcal{F}_s . Ein solches wird gebildet von den Mengen der Form

$$\bigcap_{i=1}^n \{B_{s_i} \in A_i\}, \quad 0 \leq s_1 \leq \dots \leq s_n = s, \quad A_1, \dots, A_n \in \mathcal{B}.$$

Setzt man

$$g(x_1, \dots, x_n, x) := 1_A(x) \prod_{i=1}^n 1_{A_i}(x_i),$$

so erhält man für die linke Seite

$$\begin{aligned} & \mathbb{P}(B_{s_1} \in A_1, \dots, B_{s_n} \in A_n, B_t \in A) \\ &= Eg(B_{s_1}, \dots, B_{s_n}, B_t) \\ &= \int_{A_1} \dots \int_{A_n} \int_A N(x_n, t - s)(dx) N(x_{n-1}, s_n - s_{n-1})(dx_n) \dots N(0, s_1)(dx_1) \end{aligned}$$

und mit der oben bewiesenen Hilfsaussage bei

$$g(x_1, \dots, x_n) := N(x_n, t - s)(A) \prod_{i=1}^n 1_{A_i}(x_i)$$

für die rechte Seite den Ausdruck

$$\begin{aligned} & \int_{\{B(s_1) \in A_1, \dots, B(s_n) \in A_n\}} N(B_s, t-s)(A) d\mathbb{P} \\ &= Eg(B_{s_1}, \dots, B_{s_n}) \\ &= \int_{A_1} \dots \int_{A_n} N(x_n, t-s)(A) N(x_{n-1}, s_n - s_{n-1})(dx_n) \dots N(0, s_1)(dx_1). \end{aligned}$$

Die gewünschte Gleichheit erhält man nun durch Ausintegrieren von x . \square

Die Markov-Eigenschaft (M) lässt sich auch in der Form

$$\mathbb{P}[X_{s+t} \in A | \mathcal{F}_s] = P_t(X_s, A) \quad \text{für alle } A \in \mathcal{B}, s, t \geq 0 \quad (\text{M}')$$

schreiben (die Bezeichnung ist nicht ganz korrekt, da sie die zeitliche Homogenität beinhaltet und auch die Existenz einer Übergangshalbgruppe voraussetzt). Fasst man P_t als Operator auf einem Funktionenraum auf, setzt man also in Analogie zu der Situation bei Markov-Ketten mit diskretem Zustandsraum

$$P_t f(x) := \int f(y) P_t(x, dy) \quad \text{für alle } x \in \mathbb{R}$$

(zumindest für messbare beschränkte Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$ sind diese Integrale definiert), so erhält man für die rechte Seite von (M')

$$P_t(X_s, A) = \int 1_A(y) P_t(X_s, dy) = (P_t 1_A)(X_s).$$

Da wir in dieser Vorlesung $\mathbb{P}[A | \mathcal{F}]$ nur als andere Schreibweise für $E[1_A | \mathcal{F}]$ verwenden, lässt sich (M') insgesamt umschreiben zu

$$E[f(X_{s+t}) | \mathcal{F}_s] = (P_t f)(X_s) \quad (\text{M}'')$$

für alle Funktionen f der Form $f = 1_A$ mit einem $A \in \mathcal{B}$. Die aus der Konstruktion des Lebesgue-Integrals bekannte 'übliche Maschinerie' zusammen mit der bedingten Version des Satzes von der monotonen Konvergenz zeigt, dass dann (M'') sogar für alle beschränkten und messbaren $f : \mathbb{R} \rightarrow \mathbb{R}$ gilt; man vergleiche dies mit Lemma 1.18. Umgekehrt kann man sich fragen, für welche Funktionenklasse (M'') gelten muss, um (M') folgern zu können. Im 'un'-bedingten Fall wird diese Frage bei Funktionen der Form $f(x) = e^{i\theta x}$, $\theta \in \mathbb{R}$, durch den Eindeigkeitssatz für charakteristische Funktion positiv beantwortet. Man kann dies auf bedingte charakteristische Funktionen übertragen und erhält, dass (M') erfüllt ist, wenn

$$E[\exp(i\theta X_{s+t}) | \mathcal{F}_s] = \int e^{i\theta y} P_t(X_s, dy)$$

für alle $\theta \in \mathbb{R}$ gilt. Diese Überlegungen führen auf einen direkteren Beweis zu Satz 3.12: Für alle $\theta \in \mathbb{R}$ erhält man mit Definition 3.2 (c)

$$\begin{aligned} E[\exp(i\theta B_{t+s}) | \mathcal{F}_s] &= E[\exp(i\theta(B_{t+s} - B_s + B_s)) | \mathcal{F}_s] \\ &= \exp(i\theta B_s) E \exp(i\theta(B_{t+s} - B_s)) \\ &= \exp(i\theta B_s) \exp(-\theta^2 t/2) \\ &= (P_t(e^{i\theta \cdot}))(B_s), \end{aligned}$$

denn $(P_t(e^{i\theta \cdot}))(x)$ ist der Wert in θ der charakteristischen Funktion zur Verteilung $N(x, t)$. Insgesamt ergibt sich also abermals die Eigenschaft (M) für die Brownsche Bewegung.

In der klassischen Theorie der Markov-Prozesse ist die Übergangshalbgruppe $(P_t)_{t \geq 0}$ der Ausgangspunkt. Diese kann unter bestimmten Voraussetzungen durch den zugehörigen Generator

$$G := \lim_{t \downarrow 0} \frac{1}{t} (P_t - P_0)$$

charakterisiert werden, wobei man wieder die Formel $P_t = \exp(tG)$ im Hinterkopf hat. Wir betrachten den Spezialfall der Brownschen Bewegung etwas näher. Es sei $C_b^2(\mathbb{R})$ die Menge der zweimal stetig differenzierbaren Funktionen $f: \mathbb{R} \rightarrow \mathbb{R}$ mit beschränkter zweiter Ableitung f'' .

SATZ 3.13 *Es sei $(P_t)_{t \geq 0}$ die Übergangshalbgruppe zur Brownschen Bewegung. Dann gilt*

$$\lim_{t \downarrow 0} \frac{1}{t} (P_t f - f) = \frac{1}{2} f'' \quad \text{für alle } f \in C_b^2(\mathbb{R}).$$

BEWEIS: Es sei $f \in C_b^2(\mathbb{R})$. Für alle $t \geq 0$ gilt

$$\begin{aligned} \frac{1}{t} (P_t f - f)(x) &= \int \frac{1}{t} (f(y) - f(x)) N(x, t)(dy) \\ &= \frac{1}{\sqrt{2\pi}} \int t^{-\frac{3}{2}} (f(y) - f(x)) e^{-(y-x)^2/(2t)} dy \\ &= \frac{1}{\sqrt{2\pi}} \int \frac{1}{t} (f(x + z\sqrt{t}) - f(x)) e^{-z^2/2} dz. \end{aligned}$$

Mit der Taylor-Entwicklung

$$f(x + z\sqrt{t}) = f(x) + z\sqrt{t}f'(x) + \frac{1}{2}z^2t f''(\xi),$$

wobei $\xi = \xi(x, t, z)$ zwischen x und $x + z\sqrt{t}$ liegt, erhält man

$$\frac{1}{t}(P_t f - f)(x) = \frac{1}{\sqrt{2\pi}} \frac{f'(x)}{\sqrt{t}} \int z e^{-z^2/2} dz + \frac{1}{2} \int \frac{1}{\sqrt{2\pi}} z^2 f''(\xi) e^{-z^2/2} dz.$$

Das erste Integral verschwindet aus Symmetriegründen. Bei dem zweiten benutzen wir, dass $f''(\xi)$ mit $t \rightarrow 0$ gegen $f''(x)$ konvergiert. Da f'' nach Voraussetzung beschränkt ist, lässt sich der Satz von der majorisierten Konvergenz anwenden und mit

$$\int \frac{1}{\sqrt{2\pi}} z^2 e^{-z^2/2} dz = 1$$

folgt insgesamt die Behauptung. \square

Der infinitesimale Generator zur Brownschen Bewegung ist also ein Differentialoperator. Dieser Sachverhalt, insbesondere seine Variante für mehrdimensionale Brownsche Bewegungen, stellt eine sehr fruchtbare Verbindung zu Differentialgleichungen her; ein einfaches Beispiel wird in den Übungsaufgaben behandelt.

3.5 Die starke Markov-Eigenschaft, mit Anwendungen. Ist $\tau : \Omega \rightarrow [0, \infty]$ eine Stoppzeit, so nennt man bekanntlich

$$\mathcal{F}_\tau := \{A \in \mathcal{A} : A \cap \{\tau \leq t\} \in \mathcal{F}_t \text{ für alle } t \geq 0\}$$

die σ -Algebra der τ -Vergangenheit. Durch geeignete Annahmen an die Pfade des Prozesses lassen sich Probleme vermeiden, die anderenfalls durch die überabzählbare Zeitmenge entstehen könnten (siehe auch Abschnitt 2, insbesondere den Beweis zu Lemma 2.1). Man nennt einen Prozess X mit der Messbarkeits-eigenschaft aus Teil (b) des folgenden Lemmas *progressiv messbar*.

LEMMA 3.14 (a) *Ist τ eine Stoppzeit, so wird durch*

$$\tau_n(\omega) := \begin{cases} 2^{-n} \lceil 2^n \tau(\omega) \rceil, & \text{falls } \tau(\omega) < \infty, \\ \infty, & \text{falls } \tau(\omega) = \infty, \end{cases}$$

eine Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten definiert, die mit $n \rightarrow \infty$ antiton gegen τ konvergiert. Ist τ beschränkt, so nimmt jedes τ_n nur endlich viele Werte an.

(b) *Ist $X = (X_t)_{t \geq 0}$ ein zu $(\mathcal{F}_t)_{t \geq 0}$ adaptierter Prozess mit rechtsstetigen Pfaden, so ist für jedes $t \geq 0$ die Abbildung*

$$[0, t] \times \Omega \rightarrow \mathbb{R}, \quad (s, \omega) \mapsto X_s(\omega)$$

$\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$ -messbar.

(c) Ist τ eine endliche Stoppzeit und $(X_t)_{t \geq 0}$ ein zu $(\mathcal{F}_t)_{t \geq 0}$ adaptierter Prozess mit rechtsstetigen Pfaden, so ist

$$X_\tau : \Omega \rightarrow \mathbb{R}, \quad \omega \mapsto X_{\tau(\omega)}(\omega),$$

\mathcal{F}_τ -messbar.

BEWEIS: Dass τ_n eine Stoppzeit ist, folgt aus $\{\tau_n \leq t\} = \{\tau \leq 2^{-n} \lfloor 2^n t \rfloor\}$ und $2^{-n} \lfloor 2^n t \rfloor \leq t$; die anderen Teilaussagen sind offensichtlich.

Zum Beweis von (b) überprüft man zunächst die $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$ -Messbarkeit von Abbildungen der Form

$$[0, t] \times \Omega \rightarrow \mathbb{R}, \quad (s, \omega) \mapsto 1_{[u,v)}(s) X_v(\omega),$$

mit $0 \leq u < v \leq t$: Ist A eine Borel-Menge, die 0 nicht enthält, so ergibt sich als Urbild die Rechteckmenge $[u, v) \times X_v^{-1}(A)$. Analog erhält man die Messbarkeit von $(s, \omega) \mapsto 1_{\{t\}}(s) X_t(\omega)$. Die Abbildungen $X^{(n)}$,

$$X^{(n)}(s, \omega) := \sum_{k=0}^{n-1} 1_{[kt/n, (k+1)t/n)}(s) X\left(\frac{(k+1)t}{n}, \omega\right) + 1_{\{t\}}(s) X_t(\omega),$$

sind Linearkombinationen solcher Elementarabbildungen, also ebenfalls $\mathcal{B}_{[0,t]} \otimes \mathcal{F}_t$ -messbar. Aufgrund der Rechtsstetigkeit der Pfade von X ergibt sich schließlich die interessierende Abbildung als Grenzwert der $X^{(n)}$ mit $n \rightarrow \infty$.

Bei (c) ist zu zeigen, dass $X_\tau^{-1}(A) \cap \{\tau \leq t\}$ ein Element von \mathcal{F}_t ist für alle $t \geq 0$ und alle $A \in \mathcal{B}$. Sei $t \geq 0$. Die gewünschte Aussage folgt, wenn die Verknüpfung der Abbildungen

$$\begin{aligned} \Omega &\rightarrow [0, t] \times \Omega, & \omega &\mapsto (\tau(\omega) \wedge t, \omega), \\ [0, t] \times \Omega &\rightarrow \mathbb{R}, & (s, \omega) &\mapsto X_s(\omega) \end{aligned}$$

\mathcal{F}_t -messbar ist, wobei wir auf dem Zwischen-Raum $[0, t] \times \Omega$ die σ -Algebra $\mathcal{B}[0, t] \otimes \mathcal{F}_t$ betrachten. Man hat dann nämlich $X_{\tau \wedge t}^{-1}(A) \in \mathcal{F}_t$, also wegen $\{\tau \leq t\} \in \mathcal{F}_t$

$$X_\tau^{-1}(A) \cap \{\tau \leq t\} = X_{\tau \wedge t}^{-1}(A) \cap \{\tau \leq t\} \in \mathcal{F}_t.$$

Die Messbarkeit der ersten Abbildung ergibt sich wegen (nun ist $A \in \mathcal{F}_t$)

$$\{\omega \in \Omega : (\tau(\omega) \wedge t, \omega) \in [0, s] \times A\} = A \cap \{\tau \leq s\}$$

daraus, dass τ eine Stoppzeit ist. Bei der zweiten Komponente verwendet man die bereits bewiesene Aussage (b). \square

Ist τ eine endliche Stoppzeit zu einer Brownschen Bewegung, so lassen sich wie bei Poisson-Prozessen in Abschnitt 2 der Prä- τ - und der auf Start in 0 normalisierte Post- τ -Prozess definieren. Man erhält die folgende zu Lemma 2.1 analoge Aussage, die wir mit Blick auf die Anwendung im nächsten Unterabschnitt etwas allgemeiner formulieren. Man beachte, dass der Prä- τ -Prozess $(B_{\tau \wedge t})_{t \geq 0}$ messbar ist bezüglich der σ -Algebra der τ -Vergangenheit, also impliziert die Unabhängigkeit von \mathcal{F}_τ und $(B_{\tau+t} - B_\tau)_{t \geq 0}$, dass der Prä- τ - und der Post- τ -Prozess unabhängig sind. Außerdem ist der Post- τ -Prozess wieder eine Brownsche Bewegung.

SATZ 3.15 *Es sei $(B_t, \mathcal{F}_t)_{t \geq 0}$ eine Brownsche Bewegung, τ eine endliche Stoppzeit. Dann sind die σ -Algebra \mathcal{F}_τ und der Prozess $(B_{\tau+t} - B_\tau)_{t \geq 0}$ stochastisch unabhängig und $(B_{\tau+t} - B_\tau)_{t \geq 0}$ hat dieselbe Verteilung wie $(B_t)_{t \geq 0}$.*

BEWEIS: Wir behaupten zunächst, dass für alle $t \geq 0$ der Zuwachs $B_{\tau+t} - B_\tau$ unabhängig ist von \mathcal{F}_τ und die Verteilung $N(0, t)$ hat.

Angenommen, τ nimmt nur abzählbar viele Werte an. Für alle $A \in \mathcal{F}_\tau$, $C \in \mathcal{B}$ erhält man dann unter Verwendung der in Definition 3.2 postulierten Eigenschaften der Brownschen Bewegung

$$\begin{aligned} \mathbb{P}(A \cap \{B_{\tau+t} - B_\tau \in C\}) &= \sum_{s \geq 0} \mathbb{P}(A \cap \{\tau = s\} \cap \{B_{s+t} - B_s \in C\}) \\ &= \sum_{s \geq 0} \mathbb{P}(A \cap \{\tau = s\}) \mathbb{P}(B_{s+t} - B_s \in C) \\ &= \mathbb{P}(A) N(0, t)(C). \end{aligned}$$

Mit $A = \Omega$ ergibt dies

$$\mathbb{P}(B_{\tau+t} - B_\tau \in C) = N(0, t)(C),$$

also die gewünschte Verteilungsaussage. Setzt man dies wiederum oben ein, so erhält man

$$\mathbb{P}(A \cap \{B_{\tau+t} - B_\tau \in C\}) = \mathbb{P}(A) \mathbb{P}(B_{\tau+t} - B_\tau \in C)$$

für alle $A \in \mathcal{F}_\tau$, $C \in \mathcal{B}$, also die behauptete Unabhängigkeit. Damit ist die Zwischenbehauptung für Stoppzeiten mit abzählbarem Wertebereich bewiesen. Ist nun τ eine beliebige (endliche) Stoppzeit, so ergibt der bereits bewiesene Fall für $\tau_n := 2^{-n} \lceil 2^n \tau \rceil$ wegen $\mathcal{F}_\tau \subset \mathcal{F}_{\tau_n}$ für alle $A \in \mathcal{F}_\tau$

$$\mathbb{P}(A \cap \{B_{\tau_n+t} - B_{\tau_n} \in C\}) = \mathbb{P}(A) N(0, t)(C).$$

Die Intervalle $C = (-\infty, c)$, $c \in \mathbb{R}$, bilden ein durchschnittsstabiles Erzeugendensystem zu \mathcal{B} . Die übliche Argumentation zeigt, dass die gewünschte Aussage folgen würde aus

$$\lim_{n \rightarrow \infty} \int_A 1_{(-\infty, c)}(B_{\tau_n+t} - B_{\tau_n}) d\mathbb{P} = \int_A 1_{(-\infty, c)}(B_{\tau+t} - B_{\tau}) d\mathbb{P}.$$

Die Stetigkeit der Pfade impliziert $B_{\tau_n} \rightarrow B_{\tau}$, $B_{\tau_n+t} \rightarrow B_{\tau+t}$ mit $n \rightarrow \infty$, allerdings wird dies in die nicht stetige Funktion $1_{(-\infty, c)}$ eingesetzt. Der Satz von der majorisierten Konvergenz lässt sich aber trotzdem anwenden, wenn der Randwert c nur für eine \mathbb{P} -Nullmenge eine Rolle spielt, wenn also $\mathbb{P}(B_{\tau+t} - B_{\tau} = c) = 0$ gilt. Nun impliziert die bereits erwähnte ‘punktweise’ Konvergenz von $B_{\tau_n+t} - B_{\tau_n}$ gegen $B_{\tau+t} - B_{\tau}$ auch die Konvergenz in Verteilung dieser Variablen. Da alle Folgeelemente $N(0, t)$ -verteilt sind, muss dies auch für die Limesvariable gelten, d.h. $\{c\}$ ist in der Tat eine Nullmenge für die Verteilung von $B_{\tau+t} - B_{\tau}$.

Mit Hilfe der nun vollständig bewiesenen Zwischenbehauptung soll nun die Aussage des Satzes bewiesen werden. Es seien $A \in \mathcal{F}_{\tau}$, $0 =: t_0 < t_1 < \dots < t_n$, $C_1, \dots, C_n \in \mathcal{B}$. Wendet man die bereits bewiesene Aussage nacheinander auf die Stoppzeiten $\tau + t_{n-1}, \tau + t_{n-2}, \dots, \tau + t_0$ an, so erhält man

$$\begin{aligned} \mathbb{P}(A \cap \bigcap_{k=1}^n \{B(\tau + t_k) - B(\tau + t_{k-1}) \in C_k\}) \\ = \mathbb{P}(A) \prod_{k=1}^n \mathbb{P}(B(\tau + t_k) - B(\tau + t_{k-1}) \in C_k). \end{aligned}$$

Da die Mengen der Form

$$\bigcap_{k=1}^n \{B(\tau + t_k) - B(\tau + t_{k-1}) \in C_k\}$$

ein durchschnittsstabiles Erzeugendensystem der vom Post- τ -Prozess erzeugten σ -Algebra bilden, folgt hieraus die behauptete Unabhängigkeit; die Verteilungsaussage folgt ganz analog. \square

Die folgende erste Anwendung kann als eine Invarianzeigenschaft des Wiener-Maßes angesehen werden; im Gegensatz zu den Resultaten in Satz 3.4 ist es nun eine nicht-deterministische Zeit, die die Transformation festlegt.

SATZ 3.16 (Das Spiegelungsprinzip) *Es sei $(B_t)_{t \geq 0}$ eine Brownsche Bewegung auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$, τ sei eine endliche Stoppzeit. Dann ist auch der ab τ gespiegelte Prozess $(\tilde{B}_t)_{t \geq 0}$,*

$$\tilde{B}_t(\omega) := \begin{cases} 2B_{\tau(\omega)}(\omega) - B_t(\omega), & \text{falls } t \geq \tau(\omega), \\ B_t(\omega), & \text{sonst,} \end{cases}$$

eine Brownsche Bewegung.

BEWEIS: Es seien $X = (X_t)_{t \geq 0}$ und $Y = (Y_t)_{t \geq 0}$,

$$X_t := B_{\tau \wedge t}, \quad Y_t := B_{\tau+t} - B_\tau \quad \text{für alle } t \geq 0,$$

der zu B gehörende Prä- τ - bzw. Post- τ -Prozess. Aus dem Tripel (τ, X, Y) lässt sich B rekonstruieren; genauer hat man $B = \Psi(\tau, X, Y)$ mit

$$\Psi : [0, \infty) \times C[0, \infty) \times C[0, \infty) \rightarrow C[0, \infty), \quad (a, f, g) \mapsto h,$$

wobei

$$h(t) := \begin{cases} f(t), & \text{falls } t \leq a, \\ f(a) + g(t - a), & \text{sonst} \end{cases}$$

(die hier eigentlich nötigen Messbarkeitsnachweise etc. schenken wir uns). Nach Satz 3.15 ist (τ, X) unabhängig von Y und Y ist eine Brownsche Bewegung. Nach Satz 3.4 (a) ist auch $-Y$ eine Brownsche Bewegung, die Unabhängigkeit von (τ, X) geht bei dieser deterministischen Transformation natürlich nicht verloren. Insbesondere haben die Tripel (τ, X, Y) und $(\tau, X, -Y)$ dieselbe Verteilung. Da \tilde{B} zu dem zweiten Tripel in derselben Beziehung steht wie B zu dem ersten, genauer:

$$B = \Psi(\tau, X, Y), \quad \tilde{B} = \Psi(\tau, X, -Y),$$

haben B und \tilde{B} dieselbe Verteilung. □

Als Anwendung bestimmen wir die Verteilung von M_t ,

$$M_t := \sup_{0 \leq s \leq t} B_s \quad \text{für alle } t \geq 0.$$

Hierbei machen wir von einem Spezialfall des Spiegelungsprinzips Gebrauch, bei dem der Pfad eines Prozesses ab seinem ersten Besuch in einem Wert a an der horizontalen Linie durch a gespiegelt wird. Es sei also, für $a \in \mathbb{R}$,

$$\tau_a := \inf\{t \geq 0 : B_t = a\},$$

wobei wir das Infimum über eine leere Menge wieder als ∞ lesen.

SATZ 3.17 *Es sei $t \geq 0$. Dann haben die Variablen M_t , $M_t - B_t$ und $|B_t|$ dieselbe Verteilung.*

BEWEIS: Es seien $x > 0$ und $y < x$. Spiegelung an einer horizontalen Linie durch x ab dem ersten Eintritt in x transformiert eine Funktion f mit $f(0) = 0$, $f(t) \leq y$ und $\max_{0 \leq s \leq t} f(s) \geq x$ in eine Funktion g mit $g(0) = 0$ und $g(t) \geq 2x - y$, und umgekehrt. Wendet man nun Satz 3.16 mit der Stoppzeit τ_x an (diese ist nach Satz 3.6 'nur' \mathbb{P} -fast sicher endlich, man kann aber für die hier nötige Argumentation ohne weiteres τ_x durch $\tau_x \wedge (t+1)$ ersetzen), so folgt hieraus

$$\mathbb{P}(M_t \geq x, B_t \leq y) = \mathbb{P}(B_t \geq 2x - y) \quad \text{für alle } x > 0, y < x.$$

Bezeichnet $\phi(x) = (2\pi)^{-1/2} \exp(-x^2/2)$ die Dichte zur Standardnormalverteilung, so führt dies zusammen mit $t^{-1/2}B_t \sim N(0,1)$ auf die gemeinsame Dichte

$$-\frac{\partial^2}{\partial x \partial y} \mathbb{P}(M_t \geq x, B_t \leq y) = -\frac{\partial^2}{\partial x \partial y} \mathbb{P}(B_t \geq 2x - y) = -\frac{2}{t} \phi' \left(\frac{2x - y}{\sqrt{t}} \right)$$

für $x \geq 0, y \leq x$; wegen $M_t \geq B_t$ hat die Dichte außerhalb dieses Bereiches den Wert 0. Mit der Transformationsformel für Dichten (Satz 3.12 der Vorlesung Stochastik II) erhält man hieraus als gemeinsame Dichte von M_t und $M_t - B_t$ die Funktion

$$f_{M_t, M_t - B_t}(x, y) = -\frac{2}{t} \phi' \left(\frac{x + y}{\sqrt{t}} \right), \quad x, y \geq 0.$$

Ausintegrieren liefert

$$\begin{aligned} f_{M_t}(x) &= \int f_{M_t, M_t - B_t}(x, y) dy \\ &= \int_0^\infty \frac{\partial}{\partial y} \left(-\frac{2}{\sqrt{t}} \phi \left(\frac{x + y}{\sqrt{t}} \right) \right) dy \\ &= \frac{2}{\sqrt{t}} \phi \left(\frac{x}{\sqrt{t}} \right), \quad x \geq 0. \end{aligned}$$

Vertauscht man x und y , so führt diese Rechnung auf

$$f_{M_t - B_t}(y) = \frac{2}{\sqrt{t}} \phi \left(\frac{y}{\sqrt{t}} \right), \quad y \geq 0,$$

d.h. M_t und $M_t - B_t$ haben dieselbe Dichte (insbesondere also auch dieselbe Verteilung). Dass diese wiederum auch Dichte zu $|B_t|$ ist, ist leicht zu überprüfen. \square

Als Korollar erhält man die Verteilung von bestimmten Eintrittszeiten der Brownschen Bewegung: Für $a > 0$ sind offensichtlich die Aussagen $\tau_a \leq t$ und $M_t \geq a$ äquivalent, also gilt

$$\mathbb{P}(\tau_a \leq t) = \mathbb{P}(M_t \geq a) = \mathbb{P}(|B_t| \geq a) = \mathbb{P}(|B_1| \geq at^{-1/2})$$

für alle $t > 0$, woraus sich für τ_a die Dichte

$$f_{\tau_a}(t) = \frac{a}{\sqrt{2\pi t^3}} \exp\left(-\frac{a^2}{2t}\right) \quad \text{für alle } t > 0$$

ergibt. Dass $E\tau_a = \infty$ gilt, wurde bereits in Aufgabe 34 (c) gezeigt.

Bei unserer zweiten Anwendung geht es um die Struktur der Nullstellenmengen

$$Z(\omega) := \{t \geq 0 : B_t(\omega) = 0\}$$

der Pfade $t \mapsto B_t(\omega)$ einer Brownschen Bewegung $(B_t)_{t \geq 0}$ auf $(\Omega, \mathcal{A}, \mathbb{P})$, $\omega \in \Omega$. Bekanntlich nennt man x einen Häufungspunkt der Menge A (für uns reicht: $A \subset \mathbb{R}$), wenn es eine Folge $(a_n)_{n \in \mathbb{N}} \subset A \setminus \{x\}$ gibt mit $\lim_{n \rightarrow \infty} a_n = x$; A heißt *perfekt*, wenn es mit der Menge seiner Häufungspunkte übereinstimmt ($[a, b]$ mit $a < b$ ist perfekt; (a, b) , $\{a\}$ und \mathbb{Q} sind es nicht).

SATZ 3.18 *Für \mathbb{P} -fast alle $\omega \in \Omega$ ist $Z(\omega)$ perfekt.*

BEWEIS: Nach Definition der Brownschen Bewegung gibt es ein $C \in \mathcal{A}$ mit $\mathbb{P}(C) = 1$ derart, dass $t \mapsto B_t(\omega)$ für alle $\omega \in C$ stetig ist. Für diese ist $Z(\omega)$ als Urbild der abgeschlossenen Menge $\{0\}$ abgeschlossen; es bleibt also zu zeigen, dass für \mathbb{P} -fast alle $\omega \in \Omega$ jedes $t_0 \in Z(\omega)$ Häufungspunkt von $Z(\omega)$ ist.

Bereits vor Satz 3.7 hatten wir festgestellt, dass $\mathbb{P}(A_0) = 1$ gilt mit einer Menge A_0 mit der Eigenschaft, dass für alle $\omega \in A_0$ der Wert $t_0 = 0$ Häufungspunkt von $Z(\omega)$ ist. Für jede rationale Zahl $q > 0$ sei nun A_q die Menge aller $\omega \in \Omega$ mit der Eigenschaft, dass

$$\tau_q(\omega) := \inf\{t \geq q : B_t(\omega) = 0\}$$

Häufungspunkt von $Z(\omega) \cap (\tau_q(\omega), \infty)$ ist. Da τ_q eine Stoppzeit ist, ist nach Satz 3.15 $(\tilde{B}_t)_{t \geq 0}$ mit $\tilde{B}_t(\omega) := B_{\tau_q(\omega)+t}(\omega)$ wieder eine Brownsche Bewegung (man hat $B_{\tau_q} \equiv 0$ \mathbb{P} -fast sicher). Entscheidend ist nun, dass A_q die Menge A_0 zu $(\tilde{B}_t)_{t \geq 0}$ ist, also Wahrscheinlichkeit 1 hat.

Setze nun

$$A := C \cap \bigcap_{q \in \mathbb{Q}, q \geq 0} A_q.$$

Wegen $\mathbb{P}(A) = 1$ reicht der Nachweis, dass $Z(\omega)$ für alle $\omega \in A$ perfekt ist.

Angenommen, t_0 ist kein Häufungspunkt von $Z(\omega)$. Dann muss $t_0 > 0$ gelten und es muss ein $\epsilon > 0$ existieren mit $(t_0 - \epsilon, t_0) \cap Z(\omega) = \emptyset$. Sei $q \in (t_0 - \epsilon, t_0) \cap \mathbb{Q}$. Wegen $t_0 \in A_q$ ist t_0 aber Häufungspunkt einer Folge $(t_n)_{n \in \mathbb{N}} \subset Z(\omega) \cap (t_0, \infty)$, man hat also einen Widerspruch. \square

Nach Aufgabe 36 ist $Z(\omega)$ für \mathbb{P} -fast alle $\omega \in \Omega$ eine Lebesgue-Nullmenge, enthält also insbesondere kein Intervall $[a, b]$ mit $a < b$ (auch Aufgabe 33 impliziert dies). Nun ist bekannt, dass nicht-leere perfekte Mengen überabzählbar sind (siehe beispielsweise Theorem 6.65 in Hewitt und Stromberg, *Real and Abstract Analysis*); für fast alle Pfade der Brownschen Bewegung erhält man also als Nullstellenmenge eine überabzählbare Lebesgue-Nullmenge (!). Ähnlich wie bei der Kombination von Stetigkeit und Differenzierbarkeit ist also auch hier der überraschende Fall bei der Brownschen Bewegung die Regel.

4. Das stochastische Integral

Gegenstand dieses Abschnitts ist eine der wichtigsten Entwicklungen der modernen Stochastik, die einen neuen Blick auf zahlreiche innermathematische Themen liefert, aber auch für Anwendungen in vielen Bereichen, von der Signalverarbeitung bis zur Finanzmathematik, von größter Bedeutung ist.

Unser Ziel ist es, zu zwei stochastischen Prozessen $(H_t)_{t \geq 0}$ und $(X_t)_{t \geq 0}$ das stochastische Integral

$$\int H dX = \int_0^\infty H_s dX_s$$

zu konstruieren. Wir beginnen mit drei vorbereitenden Unterabschnitten.

4.1 Gleichgradige Integrierbarkeit. Wir benötigen ein Hilfsmittel zur Vertauschung von Limes und Integral, das über den Satz von der majorisierten Konvergenz hinausgeht.

DEFINITION 4.1 Eine Familie $\{X_i : i \in I\}$ von Zufallsvariablen auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ heißt *gleichgradig integrierbar*, wenn gilt:

$$\forall \epsilon > 0 \exists C < \infty \forall i \in I : \int_{|X_i| > C} |X_i| d\mathbb{P} < \epsilon.$$

Die Bedingung der Definition kann alternativ in der Form

$$\lim_{C \rightarrow \infty} \sup_{i \in I} \int_{|X_i| > C} |X_i| d\mathbb{P} = 0$$

geschrieben werden.

BEISPIEL 4.2 (a) Ist X ein Zufallsvariable mit $E|X| < \infty$, so folgt mit dem Satz von der majorisierten Konvergenz

$$\lim_{C \rightarrow \infty} \int_{|X| > C} |X| d\mathbb{P} = 0,$$

also ist die ‘Familie’ $\{X\}$ gleichgradig integrierbar. Offensichtlich ist die endliche Vereinigung gleichgradig integrierbarer Familien wieder gleichgradig integrierbar; insbesondere sind also alle endlichen Familien von Zufallsvariablen mit existierendem Erwartungswert gleichgradig integrierbar.

(b) Ist $(X_i)_{i \in I}$ eine Familie von Zufallsvariablen, Y eine weitere Zufallsvariable mit $E|Y| < \infty$, und gilt $|X_i| \leq Y$ für alle $i \in I$, so ist $\{X_i : i \in I\}$ gleichgradig integrierbar. Dies folgt unmittelbar aus der Abschätzung

$$\int_{|X_i| > C} |X_i| d\mathbb{P} \leq \int_{Y > C} Y d\mathbb{P} \quad \text{für alle } i \in I,$$

denn die rechte Seite hängt nicht von i ab und geht mit $C \rightarrow \infty$ gegen 0 (vergleiche Teil (a)).

(c) Ist $\{X_i : i \in I\}$ L^p -beschränkt für ein $p > 1$, gilt also $\sup_{i \in I} E|X_i|^p < \infty$, so folgt mit

$$\int_{|X_i| > C} |X_i| d\mathbb{P} \leq \int_{|X_i| > C} C^{1-p} |X_i|^p d\mathbb{P} \leq C^{1-p} \sup_{i \in I} E|X_i|^p$$

die gleichgradige Integrierbarkeit der Familie: Die Oberschranke auf der rechten Seite hängt nicht von $i \in I$ ab und geht wegen $p > 1$ gegen 0 mit $C \rightarrow \infty$. Die analoge Aussage mit $p = 1$ gilt übrigens nicht (Aufgabe 39).

Einen weiteren wichtigen Typ gleichgradig integrierbarer Familien von Zufallsvariablen erhält man im Zusammenhang mit bedingten Erwartungswerten. Hierzu benötigen wir ein Hilfsmittel.

LEMMA 4.3 *Es sei X eine Zufallsvariable mit $E|X| < \infty$. Dann gilt:*

$$\forall \epsilon > 0 \quad \exists \delta > 0 \quad \forall A \in \mathcal{A} : \quad (\mathbb{P}(A) < \delta \implies \int_A |X| d\mathbb{P} < \epsilon).$$

BEWEIS: Gilt die Aussage nicht, so existieren ein $\epsilon_0 > 0$ und eine Folge $(A_n)_{n \in \mathbb{N}} \subset \mathcal{A}$ mit

$$\mathbb{P}(A_n) \leq 2^{-n} \quad \text{und} \quad \int_{A_n} |X| d\mathbb{P} \geq \epsilon_0 \quad \text{für alle } n \in \mathbb{N}.$$

Sei $A := \limsup_{n \rightarrow \infty} A_n$; das Borel-Cantelli-Lemma liefert $\mathbb{P}(A) = 0$. Andererseits folgt mit dem Lemma von Fatou:

$$\begin{aligned} \int_A |X| d\mathbb{P} &= E|X| - \int_{A^c} |X| d\mathbb{P} \\ &= E|X| - \int \liminf_{n \rightarrow \infty} (1_{A_n^c} \cdot |X|) d\mathbb{P} \\ &\geq E|X| - \liminf_{n \rightarrow \infty} \int_{A_n^c} |X| d\mathbb{P} \\ &= \limsup_{n \rightarrow \infty} \int_{A_n} |X| d\mathbb{P} \\ &\geq \epsilon_0. \end{aligned}$$

Da $1_A \cdot |X|$ \mathbb{P} -fast sicher gleich 0 ist, muss andererseits $\int_A |X| d\mathbb{P} = 0$ gelten. Man erhält also einen Widerspruch. \square

SATZ 4.4 *Ist X eine Zufallsvariable mit $E|X| < \infty$, so ist die Familie*

$$\{E[X|\mathcal{F}] : \mathcal{F} \text{ Unter-}\sigma\text{-Algebra von } \mathcal{A}\}$$

gleichgradig integrierbar.

BEWEIS: Sei $\epsilon > 0$ gegeben. Nach Lemma 4.3 existiert ein $\delta > 0$ mit

$$\forall A \in \mathcal{A} : \left(\mathbb{P}(A) < \delta \implies \int_A |X| d\mathbb{P} < \epsilon \right). \quad (1)$$

Wähle zu diesem δ ein $C > 0$ mit $C^{-1}E|X| < \delta$. Sei nun \mathcal{F} irgendeine Unter- σ -Algebra von \mathcal{A} und Y eine Version von $E[X|\mathcal{F}]$. Die Jensensche Ungleichung für bedingte Erwartungswerte liefert

$$|Y| \leq E[|X||\mathcal{F}] \quad (2)$$

und damit

$$C \mathbb{P}(|Y| \geq C) \leq E|Y| \leq E|X|,$$

also $\mathbb{P}(|Y| \geq C) < \delta$. Wegen $\{|Y| \geq C\} \in \mathcal{F}$ folgt mit (1) und (2)

$$\int_{|Y| \geq C} |Y| d\mathbb{P} \leq \int_{|Y| \geq C} E[|X||\mathcal{F}] d\mathbb{P} = \int_{|Y| \geq C} |X| d\mathbb{P} < \epsilon. \quad \square$$

Das folgende Hauptresultat dieses Unterabschnitts kann als eine bestmögliche Verallgemeinerung des Satzes von der majorisierten Konvergenz betrachtet werden; siehe hierzu auch Beispiel 4.2 (b).

SATZ 4.5 *Es seien X, X_1, X_2, \dots Zufallsvariablen mit endlichem Erwartungswert. Dann sind äquivalent:*

- (i) *Mit $n \rightarrow \infty$ konvergiert X_n im Mittel gegen X , d.h. $\lim_{n \rightarrow \infty} E|X_n - X| = 0$.*
- (ii) *Mit $n \rightarrow \infty$ konvergiert X_n in Wahrscheinlichkeit gegen X , d.h.*

$$\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| > \epsilon) = 0 \quad \text{für alle } \epsilon > 0,$$

und $\{X_n : n \in \mathbb{N}\}$ ist gleichgradig integrierbar.

BEWEIS: (i) \Rightarrow (ii): Wegen $\epsilon \mathbb{P}(|X_n - X| > \epsilon) \leq E|X_n - X|$ folgt aus der L^1 -Konvergenz die Konvergenz in Wahrscheinlichkeit. Zum Nachweis der gleichgradigen Integrierbarkeit sei $\epsilon > 0$ vorgegeben. Zu diesem ϵ existiert ein n_0 derart, dass für alle $n \geq n_0$ $E|X_n - X| < \epsilon/2$ gilt. Lemma 4.3, n_0 -mal angewendet, liefert ein $\delta > 0$ derart, dass für alle $A \in \mathcal{A}$ mit $\mathbb{P}(A) < \delta$

$$\int_A |X_n| d\mathbb{P} < \epsilon \quad \text{für } n = 1, \dots, n_0 - 1, \quad \int_A |X| d\mathbb{P} < \frac{\epsilon}{2} \quad (1)$$

gilt. Wegen $E|X_n| \leq E|X_n - X| + E|X|$ ist die Folge $(X_n)_{n \in \mathbb{N}}$ in L^1 beschränkt, also existiert ein $C > 0$ mit

$$\frac{1}{C} \sup_{n \in \mathbb{N}} E|X_n| < \delta.$$

Hieraus folgt $\mathbb{P}(|X_n| \geq C) < \delta$ für alle $n \in \mathbb{N}$, (1) kann mit $A = \{|X_n| \geq C\}$ verwendet werden, und man erhält

$$\int_{|X_n| \geq C} |X_n| d\mathbb{P} < \epsilon \quad \text{für } n = 1, \dots, n_0 - 1,$$

und für alle $n \geq n_0$

$$\begin{aligned} \int_{|X_n| \geq C} |X_n| d\mathbb{P} &\leq \int_{|X_n| \geq C} |X| d\mathbb{P} + \int_{|X_n| \geq C} |X_n - X| d\mathbb{P} \\ &\leq \frac{\epsilon}{2} + E|X_n - X| \leq \epsilon. \end{aligned}$$

Da $\epsilon > 0$ beliebig war, folgt hieraus die behauptete gleichgradige Integrierbarkeit.

(ii) \Rightarrow (i): Für alle $C > 0$ definieren wir eine Funktion $\phi_C : \mathbb{R} \rightarrow \mathbb{R}$ durch

$$\phi_C(x) := \begin{cases} C, & x > C, \\ x, & |x| \leq C, \\ -C, & x < -C. \end{cases}$$

Wir zeigen zunächst

$$\lim_{n \rightarrow \infty} E|\phi_C(X_n) - \phi_C(X)| = 0 \quad \text{für alle } C > 0. \quad (2)$$

Sei hierzu $\epsilon > 0$ vorgegeben. Mit $|\phi_C(x) - \phi_C(y)| \leq |x - y|$ folgt

$$\begin{aligned} E|\phi_C(X_n) - \phi_C(X)| &= \int_{|X_n - X| > \epsilon} |\phi_C(X_n) - \phi_C(X)| d\mathbb{P} \\ &\quad + \int_{|X_n - X| \leq \epsilon} |\phi_C(X_n) - \phi_C(X)| d\mathbb{P} \\ &\leq 2C \mathbb{P}(|X_n - X| > \epsilon) + \epsilon, \end{aligned}$$

also $\limsup_{n \rightarrow \infty} E|\phi_C(X_n) - \phi_C(X)| \leq \epsilon$. Da $\epsilon > 0$ beliebig war, folgt hieraus die Zwischenbehauptung (2).

Sei nun $\epsilon > 0$. Der Satz von der majorisierten Konvergenz liefert (vgl. Beispiel 4.2 (a))

$$E|\phi_C(X) - X| \leq \int_{|X| > C} |X| d\mathbb{P} \rightarrow 0 \quad \text{mit } C \rightarrow \infty.$$

Hiermit und mit der gleichgradigen Integrierbarkeit von $\{X_n : n \in \mathbb{N}\}$ erhält man ein $C < \infty$ mit

$$\begin{aligned} E|\phi_C(X) - X| &< \frac{\epsilon}{3}, \\ E|\phi_C(X_n) - X_n| &\leq \int_{|X_n| \geq C} |X_n| d\mathbb{P} < \frac{\epsilon}{3} \quad \text{für alle } n \in \mathbb{N}. \end{aligned}$$

Zu diesem C kann man wegen (2) ein n_0 finden derart, dass

$$E|\phi_C(X_n) - \phi_C(X)| < \frac{\epsilon}{3} \quad \text{für alle } n \geq n_0$$

gilt und somit auch

$$\begin{aligned} E|X_n - X| &\leq E|X_n - \phi_C(X_n)| + E|\phi_C(X_n) - \phi_C(X)| + E|\phi_C(X) - X| \\ &\leq \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Damit ist gezeigt, dass für alle $\epsilon > 0$ ein n_0 existiert derart, dass für alle $n \geq n_0$ $E|X_n - X| \leq \epsilon$ gilt. \square

4.2 Ergänzungen zur Martingaltheorie. Wir betrachten zunächst, wie in der Vorlesung Stochastik II, den Fall diskreter Zeit, und hier zunächst Prozesse mit Indexmenge $T = -\mathbb{N}$. Das folgende Resultat ist auch als *Rückwärtskonvergenzsatz* von Doob bekannt, die σ -Algebren werden beim Grenzübergang gröber.

SATZ 4.6 *Es sei $(X_n, \mathcal{F}_n)_{n \in -\mathbb{N}}$ ein Martingal, $\mathcal{F}_{-\infty} := \bigcap_{n=1}^{\infty} \mathcal{F}_{-n}$. Dann gilt mit $n \rightarrow -\infty$*

$$X_n \rightarrow X_{-\infty} := E[X_{-1} | \mathcal{F}_{-\infty}] \quad \mathbb{P}\text{-fast sicher und im Mittel.}$$

BEWEIS: Wie beim Beweis des Vorwärtskonvergenzsatzes erhält man eine $\mathcal{F}_{-\infty}$ -messbare, endliche Zufallsvariable $X_{-\infty}$, gegen die X_{-n} mit $n \rightarrow \infty$ fast sicher konvergiert (man betrachtet wieder die Anzahl der Überquerungen von Intervallen mit rationalen Endpunkten etc.). Im Unterschied zur Vorwärts-situation hat man nun wegen $X_{-n} = E[X_{-1} | \mathcal{F}_{-n}]$ die gleichgradige Integrierbarkeit der Familie $\{X_n : n \in -\mathbb{N}\}$ (Satz 4.4) und damit automatisch zusätzlich die L^1 -Konvergenz (Satz 4.5). Wegen

$$\left| \int_A X_{-n} d\mathbb{P} - \int_A X_{-\infty} d\mathbb{P} \right| \leq \int |X_{-n} - X_{-\infty}| d\mathbb{P}$$

liefert diese Konvergenz für alle $A \in \mathcal{F}_{-\infty}$

$$\int_A X_{-\infty} d\mathbb{P} = \lim_{n \rightarrow \infty} \int_A X_{-n} d\mathbb{P} = \int_A X_{-1} d\mathbb{P},$$

wobei wir $\mathcal{F}_{-\infty} \subset \mathcal{F}_{-n}$ und $X_{-n} = E[X_{-1} | \mathcal{F}_{-n}]$ verwendet haben. Also ist $X_{-\infty}$ auch eine Version von $E[X_{-1} | \mathcal{F}_{-\infty}]$. \square

Beim nächsten Resultat betrachten wir wieder den ‘feiner werdenden’ Fall, also eine Filtration $(\mathcal{F}_n)_{n \in \mathbb{N}}$. Es sei $\mathcal{F}_\infty := \sigma(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n)$. Offensichtlich ist auch $(\mathcal{F}_n)_{n \in \mathbb{N} \cup \{\infty\}}$ eine Filtration.

SATZ 4.7 *Es sei $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}}$ ein gleichgradig integrierbares Supermartingal. Dann existiert eine \mathcal{F}_∞ -messbare Zufallsvariable X_∞ mit den Eigenschaften*

$$X_n \rightarrow X_\infty \quad \mathbb{P}\text{-fast sicher und in } L^1$$

mit $n \rightarrow \infty$ sowie

$$X_n \geq E[X_\infty | \mathcal{F}_n] \quad \text{für alle } n \in \mathbb{N},$$

d.h. auch $(X_n, \mathcal{F}_n)_{n \in \mathbb{N} \cup \{\infty\}}$ ist ein Supermartingal. Die analogen Aussagen gelten auch für Submartingale und Martingale.

BEWEIS: Gleichgradig integrierbare Familien sind nach Aufgabe 39 L^1 -beschränkt, also lässt sich der Vorwärtskonvergenzsatz von Doob anwenden und man erhält ein X_∞ derart, dass X_n mit $n \rightarrow \infty$ \mathbb{P} -fast sicher gegen X_∞ konvergiert. Fast sichere Konvergenz impliziert die Konvergenz in Wahrscheinlichkeit, also folgt $E|X_n - X_\infty| \rightarrow 0$ mit Satz 4.5.

Sei nun $A \in \mathcal{F}_n$. Die Supermartingaleigenschaft ergibt

$$\int_A X_n d\mathbb{P} \geq \int_A X_m d\mathbb{P} \quad \text{für alle } m \geq n.$$

Wegen

$$\left| \int_A X_m d\mathbb{P} - \int_A X_\infty d\mathbb{P} \right| \leq E|X_m - X_\infty| \rightarrow 0$$

mit $m \rightarrow \infty$ folgt hieraus

$$\int_A X_n d\mathbb{P} \geq \int_A X_\infty d\mathbb{P}.$$

Da $A \in \mathcal{F}_n$ beliebig war, folgt $X_n \geq E[X_\infty | \mathcal{F}_n]$ mit Aufgabe 67 (a) der Vorlesung Stochastik II. \square

Viele Resultate zu Martingalen $(X_n, \mathcal{F}_n)_{n \in \mathbb{N}_0}$ mit diskretem Zeitparameter lassen sich, gelegentlich unter Zusatzvoraussetzungen an die Pfade, auf Martingale $(X_t, \mathcal{F}_t)_{t \geq 0}$ mit stetigem Zeitparameter verallgemeinern. Solche Zusatzvoraussetzungen stellen beispielsweise sicher, dass $\sup_{0 \leq s \leq t} X_s$ messbar ist; siehe beispielsweise Aufgabe 41, in der eine der nützlichen Martingalungleichungen auf den stetigen Fall übertragen wird. Der folgende Satz ist in diesem Sinne eine Erweiterung und Ergänzung zu Satz 4.7.

SATZ 4.8 Für ein Martingal $(X_t, \mathcal{F}_t)_{t \geq 0}$ sind äquivalent:

- (i) X_t konvergiert mit $t \rightarrow \infty$ in L^1 -Norm gegen einen Grenzwert X_∞ .
- (ii) Es gibt eine Zufallsvariable X_∞ mit der Eigenschaft

$$X_t = E[X_\infty | \mathcal{F}_t] \quad \text{für alle } t \geq 0.$$

- (iii) $\{X_t : t \geq 0\}$ ist gleichgradig integrierbar.

BEWEIS: (i) \Rightarrow (ii) Sei $A \in \mathcal{F}_\infty$. Für alle $n \in \mathbb{N}$ gilt $X_t = E[X_{t+n} | \mathcal{F}_t]$, also

$$\left| \int_A X_\infty d\mathbb{P} - \int_A X_t d\mathbb{P} \right| = \left| \int_A X_\infty d\mathbb{P} - \int_A X_{t+n} d\mathbb{P} \right| \leq \int |X_\infty - X_{t+n}| d\mathbb{P}.$$

Mit $n \rightarrow \infty$ erhält man die Oberschranke 0, also $E[X_\infty | \mathcal{F}_t] = X_t$.

(ii) \Rightarrow (iii) ergibt sich mit Satz 4.4.

(iii) \Rightarrow (i) Betrachten wir nur Zeitpunkte $t \in \mathbb{N}$, so erhalten wir ein Martingal, auf das wir Satz 4.7 anwenden können, und somit eine bzgl.

$$\mathcal{F}_\infty = \sigma\left(\bigcup_{n \in \mathbb{N}} \mathcal{F}_n\right) = \sigma\left(\bigcup_{t \geq 0} \mathcal{F}_t\right)$$

messbare Zufallsvariable X_∞ , gegen die X_n mit $n \rightarrow \infty$ fast sicher und im Mittel konvergiert. Sei nun $(t_n)_{n \in \mathbb{N}} \subset [0, \infty)$ eine beliebige Folge mit $t_n \uparrow \infty$. Diese definiert wieder ein Martingal, auf das sich Satz 4.7 anwenden lässt. Kombiniert man die Folgen, so zeigt sich, dass die jeweiligen Grenzvaren X_∞ fast sicher übereinstimmen. Man erhält also für beliebige gegen ∞ konvergierende Folgen $(t_n)_{n \in \mathbb{N}}$, dass $\lim_{n \rightarrow \infty} E|X_{t_n} - X_\infty| = 0$ gilt. \square

Ist $(X_t)_{t \geq 0}$ L^p -beschränkt für ein $p > 1$, gilt also

$$\sup_{t \geq 0} \|X_t\|_p < \infty \quad \text{mit} \quad \|X_t\|_p := (E|X_t|^p)^{1/p},$$

so ist bekanntlich Bedingung (iii) erfüllt (Beispiel 4.2 (c)). Im Falle $p = 2$ hat man sogar L^2 -Konvergenz, also $E(X_t - X_\infty)^2 \rightarrow 0$ mit $t \rightarrow \infty$ sowie $EX_\infty^2 < \infty$ (Stochastik II, Aufgabe 82). Gilt (ii), so sagt man, dass das Martingal *sich rechts abschließen lässt*. Setzt man

$$\mathcal{F}_\infty := \sigma\left(\bigcup_{t \geq 0} \mathcal{F}_t\right),$$

so lässt sich X_∞ nämlich \mathcal{F}_∞ -messbar wählen, ist dann (\mathbb{P} -fast sicher) eindeutig und $(X_t, \mathcal{F}_t)_{0 \leq t \leq \infty}$ ist ein Martingal. Für beliebige, nicht notwendigerweise endliche Stoppzeiten τ definieren wir in dieser Situation X_τ in der naheliegenden Weise durch

$$X_\tau := \begin{cases} X_{\tau(\omega)}(\omega), & \text{falls } \tau(\omega) < \infty, \\ X_\infty(\omega), & \text{falls } \tau(\omega) = \infty. \end{cases}$$

Hat X rechtsstetige Pfade, so ist X_τ nach Lemma 3.14 (c) \mathcal{F}_τ -messbar, insbesondere also eine Zufallsvariable.

Neben Ungleichungen und Grenzwertsätzen sind Aussagen zum optionalen Stoppen besonders nützlich.

SATZ 4.9 *Ist $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein gleichgradig integrierbares Martingal mit rechtsstetigen Pfaden, so ist auch die Familie $\{X_\tau : \tau \text{ Stoppzeit}\}$ gleichgradig integrierbar, und für beliebige Stoppzeiten τ_1, τ_2 mit $\tau_1 < \tau_2$ gilt*

$$X_{\tau_1} = E[X_{\tau_2} | \mathcal{F}_{\tau_1}].$$

BEWEIS: Für jedes $n \in \mathbb{N}$ werde $\tau_n : \Omega \rightarrow [0, \infty]$ definiert durch

$$\tau_n(\omega) := \begin{cases} 2^{-n} \lceil 2^n \tau(\omega) \rceil, & \text{falls } \tau(\omega) \leq n, \\ \infty, & \text{sonst.} \end{cases}$$

Dies liefert eine Folge von Stoppzeiten mit jeweils endlichem Wertebereich (incl. ∞), die antiton gegen τ konvergiert. Für jedes feste n lässt sich das Optional Stopping Theorem (kurz: OST, Satz 8.12 der Vorlesung Stochastik II) auf $(Y_k, \mathcal{G}_k)_{k \in T_n}$ mit $T_n := \{0, 1, \dots, n2^n + 1\}$,

$$Y_k := X_{k2^{-n}}, \mathcal{G}_k := \mathcal{F}_{k2^{-n}} \text{ für } k = 0, \dots, n2^n, Y_{n2^n+1} := X_\infty, \mathcal{G}_{n2^n+1} := \mathcal{F}_\infty$$

anwenden und liefert $X_{\tau_n} = E[X_\infty | \mathcal{F}_{\tau_n}]$. Dies zeigt auch, dass $(X_{\tau_{-n}}, \mathcal{F}_{\tau_{-n}})$ ein Rückwärtsmartingal ist. Mit Satz 4.6 und der Stetigkeit von rechts der Pfade erhält man $X_\tau = E[X_{\tau_1} | \mathcal{F}_\tau]$, also insgesamt

$$X_\tau = E[X_\infty | \mathcal{F}_\tau]. \quad (1)$$

Da die Familie $\{E[X_\infty | \mathcal{G}] : \mathcal{G} \text{ Unter-}\sigma\text{-Algebra von } \mathcal{A}\}$ nach Satz 4.4 gleichgradig integrierbar ist, haben wir damit den ersten Teil der Behauptung bewiesen; der zweite folgt unmittelbar aus (1) und der Stapeleigenschaft bedingter Erwartungswerte. \square

Für das nächste Resultat verwenden wir die folgende Charakterisierung der Martingaleigenschaft, die Gegenstand von Aufgabe 38 ist: $(X_t, \mathcal{F}_t)_{t \geq 0}$ ist genau dann ein Martingal, wenn $E|X_\tau| < \infty$ und $EX_\tau = EX_0$ für alle beschränkten Stoppzeiten τ gilt.

KOROLLAR 4.10 *Es sei $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein Martingal mit rechtsstetigen Pfaden.*

(a) *Sind τ_1, τ_2 beschränkte Stoppzeiten mit $\tau_1 \leq \tau_2$, so gilt $X_{\tau_1} = E[X_{\tau_2} | \mathcal{F}_{\tau_1}]$. Insbesondere hat man $E|X_{\tau_1}| < \infty$.*

(b) *Ist τ eine beliebige Stoppzeit, so ist auch der bei τ gestoppte Prozess $X^\tau := (X_t^\tau)_{t \geq 0}$, $X_t^\tau := X_{\tau \wedge t}$, bezüglich $(\mathcal{F}_t)_{t \geq 0}$ ein Martingal mit rechtsstetigen Pfaden. Hat X stetige Pfade, so auch X^τ .*

BEWEIS: (a) Man sieht leicht, dass für jede reelle Zahl $K > 0$ der Prozess $(X_{t \wedge K}, \mathcal{F}_{t \wedge K})_{t \geq 0}$ ein gleichgradig integrierbares Martingal ist. Wählt man K so groß, dass $\tau_2 \leq K$ gilt, so erhält man die Aussage mit Satz 4.9.

(b) Wir wenden das oben genannte Martingalkriterium auf X^τ an; sei also ν eine beschränkte Stoppzeit. Dann ist auch $\tau \wedge \nu$ eine beschränkte Stoppzeit, und Teil (a) liefert mit $\tau_1 \equiv 0$, $\tau_2 := \tau \wedge \nu$

$$\begin{aligned} E|X_\nu^\tau| &= E|X_{\tau \wedge \nu}| = E|X_{\tau_2}| < \infty, \\ EX_\nu^\tau &= EX_{\tau \wedge \nu} = E(E[X_{\tau \wedge \nu} | \mathcal{F}_0]) = EX_{\tau_1} = EX_0. \end{aligned}$$

Die Aussagen zu den Pfaden überprüft man leicht. \square

Das ‘Einfrieren’ bei Stoppzeiten liefert auch eine Möglichkeit, den Martingalbegriff auf Familien von Zufallsvariablen auszudehnen, zu denen der Erwartungswert nicht existiert.

DEFINITION 4.11 Es sei $X = (X_t)_{t \geq 0}$ ein zu $(\mathcal{F}_t)_{t \geq 0}$ adaptierter Prozess. Dann heißt $(X_t, \mathcal{F}_t)_{t \geq 0}$ *lokales Martingal*, wenn es eine isotone, fast sicher

gegen ∞ konvergente Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten gibt derart, dass für alle $n \in \mathbb{N}$

$$(X_t^{\tau_n} - X_0, \mathcal{F}_t)_{t \geq 0}$$

ein Martingal ist. Man nennt dann $(\tau_n)_{n \in \mathbb{N}}$ eine *reduzierende* oder *lokalisierende Folge* zu X .

Die Idee, Eigenschaften stochastischer Prozesse durch Stoppzeiten zu lokalisieren, taucht in verschiedenen Zusammenhängen auf. Nennt man beispielsweise einen stochastischen Prozess $X = (X_t)_{t \geq 0}$ beschränkt, wenn es ein $K \in \mathbb{R}$ gibt mit $|X_t(\omega)| \leq K$ für alle $t \geq 0$, $\omega \in \Omega$ (also: X ist als Abbildung von $[0, \infty) \times \Omega$ nach \mathbb{R} beschränkt), so kann man ganz analog zu Definition 4.11 lokale Beschränktheit definieren. In diesem Sinne wäre beispielsweise die Brownsche Bewegung zwar nicht beschränkt, aber lokal beschränkt. Aufgrund der Stetigkeit der Pfade lässt sich bei einem stetigen lokalen Martingal eine lokalisierende Folge finden, für die die gestoppten Prozesse sogar beschränkte Martingale sind (Aufgabe 40 (a)). Gelegentlich muss man die Lokalisierung wieder loswerden, und hierfür ist das folgende Resultat nützlich.

SATZ 4.12 *Ein stetiges lokales Martingal $(X_t, \mathcal{F}_t)_{t \geq 0}$ ist genau dann ein Martingal, wenn für alle $t \geq 0$ die Familie aller Zufallsvariablen X_τ , wobei τ die durch t beschränkten Stoppzeiten durchläuft, gleichgradig integrierbar ist.*

BEWEIS: Bei der Notwendigkeit der Bedingung verwenden wir, wie bereits im Beweis zu Satz 4.9, dass die Familie der bedingten Erwartungswerte einer festen Zufallsvariablen unter allen Unter- σ -Algebren gleichgradig integrierbar ist.

Um nachzuweisen, dass die Bedingung auch hinreichend ist, verwenden wir das vor Korollar 4.10 genannte Martingalkriterium. Es seien also ν eine Stoppzeit mit $\nu \leq t_0$, $t_0 \in \mathbb{R}$ fest, und $(\tau_n)_{n \in \mathbb{N}}$ eine lokalisierende Folge zu $(X_t, \mathcal{F}_t)_{t \geq 0}$. Nach Korollar 4.10 (b) ist dann $(X_{\tau_n \wedge \nu \wedge t} - X_0, \mathcal{F}_t)_{t \geq 0}$ ein Martingal. Mit $\tau \equiv 0$ folgt $E|X_0| < \infty$ aus der gleichgradigen Integrierbarkeit von $\{X_\tau\}$, also ist auch $(X_{\tau_n \wedge \nu \wedge t}, \mathcal{F}_t)_{t \geq 0}$ ein Martingal. Damit gilt nach Korollar 4.10 (a)

$$EX_{\tau_n \wedge \nu} = EX_{\tau_n \wedge \nu \wedge t} = EX_0.$$

Die Stetigkeit der Pfade ergibt zusammen mit $\tau_n \rightarrow \infty$, dass $X_{\tau_n \wedge \nu}$ mit $n \rightarrow \infty$ gegen X_ν konvergiert, woraus aufgrund der gleichgradigen Integrierbarkeit von $\{X_{\tau_n \wedge \nu} : n \in \mathbb{N}\}$ die Konvergenz im Mittel, also $EX_{\tau_n \wedge \nu} \rightarrow EX_\nu$ folgt (Satz 4.5). Insgesamt hat man $EX_\nu = EX_0$, d.h. die Bedingung aus dem Martingalkriterium ist erfüllt. \square

4.3 FV-Prozesse, Vorhersehbarkeit und Klammerprozess. Bei bestimmten Integratoren lässt sich das übliche Maßintegral ‘punktweise’ anwen-

den. Wir nennen X einen FV-Prozess, wenn alle Pfade von X stetig von rechts und von lokal beschränkter Totalvariation sind, also

$$V_0^a X(\cdot, \omega) < \infty \quad \text{für alle } a > 0, \omega \in \Omega$$

gilt. Für eine (rechts)stetige Funktion f von lokal beschränkter Totalvariation mit $f(0) = 0$ sind die Funktionen

$$x \mapsto V_0^x f, \quad x \mapsto V_0^x f - f(x)$$

(rechts)stetig und monoton wachsend (Aufgabe 45), durch

$$\mu_+((0, x]) := V_0^x f, \quad \mu_-((0, x]) := V_0^x f - f(x) \quad \text{für alle } x > 0$$

werden also Maße μ_+, μ_- auf den Borelschen Teilmengen von $(0, \infty]$ definiert. Insbesondere für messbare, lokal (genauer: auf jedem Intervall $[0, t], t \geq 0$) beschränkte Funktionen $g : [0, \infty) \rightarrow \mathbb{R}$ ist dann das Integral über $(0, t]$ mit Integrand g und Integrator f durch

$$\int_{(0, t]} g df := \int_{(0, t]} g d\mu_+ - \int_{(0, t]} g d\mu_-$$

für alle $t \geq 0$ wohldefiniert.

BEISPIEL 4.13 Ist $N = (N_t)_{t \geq 0}$ ein Poisson-Prozess mit Intensität λ , so sind die Pfade von N stetig von rechts und stückweise konstant mit Sprüngen der Höhe 1 in den (Zeit-)Punkten $\xi_j, j \in \mathbb{N}$, wobei $(\xi_j)_{j \in \mathbb{N}}$ die Folge der Partialsummen zu einer Folge von unabhängigen, mit Parameter λ exponentialverteilten Zufallsvariablen ist. Offensichtlich gilt $V_0^t N = N_t$; das zu einem Pfad $t \mapsto N_t(\omega)$ gehörende Maß μ_+ ist also die Summe der Einpunktmaße in den Punkten $\xi_j(\omega), j \in \mathbb{N}$; μ_- verschwindet. Man erhält beispielsweise

$$\int_{(0, t]} s dN_s = \sum_{j=1}^{\infty} 1_{(0, t]}(\xi_j) \xi_j.$$

Dass es wichtige Fälle gibt, in denen diese Konstruktion nicht reicht, folgt aus den Resultaten von Abschnitt 3.3: Die Pfade einer Brownschen Bewegung $B = (B_t)_{t \geq 0}$ sind nicht von lokal beschränkter Totalvariation, B käme also als Integrator nicht in Frage. Wie der folgende Satz zeigt, tritt dieses Phänomen grundsätzlich bei nicht-trivialen Martingalen mit stetigen Pfaden auf.

SATZ 4.14 *Ist $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein stetiges lokales Martingal und sind \mathbb{P} -fast alle Pfade von X von lokal beschränkter Totalvariation, so gilt*

$$\mathbb{P}(X_t = X_0 \text{ für alle } t \geq 0) = 1.$$

BEWEIS: Angenommen, $(Y_t, \mathcal{F}_t)_{t \geq 0}$ ist ein stetiges Martingal mit $Y_0 = 0$, für das ein $C \in \mathbb{R}$ existiert mit der Eigenschaft

$$\mathbb{P}\left(\sup_{t \geq 0} V_0^t Y \leq C\right) = 1.$$

Es sei $0 = t_0 \leq t_1 \leq \dots \leq t_k = t$ eine Zerlegung des Intervalles $[0, t]$. Dann gilt

$$EY_t^2 = E \sum_{i=1}^k (Y_{t_i}^2 - Y_{t_{i-1}}^2) = E \sum_{i=1}^k (Y_{t_i} - Y_{t_{i-1}})^2,$$

denn wegen $EY_{t_i} Y_{t_{i-1}} = E(E[Y_{t_i} Y_{t_{i-1}} | \mathcal{F}_{t_{i-1}}]) = EY_{t_{i-1}}^2$ hat man

$$E(Y_{t_i}^2 - Y_{t_{i-1}}^2 - (Y_{t_i} - Y_{t_{i-1}})^2) = -2EY_{t_{i-1}}^2 + 2EY_{t_i} Y_{t_{i-1}} = 0.$$

Mit einer bereits im Beweis zu und im Anschluss an Satz 3.8 verwendeten Ungleichung, welche Variation, quadratische Variation und gleichmäßige Stetigkeit zueinander in Beziehung setzt, ergibt sich hieraus

$$EY_t^2 \leq E V_0^t Y \sup_{1 \leq i \leq k} |Y_{t_i} - Y_{t_{i-1}}| \leq C E \sup_{1 \leq i \leq k} |Y_{t_i} - Y_{t_{i-1}}|.$$

Bei feiner werdenden Zerlegungen von $[0, t]$ geht die Oberschranke nach dem Satz von der majorisierten Konvergenz gegen 0: Das Supremum konvergiert punktweise gegen 0, da die Pfade als stetige Funktionen auf dem kompakten Intervall $[0, t]$ gleichmäßig stetig sind, und $V_0^t Y$ ist eine integrierbare (da \mathbb{P} -fast sicher beschränkte) Majorante. Mit $Y_t^2 \geq 0$ folgt also $\mathbb{P}(Y_t = 0) = 1$. Da $t \geq 0$ beliebig war, bedeutet dies zusammen mit der Stetigkeit der Pfade, dass \mathbb{P} -fast alle Pfade von Y konstant 0 sind.

Es sei nun $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein Prozess, der den Annahmen des Satzes genügt. Eine Funktion $f : [0, \infty) \rightarrow \mathbb{R}$ ist offensichtlich dann konstant, wenn für eine Folge $(t_n)_{n \in \mathbb{N}}$ mit $t_n \rightarrow \infty$ die Einschränkungen von $t \mapsto f(t) - f(0)$ auf $[0, t_n]$, $n \in \mathbb{N}$, verschwinden. Nach Übergang zu $X_t - X_0$ und geeigneter Lokalisierung können wir also annehmen, dass $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein Martingal ist und dass $X_0 = 0$ gilt. Es sei $V_t := V_0^t X$ für alle $t \geq 0$; offensichtlich ist $(V_t)_{t \geq 0}$ zu $(\mathcal{F}_t)_{t \geq 0}$ adaptiert. Insbesondere wird durch

$$\tau_n := \inf\{t \geq 0 : V_t \geq n\}$$

eine Familie von Stoppzeiten definiert; da nach Voraussetzung \mathbb{P} -fast alle Pfade von X von lokal beschränkter Totalvariation sind, gilt $\tau_n \rightarrow \infty$ \mathbb{P} -fast sicher. Für jedes $n \in \mathbb{N}$ ist X^{τ_n} nach Korollar 4.10 (b) ein stetiges Martingal, nach Konstruktion haben die Pfade dieses Prozesses eine Gesamtvariation, die n nicht übersteigt. Wendet man nun die obige Argumentation auf $Y := X^{\tau_n}$ an und lässt n gegen ∞ gehen, so erhält man insgesamt die Behauptung. \square

Wir skizzieren das weitere Vorgehen: Ausgangspunkt ist wieder ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ mit einer Filtration $(\mathcal{F}_t)_{t \geq 0}$. Zu zwei zu $(\mathcal{F}_t)_{t \geq 0}$ adaptierten Prozessen $(H_t)_{t \geq 0}$, $(X_t)_{t \geq 0}$ wollen wir unter bestimmten Voraussetzungen an den *Integranden* H und den *Integrator* X das stochastische Integral

$$\int H dX = \int_0^\infty H_s dX_s$$

als Zufallsvariable auf $(\Omega, \mathcal{A}, \mathbb{P})$ konstruieren. Zur Vermeidung technischer Komplikationen (also aus didaktischen, nicht aus mathematischen Gründen) nehmen wir an, dass X stetige Pfade hat. Um von FV-Integratoren zu Martingalintegratoren zu kommen,

- (P1) geben wir die pfadweise Definition auf (das Integral wird als L^2 -Grenzwert von Zufallsvariablen definiert),
- (P2) schränken wir die Klasse der zugelassenen Integranden ein (auf vorhersehbare Prozesse, s.u.) und
- (P3) verwenden wir die quadratische Variation (als Substitut für die unbrauchbare Totalvariation).

Im Zusammenhang mit (P1) werden wir häufig einer Folge $(X^{(n)})_{n \in \mathbb{N}}$ von stochastischen Prozessen $X^{(n)} = (X_t^{(n)})_{t \geq 0}$ begegnen mit der Eigenschaft, dass für festes $t \geq 0$ die Folge $X_t^{(n)}$ mit $n \rightarrow \infty$ fast sicher oder in Wahrscheinlichkeit oder im quadratischen Mittel gegen eine Zufallsvariable X_t konvergiert. Um zu gewährleisten, dass mit $X^{(n)}$, $n \in \mathbb{N}$, auch $X = (X_t)_{t \geq 0}$ zu der vorgegebenen Filtration adaptiert ist, machen wir generell die folgende Annahme:

- (A) \mathcal{F}_0 enthält alle Nullmengen von $\mathcal{F}_\infty := \sigma(\bigcup_{t \geq 0} \mathcal{F}_t)$.

Beim Übergang zu Modifikationen bleibt Adaptiertheit dann erhalten (siehe hierzu auch Aufgabe 27 und Aufgabe 28).

Bei der Konstruktion des stochastischen Integrals betrachten wir, wie bei der Konstruktion des Lebesgue-Integrals, zunächst Integranden von besonders einfacher Struktur. Konkret nennen wir H einen *elementaren Prozess*, wenn es eine streng monoton wachsende Folge $(t_i)_{i \in \mathbb{N}_0}$ reeller Zahlen mit $t_0 = 0$, $\lim_{i \rightarrow \infty} t_i = \infty$ sowie \mathcal{F}_{t_i} -messbare Zufallsvariablen ξ_i , $i \in \mathbb{N}_0$, und schließlich

eine weitere, \mathcal{F}_0 -messbare Zufallsvariable ξ_{-1} gibt derart, dass

$$H_t(\omega) = 1_{\{0\}}(t)\xi_{-1}(\omega) + \sum_{i=0}^{\infty} 1_{(t_i, t_{i+1}]}(t)\xi_i(\omega) \quad (\text{EL})$$

für alle $t \geq 0$ und alle $\omega \in \Omega$ gilt. Wir schreiben gelegentlich \mathcal{E} für die Klasse der elementaren Prozesse. Im Zusammenhang mit (P2) verwenden wir nun:

SATZ UND DEFINITION 4.15 Die auf $[0, \infty) \times \Omega$ von

- den elementaren Prozessen,
- den adaptierten Prozessen mit linksstetigen Pfaden,
- den adaptierten Prozessen mit stetigen Pfaden

erzeugten σ -Algebren sind gleich. Man nennt die resultierende σ -Algebra \mathcal{P} die σ -Algebra der vorhersehbaren Mengen, und nennt einen \mathcal{P} -messbaren Prozess vorhersehbar (*Englisch*: predictable).

BEWEIS: Bezeichnet man die jeweiligen σ -Algebren temporär mit \mathcal{P}_1 , \mathcal{P}_2 und \mathcal{P}_3 , so folgt $\mathcal{P}_3 \subset \mathcal{P}_2$ unmittelbar aus der Definition. Ein Prozess X mit linksstetigen Pfaden ist punktweiser Limes der Prozesse $X^{(n)}$ mit $n \rightarrow \infty$, wobei

$$X_t^{(n)}(\omega) := 1_{\{0\}}(t)X_0(\omega) + \sum_{k=0}^{\infty} 1_{(k/n, (k+1)/n]}(t)X_{k/n}(\omega)$$

(eine ähnliche Konstruktion wurde bereits im Beweis zu Lemma 3.14 (b) verwendet), man hat also auch $\mathcal{P}_2 \subset \mathcal{P}_1$. Schließlich lässt sich ein Prozess der Form $(s, \omega) \mapsto 1_{(u, v]}(s)\xi(\omega)$ durch Prozesse $(s, \omega) \mapsto f_n(s)\xi(\omega)$ approximieren, wobei $(f_n)_{n \in \mathbb{N}}$ irgendeine Folge von stetigen Funktionen ist, die punktweise gegen $1_{(u, v]}$ konvergiert (man kann beispielsweise für f_n die stückweise lineare Funktion nehmen, die in den Punkten s , $s + 1/n$, t , $t + 1/n$ die Werte 0, 1, 1 und 0 annimmt). Mit diesem Argument erhält man die noch verbleibende Implikation $\mathcal{P}_1 \subset \mathcal{P}_3$. \square

Bei den Integranden H werden wir grundsätzlich Vorhersehbarkeit voraussetzen. Mit H sind auch alle Prozesse der Form $1_{[0, t]}H$, $t \geq 0$, vorhersehbar; insbesondere wird durch das stochastische Integral dem Integranden H und dem Integrator X ein weiterer stochastischer Prozess $H \bullet X$,

$$(H \bullet X)_t = \int 1_{[0, t]}H dX = \int_0^t H_s dX_s$$

zugeordnet. Oft ist mit dem stochastischen Integral nicht die einzelne Zufallsvariable $\int_0^\infty H_s dX_s$, sondern der gesamte Prozess $H \bullet X$ gemeint.

Es bleibt (P3): Beim stochastischen Integral spielt die quadratische Variation eine entscheidende Rolle. Wir wissen aus Satz 3.8, dass diese bei der Brownschen Bewegung lokal beschränkt ist, und wollen dies auf stetige lokale Martingale verallgemeinern. Hierzu sei \mathcal{Z} die Menge aller Folgen $\Delta = (t_i)_{i \in \mathbb{N}_0}$ reeller Zahlen mit den Eigenschaften

$$t_0 = 0, \quad t_{i-1} < t_i \quad \text{für alle } i \in \mathbb{N}, \quad \lim_{i \rightarrow \infty} t_i = \infty.$$

Wir identifizieren eine solche Folge mit der Zerlegung

$$[0, \infty) = \{0\} \cup \sum_{i=1}^{\infty} (t_{i-1}, t_i]$$

und nennen

$$|\Delta| := \sup_{i \in \mathbb{N}} |t_i - t_{i-1}|$$

die *Weite* der Zerlegung Δ . Jedes $\Delta \in \mathcal{Z}$ assoziiert mit einem Prozess $Y = (Y_t)_{t \geq 0}$ einen neuen Prozess $T^\Delta(Y)$, die *quadratische Prävariation von Y entlang Δ* ,

$$T^\Delta(Y)_t := \sum_{j=1}^{i(t)} (Y_{t_j} - Y_{t_{j-1}})^2 + (Y_t - Y_{t_{i(t)}})^2,$$

wobei $i(t) := \max\{j \in \mathbb{N}_0 : t_j \leq t\}$.

LEMMA 4.16 *Es sei $\Delta \in \mathcal{Z}$. Ist $(Y_t, \mathcal{F}_t)_{t \geq 0}$ ein beschränktes Martingal mit stetigen Pfaden, so ist $(Y_t^2 - T^\Delta(Y)_t, \mathcal{F}_t)_{t \geq 0}$ ein lokal beschränktes Martingal mit stetigen Pfaden.*

BEWEIS: Adaptiertheit sowie Stetigkeit und lokale Beschränktheit der Pfade sieht man sofort. Für beliebige $s, t, u \geq 0$ mit $s \leq t < u$ gilt

$$\begin{aligned} E[(Y_u - Y_s)^2 | \mathcal{F}_t] &= E[(Y_u - Y_t + Y_t - Y_s)^2 | \mathcal{F}_t] \\ &= E[(Y_u - Y_t)^2 | \mathcal{F}_t] + (Y_t - Y_s)^2, \end{aligned}$$

denn der gemischte Term verschwindet wieder wegen der Martingaleigenschaft (im Grunde ‘Pythagoras’ im L^2). Gilt $s < t$ und $t_i \leq s, t < t_{i+1}$, liegen also s und t im gleichen Zerlegungsintervall, so folgt hieraus

$$\begin{aligned} E[T^\Delta(Y)_t - T^\Delta(Y)_s | \mathcal{F}_s] &= E[(Y_t - Y_{t_i})^2 - (Y_s - Y_{t_i})^2 | \mathcal{F}_s] \\ &= E[(Y_t - Y_s)^2 | \mathcal{F}_s] \\ &= E[Y_t^2 - Y_s^2 | \mathcal{F}_s], \end{aligned}$$

also

$$E[Y_t^2 - T^\Delta(Y)_t | \mathcal{F}_s] = Y_s^2 - T^\Delta(Y)_s.$$

Im Falle $t_{l-1} \leq s < t_l \leq t_k \leq t < t_{k+1}$ erhält man analog

$$\begin{aligned} E[T^\Delta(Y)_t - T^\Delta(Y)_s | \mathcal{F}_s] &= E[(Y_t - Y_{t_k})^2 + \sum_{i=l}^k (Y_{t_i} - Y_{t_{i-1}})^2 - (Y_s - Y_{t_{l-1}})^2 | \mathcal{F}_s] \\ &= E[(Y_t - Y_{t_{l-1}})^2 - (Y_s - Y_{t_{l-1}})^2 | \mathcal{F}_s] \\ &= E[Y_t^2 - Y_s^2 | \mathcal{F}_s], \end{aligned}$$

also wieder

$$E[Y_t^2 - T^\Delta(Y)_t | \mathcal{F}_s] = Y_s^2 - T^\Delta(Y)_s. \quad \square$$

Die Aussage des Lemmas lässt vermuten, dass im Falle der Konvergenz von $T^\Delta(Y)$ gegen einen Prozess A mit $|\Delta| \rightarrow 0$ der Prozess $Y^2 - A$ ein Martingal ist.

SATZ 4.17 *Es sei $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein stetiges lokales Martingal mit $X_0 \equiv 0$. Dann existiert ein zu $(\mathcal{F}_t)_{t \geq 0}$ adaptierter Prozess $A = (A_t)_{t \geq 0}$ mit $A_0 \equiv 0$, stetigen und isotonen Pfaden und der Eigenschaft, dass $(X_t^2 - A_t, \mathcal{F}_t)_{t \geq 0}$ ein lokales Martingal ist. Der Prozess A ist durch diese Eigenschaften bis auf Ununterscheidbarkeit eindeutig bestimmt. Ist $(\Delta_n)_{n \in \mathbb{N}}$ eine Folge von Zerlegungen aus \mathcal{Z} mit $\lim_{n \rightarrow \infty} |\Delta_n| = 0$, so gilt*

$$T^{\Delta_n}(X)_t \rightarrow A_t \quad \text{in Wahrscheinlichkeit für alle } t \geq 0.$$

BEWEIS: Die Eindeutigkeit ergibt sich direkt aus Satz 4.14, denn die Differenz zweier Darstellungen würde auf ein stetiges Martingal mit Pfaden von lokal beschränkter Totalvariation führen.

Der Beweis der Existenz verläuft einigermaßen konstruktiv: Die im letzten Teil der Behauptung gegebene Darstellung des Klammerprozesses als quadratische Variation legt es nahe, diese durch die quadratische Prävariation entlang einer lokalendlichen Partition zu approximieren. Wir können annehmen, dass $(X_t, \mathcal{F}_t)_{t \geq 0}$ ein stetiges und beschränktes Martingal ist (ggf. nach Ausführung einer geeigneten Lokalisation); konkret gelte

$$X(t, \omega) \leq K \quad \text{für alle } t \geq 0, \omega \in \Omega, \quad (1)$$

mit einer geeigneten Konstanten $K < \infty$. Eine Zerlegung \mathcal{Z} ist automatisch lokalendlich, d.h.

$$\#\{i \in \mathbb{N}_0 : a_i \leq t\} < \infty \quad \text{für alle } t \geq 0.$$

Zu $\Delta, \Delta' \in \mathcal{Z}$ sei Δ'' die Zerlegung, die man über die Vereinigung der Punkte von Δ und Δ' erhält, und

$$0 = u_0 < u_1 < \dots < u_l \leq a < u_{l+1}$$

seien die für das Intervall $[0, a]$ relevanten Punkte von Δ'' . Die obige Hilfsaussage zeigt, dass $(Y_t, \mathcal{F}_t)_{t \geq 0}$ mit

$$Y_t := T^\Delta(X)_t - T^{\Delta'}(X)_t$$

ein Martingal ist, lokal beschränkt, mit stetigen Pfaden und Start in 0, also $Y_0 \equiv 0$. Aufgrund der Beschränktheit existieren alle Momente, und man erhält mit der bereits im Beweis zu Lemma 4.16 verwendeten Argumentation

$$\begin{aligned} E(T^\Delta(X)_a - T^{\Delta'}(X)_a)^2 &= EY_a^2 \\ &= E\left(\sum_{i=1}^l (Y_{u_i}^2 - Y_{u_{i-1}}^2) + (Y_a^2 - Y_{u_l}^2)\right) \\ &= ET^{\Delta''}(Y)_a. \end{aligned}$$

Mit elementaren Manipulationen und unter Verwendung von

$$(x - y)^2 \leq 2(x^2 + y^2)$$

folgt weiter

$$\begin{aligned} T^{\Delta''}(Y)_a &= \sum_{i=1}^l (Y_{u_i} - Y_{u_{i-1}})^2 + (Y_a - Y_{u_l})^2 \\ &= \sum_{i=1}^l \left((T^\Delta(X)_{u_i} - T^\Delta(X)_{u_{i-1}}) - (T^{\Delta'}(X)_{u_i} - T^{\Delta'}(X)_{u_{i-1}}) \right)^2 \\ &\quad + \left((T^\Delta(X)_a - T^\Delta(X)_{u_l}) - (T^{\Delta'}(X)_a - T^{\Delta'}(X)_{u_l}) \right)^2 \\ &\leq 2 \left(\sum_{i=1}^l (T^\Delta(X)_{u_i} - T^\Delta(X)_{u_{i-1}})^2 + (T^\Delta(X)_a - T^\Delta(X)_{u_l})^2 \right) \\ &\quad + \sum_{i=1}^l (T^{\Delta'}(X)_{u_i} - T^{\Delta'}(X)_{u_{i-1}})^2 + (T^{\Delta'}(X)_a - T^{\Delta'}(X)_{u_l})^2 \\ &= 2 \left(T^{\Delta''}(T^\Delta(X))_a + T^{\Delta''}(T^{\Delta'}(X))_a \right). \end{aligned}$$

Wir wollen zeigen, dass bei festem $a \geq 0$ die Zufallsgröße $T^\Delta(X)_a$ mit $|\Delta| \rightarrow 0$ in L^2 konvergiert. Dies würde aus den obigen Abschätzungen folgen, wenn wir noch

$$ET^{\Delta''}(T^\Delta(X))_a \rightarrow 0 \quad \text{mit } |\Delta| + |\Delta'| \rightarrow 0$$

zeigen können.

Zu jedem $i \in \{1, \dots, l\}$ existiert ein $k = k(i)$ mit

$$t_k \leq u_i < u_{i+1} \leq t_{k+1}, \quad t_k, t_{k+1} \in \Delta,$$

und man hat

$$\begin{aligned} T^\Delta(X)_{u_{i+1}} - T^\Delta(X)_{u_i} &= (X_{u_{i+1}} - X_{t_k})^2 - (X_{u_i} - X_{t_k})^2 \\ &= (X_{u_{i+1}} - X_{u_i})(X_{u_{i+1}} + X_{u_i} - 2X_{t_k}), \end{aligned}$$

woraus

$$T^{\Delta''}(T^\Delta(X))_a \leq T^{\Delta''}(X)_a \cdot \sup_{0 \leq i < l} |X_{u_{i+1}} + X_{u_i} - 2X_{t_k}|^2$$

folgt. Die Cauchy-Schwarz-Ungleichung liefert nun

$$ET^{\Delta''}(T^\Delta(X))_a \leq \left(E \sup_{0 \leq i < l} |X_{u_{i+1}} + X_{u_i} - 2X_{t_k}|^4 \right)^{1/2} \cdot \left(ET^{\Delta''}(X)_a^2 \right)^{1/2}.$$

Aus der Stetigkeit der Pfade von X und der Beschränktheitsannahme folgt wie im Beweis zu Satz 6.8, dass der erste Faktor mit $|\Delta| + |\Delta'| \rightarrow 0$ gegen 0 geht. Wir brauchen noch eine von der Zerlegung Δ unabhängige Oberschranke für $ET^{\Delta''}(X)_a^2$.

Sind $0 = t_0 < t_1 < \dots < t_{n-1} \leq a$ die in $[0, a]$ enthaltenen Punkte von Δ , und setzen wir (zur Vereinfachung der Schreibweise) $t_n := a$, so gilt

$$\begin{aligned} T^\Delta(X)_a^2 &= \left(\sum_{k=1}^n (X_{t_k} - X_{t_{k-1}}) \right)^2 \\ &= 2 \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^2 \sum_{j=k+1}^n (X_{t_j} - X_{t_{j-1}})^2 + \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^4 \\ &= 2 \sum_{k=1}^n (T^\Delta(X)_a - T^\Delta(X)_{t_k})(T^\Delta(X)_{t_k} - T^\Delta(X)_{t_{k-1}}) \\ &\quad + \sum_{k=1}^n (X_{t_k} - X_{t_{k-1}})^4. \end{aligned}$$

Aus der Martingaleigenschaft von $X^2 - T^\Delta(X)$ folgt mit den nun vertrauten Argumenten

$$E[T^\Delta(X)_a - T^\Delta(X)_{t_k} | \mathcal{F}_{t_k}] = E[(X_a - X_{t_k})^2 | \mathcal{F}_{t_k}]$$

und damit

$$\begin{aligned} ET^\Delta(X)_a^2 &= 2 \sum_{k=1}^n E(X_a - X_{t_k})^2 (T^\Delta(X)_{t_k} - T^\Delta(X)_{t_{k-1}}) \\ &\quad + \sum_{k=1}^n E(X_{t_k} - X_{t_{k-1}})^4 \\ &\leq E\left(T^\Delta(X)_a \left(2 \sup_k |X_a - X_{t_k}|^2 + \sup_k |X_{t_k} - X_{t_{k-1}}|^2\right)\right). \end{aligned}$$

Die Beschränktheitsannahme (1) liefert für den rechten Faktor im Erwartungswert des letzten Terms die Obergrenze $2(2K)^2 + (2K)^2 = 12K^2$. Mit der Martingaleigenschaft von $X^2 - T^\Delta(X)$ folgt außerdem

$$ET^\Delta(X)_a = EX_a^2 - E(X_a^2 - T^\Delta(X)_a) = EX_a^2 - EX_0^2 \leq 2K^2,$$

d.h. man hat

$$ET^\Delta(X)_a^2 \leq 24K^4.$$

Dies ist die gewünschte, von Δ unabhängige Obergrenze, mit der nun insgesamt die L^2 -Konvergenz von $T^\Delta(X)_a$ mit $|\Delta| \rightarrow 0$ folgt.

Sei nun $(\Delta_n)_{n \in \mathbb{N}}$ mit $\Delta_n = (t_{n,i})_{i \in \mathbb{N}_0}$ eine Folge von Elementen von \mathcal{Z} . Wir setzen voraus, dass die Zerlegungen feiner werden, dass also für alle $n \in \mathbb{N}$ $(t_{n,i})_{i \in \mathbb{N}_0}$ eine Teilfolge von $(t_{n+1,i})_{i \in \mathbb{N}_0}$ ist. Sei $Q := \{t_{n,i} : n \in \mathbb{N}, i \in \mathbb{N}_0\}$ die Menge aller Zerlegungspunkte der Folge $(\Delta_n)_{n \in \mathbb{N}}$. Für alle $a \in Q$, $\epsilon > 0$ erhält man mit der Doobschen Ungleichung (Satz 8.5 der Vorlesung Stochastik II)

$$\mathbb{P}\left(\sup_{0 \leq t \leq a} |T_t^{\Delta_n} - T_t^{\Delta_m}| \geq \epsilon\right) \leq \frac{4}{\epsilon^2} E(T_a^{\Delta_n} - T_a^{\Delta_m})^2 \rightarrow 0$$

mit $m, n \rightarrow \infty$. Die zufälligen Funktionen $t \rightarrow T_t^{\Delta_n}$, $n \in \mathbb{N}$, bilden also eine Cauchy-Folge bzgl. der Konvergenz in Wahrscheinlichkeit in dem vollständigen metrischen Raum $(C[a, b], \|\cdot\|_{a, \infty})$ der stetigen reellwertigen Funktionen auf dem Intervall $[0, a]$, versehen mit der zugehörigen Supremumsnorm. Es ist aus den Stochastikvorlesungen bekannt, dass dann eine geeignete Teilfolge fast sicher konvergiert, sei $Y^a = (Y_t^a)_{0 \leq t \leq a}$ der zugehörige Limes. Die Pfade von Y^a sind Grenzwerte von gleichmäßig konvergenten Folgen stetiger Funktionen, also stetig. Die Verfeinerungsannahme impliziert, dass für alle $s, t \in Q \cap [0, a]$ mit $s \leq t$ die Ungleichung $T_s^{\Delta_n} \leq T_t^{\Delta_n}$ für alle hinreichen großen $n \in \mathbb{N}$

gilt. Da $Q \cap [0, a]$ dicht in $[0, a]$ liegt, sind somit die Pfade von Y^a schwach monoton wachsend (beachte: die Pfade von T^Δ sind im allgemeinen *nicht* schwach monoton wachsend). Die obige Hilfsaussage ergibt, zusammen mit der L^2 -Konvergenz von $T_t^{\Delta_n}$ gegen Y_t^a , dass auch $(X_t^2 - Y_t^a, \mathcal{F}_t)_{0 \leq t \leq a}$ ein Martingal ist.

Es ist nur noch zu zeigen, dass die für verschiedene $a \in Q$ erhaltenen Y -Prozesse zueinander kompatibel sind. Aus der Konstruktion folgt jedoch sofort, dass für alle $a, b, t \in Q$ mit $t \leq a \wedge b$ $\mathbb{P}(Y_t^a = Y_t^b) = 1$ gilt. Da \mathbb{P} -fast alle Pfade der beteiligten Prozesse stetig sind und da Q abzählbar ist und in $[0, \infty)$ dicht liegt, folgt

$$\mathbb{P}(Y_t^a = Y_t^b \text{ für alle } t \leq a \wedge b) = 1.$$

Dieselbe Argumentation zeigt auch, dass die zu den Lokalisierungen des Ausgangsprozesses erhaltenen Y -Prozesse zueinander kompatibel sind. \square

Aus Satz 3.8 ist bekannt, dass man bei der Brownschen Bewegung den (deterministischen) Prozess $(A_t)_{t \geq 0}$ mit $A_t \equiv t$ erhält.

DEFINITION 4.18 Man nennt den Prozess A aus Satz 4.17 den *Klammerprozess* oder auch die *quadratische Variation* zu X und verwendet die Schreibweise $A = \langle X \rangle$. Ist $(Y_t, \mathcal{F}_t)_{t \geq 0}$ ein weiteres stetiges lokales Martingal mit Start in 0, so definieren wir die *quadratische Kovariation* $\langle X, Y \rangle$ von X und Y durch

$$\langle X, Y \rangle := \frac{1}{2}(\langle X + Y \rangle - \langle X \rangle - \langle Y \rangle).$$

Man sieht leicht, dass $\langle X, Y \rangle$ ein adaptierter Prozess mit stetigen Pfaden ist, und dass $(X_t Y_t - \langle X, Y \rangle_t, \mathcal{F}_t)_{t \geq 0}$ ein lokales Martingal ist. Die Pfade von $\langle X, Y \rangle$ sind zwar im allgemeinen nicht isoton, aber als Linearkombination isotoner Funktionen von lokal beschränkter Totalvariation. Man sieht weiter leicht, dass die Zuordnung $(X, Y) \mapsto \langle X, Y \rangle$ bilinear und symmetrisch ist, und dass $\langle X \rangle = \langle X, X \rangle$ gilt. (In der Tat ergibt sich die Definition der quadratischen Kovariation aus der Definition der quadratischen Variation zwangsläufig über eine Formel der in Definition 4.18 verwendeten Art, wenn man diese Eigenschaften haben will; dieser ‘Trick’ wird auch *Polarisation* genannt.) Das Zusammenspiel dieser Abbildung mit dem ‘Einfrieren’ von Prozessen wird in den Übungsaufgaben behandelt.

Für die Konstruktion des stochastischen Integrals ist eine bestimmte Klasse von Martingalen nützlich. Es sei \mathcal{M}_0^2 die Menge aller L^2 -beschränkten Martingale $X = (X_t, \mathcal{F}_t)_{t \geq 0}$ mit Start in 0, also mit

$$X_0 \equiv 0, \quad \sup_{t \geq 0} EX_t^2 < \infty.$$

Zu jedem solchen X existiert nach Satz 6.3 und den anschließenden Bemerkungen eine (modulo \mathbb{P} -Nullmengen eindeutige) \mathcal{F}_∞ -messbare Zufallsvariable X_∞ mit $EX_\infty^2 < \infty$, $X_t \rightarrow X_\infty$ \mathbb{P} -fast sicher und in L^2 , sowie $X_t = E[X_\infty | \mathcal{F}_t]$ für alle $t \geq 0$. Da der zugehörige Klammerprozess isotone Pfade hat, können wir $\langle X \rangle$ ebenfalls ‘rechts abschliessen’, und zwar durch $\langle X \rangle_\infty := \sup_{t \geq 0} \langle X \rangle_t$. Bei allgemeinen lokalen Martingalen (beispielsweise bei der Brownschen Bewegung) kann dieses Supremum allerdings durchaus mit positiver Wahrscheinlichkeit den Wert ∞ annehmen.

LEMMA 4.19 *Es sei $X \in \mathcal{M}_0^2$. Dann ist $X^2 - \langle X \rangle$ ein gleichgradig integrierbares Martingal und es gilt $E\langle X \rangle_\infty = EX_\infty^2 < \infty$, insbesondere also*

$$\mathbb{P}(\langle X \rangle_\infty < \infty) = 1.$$

BEWEIS: Nach Satz 4.17 ist $X^2 - \langle X \rangle$ ein lokales Martingal; sei $(\tau_n)_{n \in \mathbb{N}}$ eine reduzierende Folge. Wir können annehmen, dass die Martingale $(X^2 - \langle X \rangle)^{\tau_n}$ beschränkt, also insbesondere gleichgradig integrierbar sind. Dann gilt nach Satz 4.10

$$(X^2 - \langle X \rangle)_t^{\tau_n} \rightarrow (X^2 - \langle X \rangle)_{\tau_n} \quad \text{in } L^1$$

mit $t \rightarrow \infty$, also insbesondere

$$EX_{\tau_n}^2 = E\langle X \rangle_{\tau_n} \quad \text{für alle } n \in \mathbb{N}. \quad (*)$$

Die Jensensche Ungleichung für bedingte Erwartungswerte ergibt in Verbindung mit Satz 4.11

$$0 \leq X_{\tau_n}^2 = (E[X_\infty | \mathcal{F}_{\tau_n}])^2 \leq E[X_\infty^2 | \mathcal{F}_{\tau_n}].$$

Hieraus wiederum folgt, dass $\{X_{\tau_n}^2 : n \in \mathbb{N}\}$ gleichgradig integrierbar ist, d.h. es gilt $X_{\tau_n} \rightarrow X_\infty$ in L^2 . In Verbindung mit (*) und dem Satz von der monotonen Konvergenz liefert dies $E\langle X \rangle_\infty < \infty$ (und damit die zweite Teilaussage) sowie $\langle X \rangle_t \rightarrow \langle X \rangle_\infty$ in L^1 , also insgesamt

$$(X^2 - \langle X \rangle)_{\tau_n} \rightarrow (X^2 - \langle X \rangle)_\infty \quad \text{in } L^1$$

mit $n \rightarrow \infty$. Mit der L^1 -Stetigkeit der Abbildung $Y \mapsto E[Y | \mathcal{F}_t]$ (nach der Jensenschen Ungleichung ist dies sogar eine schwache Kontraktion) ergibt sich hieraus

$$(X^2 - \langle X \rangle)_{\tau_n \wedge t} \rightarrow E[(X^2 - \langle X \rangle)_\infty | \mathcal{F}_t] \quad \text{in } L^1$$

mit $n \rightarrow \infty$, also insgesamt

$$(X^2 - \langle X \rangle)_t = E[X_\infty^2 - \langle X \rangle_\infty | \mathcal{F}_t] \quad \text{für alle } t \geq 0.$$

Nach Satz 4.10 ist somit $X^2 - \langle X \rangle$ ein gleichgradig integrierbares Martingal. \square

Zu jedem $X \in \mathcal{M}_0^2$ gehört also ein $X_\infty \in L^2$. Die Menge der quadratisch integrierbaren Zufallsvariablen auf $(\Omega, \mathcal{F}_\infty, \mathbb{P})$ bildet, nach ‘Herausdividieren der Nullmengen’, einen Banach-Raum, sogar einen Hilbert-Raum. Was bedeutet die Konvergenz der abschließenden Variablen für die zugehörigen Martingale?

LEMMA 4.20 *Auf \mathcal{M}_0^2 wird durch*

$$\|X\|_{\mathcal{M}} := \|X_\infty\|_2 \quad \left(= (EX_\infty^2)^{1/2} = (E\langle X \rangle_\infty)^{1/2} \right)$$

eine Seminorm definiert. Jede Cauchy-Folge bezüglich dieser Seminorm hat einen bis auf Ununterscheidbarkeit eindeutigen Grenzwert in \mathcal{M}_0^2 .

BEWEIS: Man hat

$$(\alpha X)_\infty = \alpha X_\infty, \quad (X + Y)_\infty = X_\infty + Y_\infty \quad \mathbb{P}\text{-fast sicher,}$$

woraus leicht folgt, dass $\|\cdot\|_{\mathcal{M}}$ eine Seminorm auf \mathcal{M}_0^2 ist. Es sei nun $(X^{(n)})_{n \in \mathbb{N}}$ eine Cauchy-Folge in $(\mathcal{M}_0^2, \|\cdot\|_{\mathcal{M}})$. Nach Definition der Seminorm ist dann die Folge $(X_\infty^{(n)})_{n \in \mathbb{N}}$ der abschließenden Variablen eine Cauchy-Folge in $\mathcal{L}^2(\Omega, \mathcal{F}_\infty, \mathbb{P})$, d.h. es gibt eine Zufallsvariable X_∞ mit $X_\infty^{(n)} \rightarrow X_\infty$ in $\|\cdot\|_2$. Die behauptete Vollständigkeit von \mathcal{M}_0^2 bezüglich $\|\cdot\|_{\mathcal{M}}$ folgt nun, wenn gezeigt werden kann, dass X_∞ die abschließende Variable eines quadratisch integrierbaren Martingals X mit stetigen Pfaden und $X_0 \equiv 0$ ist. Wir setzen $Y_t := E[X_\infty | \mathcal{F}_t]$ für alle $t \in \mathbb{Q}$, sei

$$\xi_{n,m} := \sup_{t \in \mathbb{Q} \cap [0,m]} |X_t^{(n)} - Y_t|, \quad m, n \in \mathbb{N}.$$

Da $(X_t^{(n)} - Y_t, \mathcal{F}_t)_{t \in \mathbb{Q}}$ ein Martingal ist, folgt mit einer geeigneten Variante der Doobischen Ungleichung (siehe beispielsweise Aufgabe 40) und der Jensenschen Ungleichung

$$\begin{aligned} \mathbb{P}(\xi_{n,m} \geq \epsilon) &\leq \frac{1}{\epsilon^2} E(X_m^{(n)} - Y_m)^2 \\ &= \frac{1}{\epsilon^2} E(E[X_\infty^{(n)} - X_\infty | \mathcal{F}_m])^2 \\ &\leq \frac{1}{\epsilon^2} E(E[(X_\infty^{(n)} - X_\infty)^2 | \mathcal{F}_m]) \\ &= \frac{1}{\epsilon^2} E(X_\infty^{(n)} - X_\infty)^2 \\ &\rightarrow 0 \quad \text{mit } n \rightarrow \infty, \end{aligned}$$

$\xi_{n,m}$ konvergiert also bei festem $m \in \mathbb{N}$ mit $n \rightarrow \infty$ in Wahrscheinlichkeit gegen 0. Nach Übergang zu einer Teilfolge erhält man die fast sichere Konvergenz und damit die gleichmäßige (!) Stetigkeit der Pfade $t \mapsto Y_t(\omega)$ auf $[0, m] \cap \mathbb{Q}$ für alle ω außerhalb einer \mathbb{P} -Nullmenge N_m . Ein Standardargument aus der Analysis zeigt, dass dann für alle $\omega \notin N := \bigcup_{m=1}^{\infty} N_m$ die Funktion $t \mapsto Y_t(\omega)$ zu einer stetigen Funktion $t \mapsto X_t(\omega)$ fortgesetzt werden kann; auf N setzen wir $X \equiv 0$ (hier ist beispielsweise unsere Annahme, dass \mathcal{F}_0 alle \mathbb{P} -Nullmengen von \mathcal{A} enthält, nützlich, sonst würde hierbei u.U. ein nicht mehr zu $(\mathcal{F}_t)_{t \geq 0}$ adaptierter Prozess herauskommen). Nach Konstruktion gilt

$$X_t = E[X_\infty | \mathcal{F}_t] \quad \text{für alle } t \in \mathbb{Q} \cap [0, \infty).$$

Der Rückwärtskonvergenzsatz für Martingale zeigt, dass dies dann sogar für alle $t \geq 0$ gilt; insbesondere ist X ein Martingal. Man hat auch

$$EX_0^2 = E(X_n^{(0)} - X_0)^2 \leq E(X_\infty^{(n)} - X_\infty)^2 \rightarrow 0 \quad \text{mit } n \rightarrow \infty,$$

also $X_0 = 0$ \mathbb{P} -fast sicher. □

Würde man, ähnlich wie in der Vorlesung Stochastik II beim Übergang von \mathcal{L}^p zu L^p , ununterscheidbare Prozesse identifizieren, so würde die Seminorm zur Norm und man erhielte einen Banach-Raum (dass $\|X - Y\|_{\mathcal{M}} = 0$ zur Ununterscheidbarkeit Prozesse X und Y äquivalent ist, ist Gegenstand einer Übungsaufgabe). Wesentlich ist, dass man in den jeweiligen Äquivalenzklassen einen Vertreter mit stetigen Pfaden findet.

4.4 Konstruktion des stochastischen Integrals. Wir definieren zunächst das elementare stochastische Integral $H \bullet X$ für Integranden $X \in \mathcal{M}_0^2$ und beschränkte elementare Integranden, also Prozesse der Form

$$H_t(\omega) = 1_{\{0\}}(t)\xi_{-1}(\omega) + \sum_{i=0}^{\infty} 1_{(t_i, t_{i+1}]}(t)\xi_i(\omega) \quad (\text{EL})$$

(siehe Abschnitt 4.3), die zusätzlich mit einem geeigneten $K \in \mathbb{R}$ der Bedingung

$$|H(t, \omega)| \leq K \quad \text{für alle } t \geq 0, \omega \in \Omega$$

genügen; wir schreiben $b\mathcal{E}$ für die Klasse dieser beschränkten elementaren Prozesse. Bei den jeweiligen Konstruktionsschritten wird der Klammerprozess $\langle X \rangle$ zum Integrator eine wichtige Rolle spielen. Da dieser stetige und isotone Pfade hat, also ein FV-Prozess ist, kann das Integral mit Integrator $\langle X \rangle$ wie in Abschnitt 4.3 ausgeführt werden.

SATZ UND DEFINITION 4.21 Zu $X \in \mathcal{M}_0^2$, $H \in b\mathcal{E}$ von der Form (EL) definieren wir das elementare stochastische Integral $H \bullet X$ durch

$$H \bullet X := \sum_{i=0}^{\infty} \xi_i (X^{t_{i+1}} - X^{t_i}).$$

Der Prozess $H \bullet X$ hängt nicht von der Wahl der Darstellung (EL) ab. Es gilt $H \bullet X \in \mathcal{M}_0^2$ sowie

$$E(H \bullet X)_{\infty}^2 = E \int_0^{\infty} H_t^2 d\langle X \rangle_t. \quad (\text{iso}_0)$$

BEWEIS: Ist $t_2 \in (t_1, t_3)$, so folgt aus

$$\xi_0 \mathbf{1}_{(t_1, t_2]} + \xi_1 \mathbf{1}_{(t_2, t_3]} = \xi_2 \mathbf{1}_{(t_1, t_3]}$$

durch Einsetzen von $t = t_2$, $t = t_3$, dass $\xi_0 = \xi_1 = \xi_2$ gilt und damit

$$\xi_0 (X_{t_2 \wedge t} - X_{t_1 \wedge t}) + \xi_1 (X_{t_3 \wedge t} - X_{t_2 \wedge t}) = \xi_2 (X_{t_3 \wedge t} - X_{t_1 \wedge t}).$$

Mit der (offensichtlichen) Linearität von $H \mapsto H \bullet X$ zeigt dies, dass zwei Darstellungen der Form (EL) mit Zerlegungen Δ, Δ' auf dasselbe $H \bullet X$ führen wie die Darstellung mit der kleinsten gemeinsamen Verfeinerung Δ'' .

Es seien nun $s, t \geq 0$ mit $s \leq t$. Dann existieren $l, m \in \mathbb{N}_0$ mit $l \leq m$ und

$$t_l \leq s < t_{l+1}, \quad t_m \leq t < t_{m+1},$$

und die Definition von $H \bullet X$ liefert

$$\begin{aligned} E[(H \bullet X)_t | \mathcal{F}_s] &= \sum_{k=0}^{l-1} E[\xi_k (X_{t_{k+1}} - X_{t_k}) | \mathcal{F}_s] + E[\xi_l (X_{t_{l+1}} - X_{t_l}) | \mathcal{F}_s] \\ &\quad + \sum_{k=l+1}^{m-1} E[\xi_k (X_{t_{k+1}} - X_{t_k}) | \mathcal{F}_s] + E[\xi_m (X_t - X_{t_m}) | \mathcal{F}_s]. \end{aligned}$$

Beim ersten Term auf der rechten Seite erhält man $\sum_{k=0}^{l-1} \xi_k (X_{t_{k+1}} - X_{t_k})$, da wegen $t_l \leq s$ die beteiligten Zufallsvariablen \mathcal{F}_s -messbar sind. Beim zweiten Term erhalten wir

$$\begin{aligned} E[\xi_l (X_{t_{l+1}} - X_{t_l}) | \mathcal{F}_s] &= E[\xi_l (X_{t_{l+1}} - X_s) | \mathcal{F}_s] + E[\xi_l (X_s - X_{t_l}) | \mathcal{F}_s] \\ &= \xi_l (E[X_{t_{l+1}} | \mathcal{F}_s] - X_s) + \xi_l (X_s - X_{t_l}) \\ &= \xi_l (X_s - X_{t_l}). \end{aligned}$$

Die Summe dieser beiden Terme ist also gerade $(H \bullet X)_s$, d.h. $H \bullet X$ ist ein Martingal, wenn der dritte und der vierte Term verschwinden. Man erhält für $l+1 \leq k \leq m$

$$E[\xi_k(X_{t_{k+1}} - X_{t_k}) | \mathcal{F}_s] = E[E[\xi_k(X_{t_{k+1}} - X_{t_k}) | \mathcal{F}_{t_k}] | \mathcal{F}_s] = 0,$$

und erledigt den vierten Term analog.

Damit ist also gezeigt, dass $H \bullet X$ ein Martingal ist. Nach Konstruktion sind die Pfade stetig, und auch $(H \bullet X)_0 \equiv 0$ ist direkt zu sehen. Für alle $t \geq 0$ erhält man, wenn man l so wählt, dass $t_l \leq t < t_{l+1}$ gilt,

$$\begin{aligned} E(H \bullet X)_t^2 &= E\left(\sum_{k=0}^{l-1} \xi_k(X_{t_{k+1}} - X_{t_k}) + \xi_l(X_t - X_{t_l})\right)^2 \\ &= \sum_{k=0}^{l-1} E\xi_k^2(X_{t_{k+1}} - X_{t_k})^2 + E\xi_l^2(X_t - X_{t_l})^2, \end{aligned}$$

denn die gemischten Terme verschwinden wieder. Nach Lemma 4.12 ist der Prozess $X^2 - \langle X \rangle$ ein Martingal, also folgt weiter

$$\begin{aligned} E\xi_k^2(X_{t_{k+1}} - X_{t_k})^2 &= E(\xi_k^2 E[(X_{t_{k+1}} - X_{t_k})^2 | \mathcal{F}_{t_k}]) \\ &= E(\xi_k^2 (E[X_{t_{k+1}}^2 | \mathcal{F}_{t_k}] - X_{t_k}^2)) \\ &= E\left(\xi_k^2 \left(E[X_{t_{k+1}}^2 - \langle X \rangle_{t_{k+1}} | \mathcal{F}_{t_k}] \right. \right. \\ &\quad \left. \left. - (X_{t_k}^2 - \langle X \rangle_{t_k}) + E[\langle X \rangle_{t_{k+1}} - \langle X \rangle_{t_k} | \mathcal{F}_{t_k}]\right)\right) \\ &= E\xi_k^2(\langle X \rangle_{t_{k+1}} - \langle X \rangle_{t_k}) \end{aligned}$$

und damit

$$E(H \bullet X)_t^2 = E \sum_{k=0}^{\infty} \xi_k^2(\langle X \rangle_{t_{k+1} \wedge t} - \langle X \rangle_{t_k \wedge t}) = E \int_0^t H_s^2 d\langle X \rangle_s \quad (*)$$

für alle $t \geq 0$, wobei sich die zweite Gleichheit daraus ergibt, dass H stückweise konstante Pfade hat. Hieraus können nun die verbleibenden Aussagen gefolgert werden: Hat man $|H| \leq K$, so folgt

$$E(H \bullet X)_t^2 \leq K^2 E\langle X \rangle_t \leq K^2 E\langle X \rangle_\infty < \infty,$$

d.h. der Prozess $H \bullet X$ ist L^2 -beschränkt. Dies wiederum impliziert die Konvergenz von $(H \bullet X)_t$ mit $t \rightarrow \infty$ gegen $(H \bullet X)_\infty$ in L^2 , also

$$E(H \bullet X)_\infty^2 = \lim_{t \rightarrow \infty} E(H \bullet X)_t^2.$$

Der Satz von der monotonen Konvergenz liefert

$$E \int_0^\infty H_t^2 d\langle X \rangle_t = \lim_{t \rightarrow \infty} E \int_0^t H_s^2 d\langle X \rangle_s,$$

also folgt (iso₀) durch Grenzübergang aus (*). □

Man sieht, dass für das elementare stochastische Integral der Wert des Integranden in 0 keine Rolle spielt ($\int_0^t \dots = \int_{(0,t]} \dots$). Bei der Fortsetzung des Integrals auf interessantere Integranden verwenden wir die in (iso₀) enthaltene Isometrie: Zu $H \in b\mathcal{E}$, $X \in \mathcal{M}_0^2$ erhalten wir die Zufallsvariablen

$$\left(\int_0^\infty H_t^2 d\langle X \rangle_t \right)^{1/2} \quad \text{und} \quad (H \bullet X)_\infty.$$

Beide sind quadratisch integrierbar und haben dieselbe L^2 -Norm, für die wir $\|H\|_X$ schreiben, also

$$\|H\|_X := \left(E \int_0^\infty H_t^2 d\langle X \rangle_t \right)^{1/2} = (E(H \bullet X)_\infty^2)^{1/2}.$$

Darüberhinaus ist $H \bullet X$ ein Element des vollständigen metrischen Raums \mathcal{M}_0^2 , siehe Lemma 4.13. Nun beinhaltet die linke Zufallsvariable ja nur das gewöhnliche Integral, lässt sich also ohne weiteres auch für nicht-elementare Prozesse H definieren. Die Isometrie ermöglicht uns, eine Fortsetzung auf dieser Seite auf die uns eigentlich interessierende rechte Seite zu transferieren. Hierbei ist die folgende Hilfsaussage nützlich.

LEMMA 4.22 *Es seien $X \in \mathcal{M}_0^2$ und H ein vorhersehbarer Prozess mit*

$$E \int_0^\infty H_t^2 d\langle X \rangle_t < \infty.$$

Dann existiert eine Folge $(H^{(n)})_{n \in \mathbb{N}}$ in $b\mathcal{E}$ mit der Eigenschaft

$$\lim_{n \rightarrow \infty} \|H^{(n)} - H\|_X = 0.$$

BEWEIS: Setzt man $\tilde{H}^{(n)} := H \wedge n$, so erhält man $\lim_{n \rightarrow \infty} \|\tilde{H}^{(n)} - H\|_X = 0$ mit dem Satz von der monotonen Konvergenz; es reicht also, die Aussage für beschränkte Prozesse H zu beweisen. Es sei \mathcal{H} die Menge aller vorhersehbaren Prozesse, die sich bezüglich $\|\cdot\|_X$ durch eine Folge aus $b\mathcal{E}$ approximieren lassen. Wir behaupten, dass \mathcal{H} die folgenden Eigenschaften hat:

- (i) \mathcal{H} ist ein Vektorraum,
- (ii) die Funktion $H \equiv 1$ ist ein Element von \mathcal{H} ,
- (iii) $1_E \in \mathcal{H}$, wobei E ein durchschnittsstabiles Erzeugendensystem der σ -Algebra \mathcal{P} der vorhersehbaren Mengen durchläuft,
- (iv) existiert zu einem beschränkten Prozess H eine isotone, nicht-negative Folge $(H^{(n)})_{n \in \mathbb{N}}$ in \mathcal{H} mit $H^{(n)} \uparrow H$, so gilt $H \in \mathcal{H}$.

Ein Hilfsmittel aus der Maßtheorie (Aufgabe 47) impliziert, dass dann \mathcal{H} alle beschränkten vorhersehbaren Prozesse enthält.

Beim Nachweis von (i) verwendet man, dass $b\mathcal{E}$ ein Vektorraum ist; (ii) ist trivial, da $H \equiv 1$ ein Element von $b\mathcal{E}$ ist. Ein im Zusammenhang mit (iii) geeignetes Erzeugendensystem wird von den vorhersehbaren Rechtecken

$$(s, t] \times A, \quad 0 \leq s < t < \infty, \quad A \in \mathcal{F}_s$$

geliefert; bei (iv) kann wieder der Satz von der monotonen Konvergenz verwendet werden. \square

SATZ UND DEFINITION 4.23 *Es seien $X \in \mathcal{M}_0^2$ und H ein vorhersehbarer Prozess mit $E \int H_t^2 d\langle X \rangle_t < \infty$. Weiter sei $(H^{(n)})_{n \in \mathbb{N}} \subset b\mathcal{E}$ eine Folge mit $\lim_{n \rightarrow \infty} \|H^{(n)} - H\|_X = 0$. Dann ist $(H^{(n)} \bullet X)_{n \in \mathbb{N}}$ eine Cauchy-Folge in $(\mathcal{M}_0^2, \|\cdot\|_{\mathcal{M}})$. Der Grenzwert ist das stochastische Integral $H \bullet X$ von H bezüglich X und hängt, bis auf Ununterscheidbarkeit, nicht von der Wahl der Folge $(H^{(n)})_{n \in \mathbb{N}}$ ab. Außerdem gilt*

$$E(H \bullet X)_\infty^2 = E \int_0^\infty H_t^2 d\langle X \rangle_t. \quad (\text{iso})$$

BEWEIS: Die Abbildung $H \mapsto H \bullet X$ ist (bei festem $X \in \mathcal{M}_0^2$) wegen der Eigenschaft (iso₀) aus Satz 4.14 eine Isometrie von dem normierten Raum $(b\mathcal{E}, \|\cdot\|_X)$ in den normierten Raum $(\mathcal{M}_0^2, \|\cdot\|_{\mathcal{M}})$. Standardargumente der Topologie und Funktionalanalysis zeigen, dass eine solche Isometrie auf den in Lemma 4.22 beschriebenen Abschluss von $b\mathcal{E}$ fortgesetzt werden kann. (Eine Isometrie ist gleichmäßig stetig; es geht also um das bereits im Beweis zu Lemma 4.20 verwendete Argument.) \square

Im nächsten Schritt wollen wir durch Lokalisieren die Bedingung $X \in \mathcal{M}_0^2$ abschwächen (man beachte, dass beispielsweise die Brownsche Bewegung dieser Bedingung nicht genügt). In diesem Zusammenhang ist das Verhalten des stochastischen Integrals beim Stoppen des Integrators und beim Abschneiden des Integranden von Bedeutung. Wir verallgemeinern zunächst die Formel

$$(\xi_{1(u,v]}) \bullet X = \xi(X^v - X^u), \quad 0 \leq u < v < \infty, \quad \xi \mathcal{F}_u\text{-messbar},$$

von festen Zeiten u, v auf Stoppzeiten.

LEMMA 4.24 *Es seien $X \in \mathcal{M}_0^2$, ν und τ Stoppzeiten mit $\nu \leq \tau < \infty$ sowie ξ eine beschränkte, \mathcal{F}_ν -messbare Zufallsvariable. Dann gilt*

$$(\xi_{1(\nu,\tau]}) \bullet X = \xi(X^\tau - X^\nu).$$

BEWEIS: Es sei $K \in \mathbb{R}$ mit $|\xi| \leq K$ \mathbb{P} -fast sicher. Mit $\nu_n := n^{-1}\lceil n\nu \rceil$, $\tau_n := n^{-1}\lceil n\tau \rceil$ erhält man zwei Folgen von Stoppzeiten, für die $\nu_n \rightarrow \nu$, $\tau_n \rightarrow \tau$ mit $n \rightarrow \infty$ sowie $\nu_n \leq \tau_n$ für alle $n \in \mathbb{N}$ gilt. Setzt man

$$H^{(n)} := \xi 1_{(\nu_n, \tau_n]}, \quad n \in \mathbb{N},$$

so erhält man, da wegen $\nu \leq \nu_n$ die Zufallsvariable ξ \mathcal{F}_{ν_n} -messbar ist, eine Folge vorhersehbarer Prozesse mit

$$\begin{aligned} \|H^{(n)} - H\|_X^2 &= E \int_0^\infty (H^{(n)} - H)_t^2 d\langle X \rangle_t \\ &= E \int_0^\infty \xi^2 |1_{(\nu_n, \tau_n]}(t) - 1_{(\nu, \tau]}(t)|^2 d\langle X \rangle_t \\ &\leq K^2 E \int_0^\infty (1_{(\nu, \nu_n]}(t) + 1_{(\tau, \tau_n]}(t)) d\langle X \rangle_t \\ &= K^2 \left(E(\langle X \rangle_{\nu_n} - \langle X \rangle_\nu) + E(\langle X \rangle_{\tau_n} - \langle X \rangle_\tau) \right) \\ &\rightarrow 0 \quad \text{mit } n \rightarrow \infty, \end{aligned}$$

denn wegen $E\langle X \rangle_\infty < \infty$ (Lemma 4.19) und der Stetigkeit der Pfade von $\langle X \rangle$ lässt sich der Satz von der majorisierten Konvergenz anwenden. Für $t \in (k/n, (k+1)/n]$ hat man

$$H^{(n)}(t, \omega) = \begin{cases} \xi(\omega), & \text{wenn } \nu_n(\omega) \leq k/n \text{ und } \tau_n(\omega) \geq (k+1)/n, \\ 0, & \text{sonst,} \end{cases}$$

d.h. es gilt

$$H^{(n)}(t, \omega) = \sum_{k=0}^{\infty} \left(\xi(\omega) 1_{[0, k/n]}(\nu_n(\omega)) 1_{[(k+1)/n, \infty)}(\tau_n(\omega)) \right) 1_{(k/n, (k+1)/n]}(t).$$

Wegen $1_{[0, k/n]}(\nu_n) = 1_{\{\nu \leq k/n\}}$ ist $\xi 1_{[0, k/n]} \circ \nu_n$ nach Aufgabe 48 (a) $\mathcal{F}_{k/n}$ -messbar. Außerdem gilt

$$1_{[(k+1)/n, \infty)} \circ \tau_n = 1_{\{\tau > k/n\}} = 1 - 1_{\{\tau \leq k/n\}},$$

also ist auch dieser Faktor $\mathcal{F}_{k/n}$ -messbar. Insgesamt zeigen diese Überlegungen, dass $H^{(n)} \in b\mathcal{E}$ gilt. Da Prozesse aus \mathcal{M}_0^2 durch ihre ‘finalen Werte’ charakterisiert werden, reicht es nach Satz 4.16 zu zeigen, dass mit $n \rightarrow \infty$

$$(H^{(n)} \bullet X)_\infty \rightarrow \xi(X_\tau - X_\nu) \quad \mathbb{P}\text{-fast sicher}$$

gilt (dies identifiziert den L^2 -Grenzwert). Satz 4.14 liefert

$$\begin{aligned} (H^{(n)} \bullet X)_\infty &= \sum_{k=0}^{\infty} \xi (1_{[0, k/n]} \circ \nu_n) (1_{[(k+1)/n, \infty)} \circ \tau_n) (X_{(k+1)/n} - X_{k/n}) \\ &= \xi (X_{n^{-1}\lceil n\tau \rceil} - X_{n^{-1}\lceil n\nu \rceil}) \end{aligned}$$

(wenn man ξ herauszieht, so bleibt eine teleskopierende Summe übrig), woraus die gewünschte Konvergenz mit der Stetigkeit der Pfade von X folgt. \square

LEMMA 4.25 *Es seien $X \in \mathcal{M}_0^2$, τ eine Stoppzeit und H ein vorhersehbarer Prozess mit $E \int_0^\infty H_t^2 d\langle X \rangle_t < \infty$. Dann gilt*

$$(H \bullet X)^\tau = (H1_{(0,\tau]}) \bullet X = H \bullet X^\tau.$$

BEWEIS: Man muss sich zunächst einmal klarmachen, dass die im zweiten und dritten Ausdruck auftauchenden stochastischen Integrale durch den bisher, also mit Satz 4.16, erreichten Rahmen abgedeckt sind; dies ist aber leicht mit Aufgabe 48 (b) ($H1_{(0,\tau]}$ ist vorhersehbar) und Aufgabe 49 (c) ($\langle X^\tau \rangle = \langle X \rangle^\tau$) einzusehen. Insbesondere sind also alle drei Prozesse Elemente von \mathcal{M}_0^2 und somit über ihre ‘finalen Werte’ identifizierbar.

Wir betrachten zunächst Prozesse H von der Form

$$H(t, \omega) = \xi(\omega)1_{(u,v]}(t), \quad 0 \leq u < v < \infty, \quad \xi \mathcal{F}_u\text{-messbar und beschränkt.}$$

Dann gilt $((H \bullet X)^\tau)_\infty = \xi(X_{\tau \wedge v} - X_{\tau \wedge u})$ und man hat

$$H1_{(0,\tau]} = \xi 1_{(u,v]}1_{(0,\tau]} = \xi 1_{(u,(v \wedge \tau) \vee u]}.$$

Mit Lemma 4.24 folgt also

$$((H1_{(0,\tau]}) \bullet X)_\infty = \xi(X^{(v \wedge \tau) \vee u} - X^u)_\infty = \xi(X_{\tau \wedge v} - X_{\tau \wedge u})$$

(man überprüft einfach die Fälle $\tau \leq u$, $u < \tau \leq v$ und $\tau > v$). Schließlich gilt, da X^τ ein Element von \mathcal{M}_0^2 ist und somit Satz 4.21 angewendet werden kann,

$$H \bullet X^\tau = \xi((X^\tau)^v - (X^\tau)^u) = \xi(X^{\tau \wedge v} - X^{\tau \wedge u}).$$

Damit ist die Behauptung für diese spezielle Klasse von Integranden bewiesen. Linearität und Isometrie erlauben, wie man sich leicht klarmachen kann (und sollte), den Übergang auf allgemeinere H -Prozesse. \square

Wir erinnern daran, dass stetige lokale Martingale mit Start in 0 so lokalisiert werden können, dass man beschränkte Martingale erhält; insbesondere können wir also annehmen, dass die lokalisierten Prozesse Elemente von \mathcal{M}_0^2 sind. Nachdem wir die Verallgemeinerung bei den Integranden durchgeführt haben, dehnen wir nun die Klasse der Integratoren aus.

SATZ UND DEFINITION 4.26 *Es seien X ein stetiges lokales Martingal mit $X_0 \equiv 0$ und H ein vorhersehbarer Prozess mit*

$$\mathbb{P}\left(\int_0^t H_s^2 d\langle X \rangle_s < \infty\right) = 1 \quad \text{für alle } t \geq 0.$$

Dann existiert eine Folge $(\tau_n)_{n \in \mathbb{N}}$ von Stoppzeiten mit $\tau_n \uparrow \infty$ und

$$X^{\tau_n} \in \mathcal{M}_0^2, \quad E \int_0^\infty H_t^2 d\langle X^{\tau_n} \rangle_t < \infty \quad \text{für alle } n \in \mathbb{N}.$$

Wir definieren das stochastische Integral $H \bullet X$ von H bezüglich X durch

$$(H \bullet X)_t = (H \bullet X^{\tau_n})_t \quad \text{auf } \{\tau_n > t\}.$$

Der Prozess $H \bullet X$ ist unabhängig von der Wahl der Folge $(\tau_n)_{n \in \mathbb{N}}$ und ist ein stetiges lokales Martingal mit $(H \bullet X)_0 \equiv 0$.

BEWEIS: Zu X existiert eine lokalisierende Folge $(\tau_n^{(1)})_{n \in \mathbb{N}}$ derart, dass $X^{\tau_n^{(1)}} \in \mathcal{M}_0^2$ gilt für alle $n \in \mathbb{N}$. Wir definieren $(\tau_n^{(2)})_{n \in \mathbb{N}}$ durch

$$\tau_n^{(2)} := \inf \left\{ t \geq 0 : \int_0^t H_s^2 d\langle X \rangle_s \geq n \right\},$$

wobei wie immer $\inf \emptyset = \infty$. Dann ist auch $(\tau_n)_{n \in \mathbb{N}}$ mit $\tau_n := \tau_n^{(1)} \wedge \tau_n^{(2)}$ eine lokalisierende Folge zu X und für alle $n \in \mathbb{N}$ genügt wegen $\langle X^{\tau_n} \rangle = \langle X \rangle^{\tau_n}$ die Kombination von Integrand H und Integrator X^{τ_n} den Bedingungen in Satz 4.14. Insbesondere sind die stochastischen Integrale $H \bullet X^{\tau_n}$ definiert und Elemente von \mathcal{M}_0^2 . Mit Lemma 4.25 erhält man

$$(H \bullet X^{\tau_{n+1}})^{\tau_n} = H \bullet X^{\tau_{n+1} \wedge \tau_n} = H \bullet X^{\tau_n},$$

die Prozesse sind also zueinander kompatibel. Ganz analog erhält man auch, dass $H \bullet X$ unabhängig von der Wahl der Folge $(\tau_n)_{n \in \mathbb{N}}$ ist, wobei wie immer solche Aussagen modulo Ununterscheidbarkeit zu interpretieren sind.

Genügt das Paar (t, ω) der Bedingung $\tau_n(\omega) > t$, so gilt $(H \bullet X)(t, \omega) = (H \bullet X^{\tau_n})(t, \omega)$. Es gilt also $(H \bullet X)^{\tau_n} = H \bullet X^{\tau_n}$, d.h. wegen $H \bullet X^{\tau_n} \in \mathcal{M}_0^2$, dass $(\tau_n)_{n \in \mathbb{N}_0}$ eine lokalisierende Folge und somit $H \bullet X$ ein lokales Martingal ist. Die Stetigkeit der Pfade und $(H \bullet X)_0 \equiv 0$ schließlich ergeben sich aus der Konstruktion. \square

BEISPIEL 4.27 Es sei $B = (B_t)_{t \geq 0}$ die Brownsche Bewegung, wir wollen $B \bullet B$ berechnen. Zunächst sind die Voraussetzungen von Satz 4.26 zu überprüfen. Bekannt ist, dass B ein Martingal ist mit Klammerprozess $\langle B \rangle_t = t$, allerdings ist B wegen $EB_t^2 = t$ nicht L^2 -beschränkt. Setzt man $\tau_n \equiv n$, so gilt

$$\sup_{t \geq 0} E(B_t^{\tau_n})^2 = \sup_{t \geq 0} EB_{t \wedge n}^2 = n < \infty,$$

die Folge $(\tau_n)_{n \in \mathbb{N}}$ lokalisiert also B ‘in den Raum \mathcal{M}_0^2 hinein’. Die Brownsche Bewegung hat stetige und somit auf kompakten Intervallen beschränkte Pfade, woraus

$$\int_0^t B_s^2 d\langle B \rangle_s < \infty \quad \text{für alle } t \geq 0$$

folgt (eine etwaige Ausnahmenullmenge ist wieder ohne Relevanz, da $B \bullet B$ sowieso nur modulo Ununterscheidbarkeit bestimmbar ist). Die Folge $(\tau_n)_{n \in \mathbb{N}}$ führt darüberhinaus auf

$$\begin{aligned} E \int_0^\infty B_s^2 d\langle B^{\tau_n} \rangle_s &= E \int_0^n B_s^2 ds \\ &= \int_0^n E B_s^2 ds \quad (\text{Fubini, alles } \geq 0) \\ &= \int_0^n s ds = n^2/2 < \infty. \end{aligned}$$

Wir können also Satz 4.26 anwenden und erhalten $(B \bullet B)_t = (B \bullet B^{\tau_n})_t$ auf $t < n$. Seien also $n \in \mathbb{N}$, $t \geq 0$ mit $t < n$. Zur Berechnung von $(B \bullet B^{\tau_n})_t$ verwenden wir Satz 4.23 und definieren hierzu Partitionen über Vielfache von $1/m$ und elementare Prozesse $H^{(m)}$ durch

$$H^{(m)} := \sum_{i=1}^{\infty} B_{(i-1)/m} 1_{((i-1)/m, i/m]}.$$

Wir haben $H = B$ und erhalten mit Fubini und $E(B_t - B_s)^2 = |t - s|$

$$\begin{aligned} \|H^{(m)} - H\|_{B^{\tau_n}} &= \sum_{i=1}^{mn} E \left(\int_{(i-1)/m}^{i/m} (B_{(i-1)/m} - B_t)^2 dt \right) \\ &= \sum_{i=1}^{mn} \int_{(i-1)/m}^{i/m} \left(t - \frac{i-1}{m} \right) dt \\ &\leq \frac{n}{m} \rightarrow 0 \quad \text{mit } m \rightarrow \infty, \end{aligned}$$

d.h. $(B \bullet B^{\tau_n})_\infty$ ist der L^2 -Limes von

$$Y_m := \sum_{i=1}^{mn} B_{(i-1)/m} (B_{i/m} - B_{(i-1)/m})$$

mit $m \rightarrow \infty$. Es sei

$$Z_m := \sum_{i=1}^{mn} (B_{i/m} - B_{(i-1)/m})^2.$$

Aus dem Beweis zu Satz 4.8 geht hervor, dass Z_m mit $m \rightarrow \infty$ in L^2 gegen n konvergiert. Es gilt

$$\begin{aligned} Z_m &= \sum_{i=1}^{mn} B_{i/m}^2 + \sum_{i=1}^{mn} B_{(i-1)/m}^2 - 2 \sum_{i=1}^{mn} B_{i/m} B_{(i-1)/m} \\ &= \sum_{i=1}^{mn} B_{i/m}^2 - \sum_{i=1}^{mn} B_{(i-1)/m}^2 - 2 \sum_{i=1}^{mn} B_{(i-1)/m} (B_{i/m} - B_{(i-1)/m}) \\ &= B_n^2 - B_0^2 - 2Y_m, \end{aligned}$$

also erhält man

$$Y_m \rightarrow \frac{1}{2}(B_n^2 - n) \quad \text{in } L^2 \text{ mit } m \rightarrow \infty.$$

Da $(B_t^2 - t)_{t \geq 0}$ ein Martingal ist (man kann dies leicht direkt nachrechnen, oder man erinnert sich daran, wie der Klammerprozess zur Brownschen Bewegung aussieht), folgt nun mit Lemma 4.20

$$(B \bullet B^{\tau_n})_t = E[(B \bullet B^{\tau_n})_\infty | \mathcal{F}_t] = E\left[\frac{1}{2}(B_n^2 - n) \middle| \mathcal{F}_t\right] = \frac{1}{2}(B_t^2 - t)$$

für $t < n$, also insgesamt

$$\int_0^t B_s dB_s = \frac{1}{2}(B_t^2 - t) \quad \text{für alle } t \geq 0.$$

Interessant ist an dieser Stelle der Vergleich mit stetigen FV-Prozessen $(A_t)_{t \geq 0}$ mit $A_t \equiv 0$, bei denen man

$$\int_0^t A_s dA_s = \frac{1}{2} A_t^2 \quad \text{für alle } t \geq 0$$

erhält (Aufgabe 50). Man kann das obige Resultat auch in der Form

$$B \bullet B = \frac{1}{2}(B^2 - \langle B \rangle)$$

schreiben und erkennt dann das Auftreten der quadratischen Variation als ‘Korrekturgröße’, die bei FV-Prozessen verschwindet.

Damit haben wir das stochastische Integral für Martingalintegratoren konstruiert und ein wichtiges Beispiel im Detail durchgerechnet. Wir bemerken noch, ohne den etwas langweiligen Beweis auszuführen, dass das so definierte Integral bilinear ist im Sinne von

$$\begin{aligned} H \bullet (X + Y) &= H \bullet X + H \bullet Y, & (G + H) \bullet X &= G \bullet X + H \bullet X, \\ (\alpha H) \bullet X &= H \bullet (\alpha X) = \alpha (H \bullet X) & (\alpha \in \mathbb{R}), \end{aligned}$$

alles modulo Ununterscheidbarkeit und unter der Voraussetzung, dass die beteiligten stochastischen Integrale existieren. Darüberhinaus übertragen sich die Lokalisierungseigenschaften aus Lemma 4.25 auf die in Satz 4.26 betrachtete Situation, d.h. es gilt

$$(H \bullet X)^\tau = (H1_{(0,\tau]}) \bullet X = H \bullet X^\tau$$

für alle Stoppzeiten τ , mit denselben Nebenbemerkungen wie bei der Bilinearität.

4.5 Semimartingale und Itô-Formel. Wir kombinieren zunächst die Konstruktionen aus den Abschnitten 4.3 und 4.4. Es sei weiterhin $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum mit einer Filtration $(\mathcal{F}_t)_{t \geq 0}$, wobei \mathcal{F}_0 alle \mathbb{P} -Nullmengen von \mathcal{A} enthalte.

DEFINITION 4.28 Ein zu $(\mathcal{F}_t)_{t \geq 0}$ adaptierter Prozess $X = (X_t)_{t \geq 0}$ heisst *stetiges Semimartingal*, wenn es Prozesse $A = (A_t)_{t \geq 0}$, $M = (M_t)_{t \geq 0}$ gibt mit stetigen Pfaden und $A_0 \equiv M_0 \equiv 0$ derart, dass A ein FV-Prozess und M ein lokales Martingal ist und dass gilt:

$$X_t = X_0 + A_t + M_t \quad \text{für alle } t \geq 0.$$

Bis auf Ununterscheidbarkeit ist die Zerlegung in dieser Definition eindeutig: Hat man

$$X_t = X_0 + A_t^{(1)} + M_t^{(1)} = X_0 + A_t^{(2)} + M_t^{(2)} \quad \text{für alle } t \geq 0,$$

so folgt $M^{(1)} - M^{(2)} = A^{(2)} - A^{(1)}$, das lokale Martingal $M^{(1)} - M^{(2)}$ hätte also stetige Pfade von lokal beschränkter Totalvariation und wäre somit nach Satz 4.14 wegen $M_0^{(1)} = M_0^{(2)}$ vom Nullprozess ununterscheidbar. Insbesondere können wir den Klammerprozess $\langle X \rangle$ zu einem Semimartingal X mit der Darstellung $X = X_0 + A + M$ sinnvoll definieren durch $\langle X \rangle := \langle M \rangle$; die Interpretation als quadratische Variation bleibt hierbei erhalten. Semimartingale bilden

offensichtlich einen linearen Raum, wir können also auch die quadratische Kovariation zweier Semimartingale X und Y wieder über den Polarisationsstrick definieren als

$$\langle X, Y \rangle := \frac{1}{2}(\langle X + Y \rangle - \langle X \rangle - \langle Y \rangle).$$

Die quadratische Kovariation zu X und Y ist der (eindeutige) FV-Prozess A , mit dem $XY - A$ ein stetiges lokales Martingal ist; bei FV-Prozessen A erhält man $\langle A, Y \rangle \equiv 0$.

Für Prozesse H , für die sowohl das Integral bezüglich des FV-Anteils A als auch das stochastische Integral bezüglich des Martingal-Anteils M definiert sind, setzen wir in naheliegender Weise

$$H \bullet X := ((H \bullet X)_{t \geq 0}) \quad \text{mit} \quad (H \bullet X)_t := \int_0^t H_s dA_s + \int_0^t H_s dM_s.$$

Dies ist sicher dann möglich, wenn H ein vorhersehbarer Prozess ist, dessen Pfade auf kompakten Intervallen beschränkt sind (alle adaptierten Prozesse mit stetigen Pfaden werden hierdurch erfasst); für solche Integranden ist $H \bullet X$ wieder ein Semimartingal (!). Die am Ende von Abschnitt 4.4 aufgelisteten Bilinearitätseigenschaften übertragen sich leicht auf Semimartingalintegratoren, und es gilt weiterhin

$$(H \bullet X)^\tau = (H 1_{(0, \tau]}) \bullet X = H \bullet X^\tau$$

für alle Stoppzeiten τ .

Da $H \bullet X$ wieder ein Semimartingal ist, ergeben sich unmittelbar zwei Fragen: Wie sieht der zugehörige Klammerprozess aus, und was passiert bei nochmaliger Integration? Mit der oben eingeführten Notation lassen sich die Antworten sehr kompakt und elegant formulieren.

SATZ 4.29 *Sind X und Y stetige Semimartingale, so gilt*

$$\langle H \bullet X, Y \rangle = H \bullet \langle X, Y \rangle$$

für alle Prozesse H , für die die beteiligten Integrale existieren.

BEWEIS: Es seien zunächst $X, Y \in \mathcal{M}_0^2$ und H von der Form

$$H_t(\omega) = \xi(\omega) 1_{(u, v]}(t)$$

mit einer beschränkten, \mathcal{F}_u -messbaren Zufallsvariablen ξ . Dann gilt

$$H \bullet X = \xi(X^v - X^u), \quad H \bullet \langle X, Y \rangle = \xi(\langle X, Y \rangle^v - \langle X, Y \rangle^u),$$

also $(H \bullet X)Y - H \bullet \langle X, Y \rangle = \xi M$ mit

$$M := (X^v Y - \langle X, Y \rangle^v) - (X^u Y - \langle X, Y \rangle^u).$$

Wegen $\langle X, Y \rangle^v = \langle X^v, Y \rangle$ und $\langle X, Y \rangle^u = \langle X^u, Y \rangle$ ist M die Differenz zweier Martingale, also wieder ein Martingal. Da M für Zeitpunkte $t \leq u$ verschwindet und ξ \mathcal{F}_u -messbar ist, ist auch ξM ein Martingal. Subtrahiert man also den FV-Prozess (!) $H \bullet \langle X, Y \rangle$ vom Produkt der Prozesse $H \bullet X$ und Y , so erhält man ein Martingal. Dies bedeutet, dass $H \bullet \langle X, Y \rangle$ die quadratische Kovariation von $H \bullet X$ und Y ist.

Damit ist die Behauptung unter den obigen Zusatzbedingungen an H , X und Y bewiesen. Diese Zusatzbedingungen kann man nun in der üblichen Weise unter Verwendung der Linearität, der Isometrie aus Satz 4.23, einer Lokalisierung und einer Übertragung von lokalen Martingalen auf Semimartingale beseitigen. \square

Aus Satz 4.29 folgt in Verbindung mit Aufgabe 49 (b), dass das stochastische Integral $H \bullet X$ für Integratoren $X \in \mathcal{M}_0^2$ durch

$$\langle H \bullet X, Y \rangle = H \bullet \langle X, Y \rangle \quad \text{für alle } Y \in \mathcal{M}_0^2$$

charakterisiert werden kann. Wir haben uns das stochastische Integral konstruktiv erarbeitet, man kann aber auch alternativ von dieser Eigenschaft ausgehen und erhält auf diese Weise einen etwas abstrakteren Zugang. Lokalisierung verallgemeinert diese Aussage auf stetige lokale Martingale X mit $X \equiv 0$.

SATZ 4.30 (Assoziativität des Integrals) *Ist X ein stetiges Semimartingal, so gilt*

$$G \bullet (H \bullet X) = (GH) \bullet X$$

für alle Prozesse G, H , für die die beteiligten Integrale existieren.

BEWEIS: Es reicht offensichtlich, diese Aussage separat für FV-Prozesse A mit $A_0 \equiv 0$ und stetige lokale Martingale X mit $X_0 \equiv 0$ zu beweisen. Der erste Teil ist Aufgabe 52, bei dem zweiten verwenden wir die oben angegebene Charakterisierung über die quadratische Kovariation: Für alle stetigen lokalen Martingale Y mit $Y_0 \equiv 0$ erhält man mit Satz 4.29 und unter Verwendung der Assoziativität im FV-Fall

$$\begin{aligned} \langle G \bullet (H \bullet X), Y \rangle &= G \bullet \langle H \bullet X, Y \rangle = G \bullet (H \bullet \langle X, Y \rangle) \\ &= (GH) \bullet \langle X, Y \rangle = \langle (GH) \bullet X, Y \rangle. \end{aligned}$$

\square

Zusammengefasst liefern die letzten beiden Sätze auch

$$\langle H \bullet X \rangle = \langle H \bullet X, H \bullet X \rangle = H \bullet \langle X, H \bullet X \rangle = H \bullet (H \bullet \langle X, X \rangle) = H^2 \bullet \langle X \rangle.$$

Betrachtet man die zu den Klammerprozessen gehörenden Maße, so sieht man, dass das Maß zum Klammerprozess von $H \bullet X$ Dichte H^2 bezüglich des Maßes zum Klammerprozess von X hat.

Die bereits im Beispiel 4.27 aufgetretene Korrekturgröße macht sich auch bei der partiellen Integration bemerkbar (man vergleiche das folgende Resultat mit der entsprechenden Formel aus Aufgabe 50 für den FV-Fall; im Falle $X = Y$ erhalten wir eine Formel, die die Rechnung aus Beispiel 4.27 verallgemeinert).

SATZ 4.31 *Sind X und Y stetige Semimartingale, so gilt*

$$XY = X_0Y_0 + X \bullet Y + Y \bullet X + \langle X, Y \rangle.$$

BEWEIS: Ist A ein FV-Prozess und M ein lokales Martingal, so hat man bekanntlich $\langle A, M \rangle \equiv 0$. Für FV-Prozesse ist die Aussage Gegenstand der bereits erwähnten Aufgabe 50; es reicht also, den Fall zu betrachten, in dem X und Y stetige lokale Martingale mit Start in 0 sind. Wir können darüberhinaus sogar voraussetzen, dass X und Y beschränkte Martingale sind (sonst wird geeignet lokalisiert) und brauchen wegen des Polarisationsstricks nur den Fall $X = Y$, also

$$X_t^2 = 2 \int_0^t X_s dX_s + \langle X \rangle_t \quad \text{für alle } t \geq 0$$

zu behandeln. Zum Beweis dieser Aussage sei $t \geq 0$ fest. Für alle $n \in \mathbb{N}$ gilt

$$\sum_{i=1}^n (X_{it/m} - X_{(i-1)t/m})^2 = X_t^2 - X_0^2 - 2 \sum_{i=1}^n X_{(i-1)t/n} (X_{it/m} - X_{(i-1)t/m})$$

(diese elementare Umformung wurde bereits in Beispiel 4.27 benutzt). Nach Satz 4.17 konvergiert die linke Seite mit $n \rightarrow \infty$ in Wahrscheinlichkeit gegen $\langle X \rangle_t$. Die Summe auf der rechten Seite konvergiert nach Satz 4.23 gegen $\int_0^t X_s dX_s$, wenn

$$E \int_0^t (X_{\lfloor \frac{ms}{t} \rfloor \frac{t}{m}} - X_s)^2 d\langle X \rangle_s \rightarrow 0 \quad \text{mit } n \rightarrow \infty$$

gilt. Stetigkeit der Pfade von X liefert die punktweise Konvergenz der Integranden gegen 0. Da X beschränkt ist und somit auch $E\langle X \rangle_t < \infty$ gilt (Lemma 4.19), lässt sich der Satz von der majorisierten Konvergenz anwenden und liefert die gewünschte Konvergenz. \square

Die obige Formel für die partielle Integration opfert strikte Legalität zugunsten eleganter Kompaktheit. Ausgeschrieben ist natürlich

$$X_t Y_t = X_0 Y_0 + \int_0^t X_s dY_s + \int_0^t Y_s dX_s + \langle X, Y \rangle_t \quad \text{für alle } t \geq 0$$

gemeint, und auch der obligatorische Zusatz ‘modulo Ununterscheidbarkeit’ darf eigentlich nicht fehlen*. Wir kommen nun zu dem zentralen Resultat der Theorie, bei dem wir, was die Notation betrifft, ähnlich verfahren, und beispielsweise $f(X)$ für den stochastischen Prozess $(f(X_t))_{t \geq 0}$ schreiben.

SATZ 4.32 (Die Itô-Formel) *Ist X ein stetiges Semimartingal und $f : \mathbb{R} \rightarrow \mathbb{R}$ eine zweimal stetig differenzierbare Funktion, so gilt*

$$f(X) = f(X_0) + f'(X) \bullet X + \frac{1}{2} f''(X) \bullet \langle X \rangle.$$

Inbesondere ist $f(X)$ wieder ein stetiges Semimartingal.

BEWEIS: Wir betrachten die Menge aller f , für die die behauptete Formel bei gegebenem X erfüllt ist. Offensichtlich ist diese Menge ein linearer Raum, der die Funktion $f \equiv 1$ enthält. Gilt die Formel für ein f , so ist $f(X)$ ein zugelassener Integrator, und die Assoziativität aus Satz 4.30 liefert

$$\begin{aligned} X \bullet f(X) &= X \bullet \left(f'(X) \bullet X + \frac{1}{2} f''(X) \bullet \langle X \rangle \right) \\ &= (X f'(X)) \bullet X + \left(\frac{1}{2} X f''(X) \right) \bullet \langle X \rangle. \end{aligned}$$

Andererseits erhält man mit einer partiellen Integration (Satz 4.31)

$$X f(X) = X_0 f(X_0) + X \bullet f(X) + f(X) \bullet X + \langle X, f(X) \rangle.$$

In Kombination ergibt sich für $g : \mathbb{R} \rightarrow \mathbb{R}$, $x \mapsto x f(x)$,

$$\begin{aligned} g(X) &= g(X_0) + (X f'(X)) \bullet X + f(X) \bullet X \\ &\quad + \left(\frac{1}{2} X f''(X) \right) \bullet \langle X \rangle + f'(X) \bullet \langle X \rangle \\ &= g(X_0) + g'(X) \bullet X + \frac{1}{2} g''(X) \bullet \langle X \rangle, \end{aligned}$$

* Man könnte auch X^0 anstelle von X_0 schreiben — wobei dann natürlich der zur Zeit 0 gestoppte Prozess und nicht irgendeine nullte Potenz gemeint wäre.

wobei wir

$$\langle X, f(X) \rangle = \langle X, f'(X) \bullet X \rangle = f'(X) \bullet \langle X \rangle$$

verwendet haben. Mit f gilt also die Itô-Formel auch für g , der interessierende Raum enthält somit alle Polynome.

Wir setzen nun zusätzlich voraus, dass ein $K < \infty$ existiert mit

$$|X_t| \leq K, \quad \langle X \rangle_t (= \langle M \rangle_t) \leq K \quad \text{für alle } t \geq 0, \quad (*)$$

wobei M den Martingalteil aus der Zerlegung von X bezeichnet. Zu einer beliebigen zweimal stetig differenzierbaren Funktion f existiert nach dem Satz von Weierstraß eine Folge $(f_n)_{n \in \mathbb{N}}$ von Polynomen mit der Eigenschaft

$$\lim_{n \rightarrow \infty} \sup_{|x| \leq K} (|f_n(x) - f(x)| + |f'_n(x) - f'(x)| + |f''_n(x) - f''(x)|) = 0.$$

Sei $t \geq 0$. Wegen $(*)$ können wir den Satz von der majorisierten Konvergenz auf die im Zusammenhang mit dem Integral bis t relevanten FV-Integrale anwenden und erhalten

$$\int_0^t f'_n(X_s) dA_s \rightarrow \int_0^t f'(X_s) dA_s, \quad \int_0^t f''_n(X_s) d\langle X \rangle_s \rightarrow \int_0^t f''(X_s) d\langle X \rangle_s,$$

jeweils punktweise in ω und mit $n \rightarrow \infty$, wobei A den FV-Teil von X bezeichnet. Die Isometrie aus Satz 4.23 liefert außerdem

$$\begin{aligned} E \left(\int_0^t f'_n(X_s) dM_s - \int_0^t f'(X_s) dM_s \right)^2 &= E \int_0^t (f'_n(X_s) - f'(X_s))^2 d\langle M \rangle_s \\ &\leq \left(\sup_{|x| \leq K} |f'_n(x) - f'(x)| \right)^2 E \langle M \rangle_t \\ &\rightarrow 0 \quad \text{mit } n \rightarrow \infty, \end{aligned}$$

also insbesondere $\int_0^t f'_n(X_s) dM_s \rightarrow \int_0^t f'(X_s) dM_s$ in Wahrscheinlichkeit. Insgesamt ergibt dies

$$\begin{aligned} \int_0^t f'_n(X_s) dX_s + \frac{1}{2} \int_0^t f''_n(X_s) d\langle X \rangle_s \\ \rightarrow \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s, \end{aligned}$$

wieder in Wahrscheinlichkeit. Auf der linken Seite kann aber die für Polynome bereits bewiesene Itô-Formel angewendet werden, die $f_n(X_t) - f_n(X_0)$

liefert, was wiederum für alle $\omega \in \Omega$ gegen $f(X_t) - f(X_0)$ konvergiert. Da die Grenzwerte bei der Konvergenz in Wahrscheinlichkeit fast sicher eindeutig sind, folgt

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s \quad \mathbb{P}\text{-fast sicher}$$

für alle $t \geq 0$. Rechts und links stehen Prozesse mit stetigen Pfaden, man hat also sogar Ununterscheidbarkeit.

Damit ist die Itô-Formel für solche Semimartingale bewiesen, die der Zusatzbedingung (*) genügen, es bleibt die Lokalisierung. Sei also X ein beliebiges Semimartingal. Für alle $K \in \mathbb{N}$ sei

$$\tau_K := \inf\{t \geq 0 : |X_t| \geq K \text{ oder } \langle X \rangle_t \geq K\}.$$

Dies ergibt eine Folge von Stoppzeiten, die isoton gegen ∞ konvergiert. Die gestoppten Prozesse genügen der Bedingung (*), also lässt sich die Itô-Formel anwenden und man erhält

$$f(X^{\tau_K}) = f(X_0^{\tau_K}) + f'(X^{\tau_K}) \bullet X^{\tau_K} + \frac{1}{2} f''(X^{\tau_K}) \bullet \langle X^{\tau_K} \rangle \quad \text{für alle } K \in \mathbb{N}.$$

Man macht sich leicht klar, dass beispielsweise $f(X^{\tau_K}) = f(X)^{\tau_K}$ gilt. Mit dem bekannten Zusammenspiel von Integration und Einfrieren (Lemma 4.25 und die später erwähnten Verallgemeinerungen) sowie Satz 4.26 folgt hieraus

$$f(X)^{\tau_K} = \left(f(X_0) + f'(X) \bullet X + \frac{1}{2} f''(X) \bullet \langle X \rangle \right)^{\tau_K} \quad \text{für alle } K \in \mathbb{N},$$

und somit wegen $\tau_K \uparrow \infty$ die Behauptung. \square

BEMERKUNG 4.33 (a) Die Punktnotation dient zur kompakten Schreibweise ‘unbestimmter’ Integrale, ausgeschrieben lautet die Itô-Formel

$$f(X_t) = f(X_0) + \int_0^t f'(X_s) dX_s + \frac{1}{2} \int_0^t f''(X_s) d\langle X \rangle_s \quad \text{für alle } t \geq 0,$$

was wiederum in ‘Differentialnotation’ als

$$df(X_t) = f'(X_t) dX_t + \frac{1}{2} f''(X_t) d\langle X \rangle_t$$

geschrieben werden kann.

(b) Zur Itô-Formel gibt es auch eine mehrdimensionale Version, die wir hier ohne Beweis angeben: Ist $(X_t)_{t \geq 0}$ ein \mathbb{R}^d -wertiger Prozess, dessen Komponenten $(X_t^{(1)})_{t \geq 0}, \dots, (X_t^{(d)})_{t \geq 0}$ den Voraussetzungen von Satz 4.31 genügen, und ist $f : \mathbb{R}^d \rightarrow \mathbb{R}$ zweimal stetig differenzierbar, so gilt mit $D_i f := \partial f / \partial x_i$ etc.:

$$f(X_t) = f(X_0) + \sum_{i=1}^d \int_0^t (D_i f)(X_s) dX_s^{(i)} + \frac{1}{2} \sum_{i,j=1}^d \int_0^t (D_i D_j f)(X_s) d\langle X^{(i)}, X^{(j)} \rangle_s.$$

(c) Die mehrdimensionale Itô-Formel aus (b) kann benutzt werden, um in die eindimensionale Formel eine zusätzliche Zeitabhängigkeit einzubauen: Ist $(X_t)_{t \geq 0}$ ein eindimensionales Semimartingal und definiert man einen \mathbb{R}^2 -wertigen Prozess $(Y_t)_{t \geq 0}$ durch $Y_t := (t, X_t)$, so erhält man wegen $\langle t, X_t \rangle = 0$

$$f(t, X_t) = f(0, X_0) + \int_0^t f'_t(s, X_s) ds + \int_0^t f'_x(s, X_s) dX_s + \frac{1}{2} \int_0^t f''_{xx}(X_s) d\langle X \rangle_s,$$

wobei f'_t, f'_x die Ableitungen nach dem ersten bzw. zweiten Argument der Funktion $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ bezeichnen. Die Zeitabhängigkeit liefert also den zusätzlichen Term $\int_0^t f'_t(s, X_s) ds$. In Differentialnotation wird dies zu

$$df(t, X_t) = f'_t(t, X_t) dt + f'_x(t, X_t) dX_t + \frac{1}{2} f''_{xx}(t, X_t) d\langle X \rangle_t.$$

4.6 Eine Anwendung. Wir beginnen mit einer Charakterisierung der Brownschen Bewegung.

SATZ 4.33 (Lévy) *Es sei $X = (X_t)_{t \geq 0}$ ein stetiges lokales Martingal mit $X_0 \equiv 0$ und*

$$\langle X \rangle_t = t \quad \text{für alle } t \geq 0.$$

Dann ist X eine Brownsche Bewegung.

BEWEIS: Es sei $\theta \in \mathbb{R}$ und

$$f : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{C}, \quad (t, x) \mapsto \exp(i\theta x + \theta^2 t^2 / 2).$$

Wenden wir die Itô-Formel (Satz 4.32 in Verbindung mit Bemerkung 4.33) auf Real- und Imaginärteil an, so folgt

$$\begin{aligned} f(t, X_t) &= f(0, X_0) + \int_0^t f'_t(s, X_s) ds + \int_0^t f'_x(s, X_s) dX_s \\ &\quad + \frac{1}{2} \int_0^t f''_{xx}(s, X_s) d\langle X \rangle_s \\ &= 1 + \frac{1}{2} \theta^2 \int_0^t f(s, X_s) ds + i\theta \int_0^t f(s, X_s) dX_s - \frac{1}{2} \theta^2 \int_0^t f(s, X_s) ds \\ &= 1 + i\theta \int_0^t f(s, X_s) dX_s. \end{aligned}$$

Da nur das stochastische Integral bzgl. eines lokalen Martingals übrigbleibt, folgt hieraus, dass $Y = (Y_t)_{t \geq 0}$ mit $Y_t = f(t, X_t)$ ein lokales Martingal ist. Wegen $|Y_t| \leq \exp(\theta^2 t/2)$ lässt sich Satz 4.12 anwenden (eine durch einen festen, endlichen Wert dominierte Familie von Zufallsvariablen ist natürlich gleichgradig integrierbar), Y ist also sogar ein ‘richtiges’ Martingal. Insbesondere gilt für alle $s, t \geq 0$ mit $s \leq t$

$$E[\exp(i\theta(X_t - X_s)) | \mathcal{F}_s] = \exp(-\theta^2(t-s)/2).$$

Nimmt man hiervon den Erwartungswert, so erhält man die charakteristische Funktion zu $X_t - X_s$ und damit $X_t - X_s \sim N(0, t-s)$. Da außerdem der bedingte Erwartungswert nicht von ω abhängt, hat man

$$E 1_A \exp(i\theta(X_t - X_s)) = \mathbb{P}(A) E \exp(i\theta(X_t - X_s))$$

für alle $\theta \in \mathbb{R}$, $A \in \mathcal{F}_s$. Die Ideen, die dem Eindeutigkeitsatz für charakteristische Funktionen zugrunde liegen, liefern nun

$$\mathbb{P}(A \cap \{X_t - X_s \in B\}) = \mathbb{P}(A) \cdot \mathbb{P}(X_t - X_s \in B)$$

für alle $A \in \mathcal{F}_s, B \in \mathcal{B}$, d.h. $X_t - X_s$ und \mathcal{F}_s sind unabhängig.

Da die Stetigkeit der Pfade von X und $X_0 = 0$ vorausgesetzt wurden, sind damit alle Bedingungen aus Definition 3.2 erfüllt. \square

Bereits aus dem dritten Abschnitt ist bekannt, dass eine Brownsche Bewegung $(B_t)_{t \geq 0}$ ein Martingal ist, und dass auch $(B_t^2 - t)_{t \geq 0}$ ein Martingal ist. Der obige Satz zeigt, dass diese beiden Aspekte bereits die Brownsche Bewegung in der Klasse der stetigen Prozesse charakterisieren.

Mit diesem Resultat zeigen wir nun, dass sich stetige Martingale als zeitlich transformierte Brownsche Bewegungen darstellen lassen, wobei der Klammerprozess als ‘stochastische Uhr’ dient.

SATZ 4.34 *Es sei M ein stetiges lokales Martingal mit den Eigenschaften*

$$M_0 \equiv 0, \quad \langle M \rangle_s < \langle M \rangle_t \text{ für } s < t, \quad \langle M \rangle_t \rightarrow \infty \text{ mit } t \rightarrow \infty.$$

Für alle $t \geq 0$ sei

$$\tau_t := \inf\{s \geq 0 : \langle M \rangle_s \geq t\}, \quad \mathcal{G}_t := \mathcal{F}_{\tau_t}, \quad B_t := M_{\tau_t}.$$

Dann ist $(B_t, \mathcal{G}_t)_{t \geq 0}$ eine Brownsche Bewegung. Weiter ist für alle $t \geq 0$ $\langle M \rangle_t$ eine Stoppzeit bzgl. $(\mathcal{G}_t)_{t \geq 0}$ und es gilt $M_t = B_{\langle M \rangle_t}$.

BEWEIS: Wegen $\tau_s \leq \tau_t$ für $s \leq t$ ist $(\mathcal{G}_t)_{t \geq 0}$ eine Filtration; $(B_t)_{t \geq 0}$ ist nach Lemma 3.14 (b) zu dieser adaptiert. Nach Definition ist $t \mapsto \tau_t$ die Umkehrfunktion zu $t \mapsto \langle M \rangle_t$. Insbesondere sind aufgrund der strengen Monotonie der Pfade von $\langle M \rangle$ die Pfade von $(\tau_t)_{t \geq 0}$ stetig. Als Hintereinanderschaltung stetiger Abbildungen sind also die Pfade des Prozesses B stetig. Aufgrund der Voraussetzungen an M gilt $\{\langle M \rangle_t \leq u\} = \{\tau_{\langle M \rangle_t} \leq \tau_u\} = \{\tau_u \geq t\} \in \mathcal{F}_{\tau_u} = \mathcal{G}_u$, also ist $\langle M \rangle_t$ eine Stoppzeit bzgl. $(\mathcal{G}_t)_{t \geq 0}$. Mit $\langle M \rangle_{\tau_t} = t$ folgt auch unmittelbar $B_{\langle M \rangle_t} = M_t$.

Wegen Satz 4.33 genügt es, noch zu zeigen, dass $(B_t, \mathcal{G}_t)_{t \geq 0}$ und $(B_t^2 - t, \mathcal{G}_t)_{t \geq 0}$ lokale Martingale sind. Hierzu setzen wir $\nu_n := \inf\{t \geq 0 : |M_t| \geq n\}$ und $\rho_n := \langle M \rangle_{\nu_n}$. Aus Aufgabe 41 (a) ist bekannt, dass $(\nu_n)_{n \in \mathbb{N}}$ eine lokalisierende Folge zu M ist. Mit

$$\begin{aligned} \{\rho_n \leq t\} &= \{\langle M \rangle_{\nu_n} \leq t\} = \{\tau_{\langle M \rangle_{\nu_n}} \leq \tau_t\} \\ &= \{\nu_n \leq \tau_t\} \in \mathcal{F}_{\nu_n} \cap \mathcal{F}_{\tau_t} \subset \mathcal{F}_{\tau_t} = \mathcal{G}_t \text{ für alle } t \geq 0 \end{aligned}$$

folgt, dass ρ_n eine $(\mathcal{G}_t)_{t \geq 0}$ -Stoppzeit ist ('vergessene Übungsaufgabe': Sind τ und ρ Stoppzeiten zur Filtration $(\mathcal{F}_t)_{t \geq 0}$, so gilt $\{\tau \leq \rho\} \in \mathcal{F}_\tau \cap \mathcal{F}_\rho$), und es gilt

$$B_t^{(\rho_n)} = B_{\rho_n \wedge t} = M_{\tau_{\rho_n \wedge t}} = M_{\tau_{\rho_n} \wedge \tau_t} = M_{\nu_n \wedge \tau_t} = M_{\tau_t}^{(\nu_n)}.$$

Insgesamt erhalten wir nun mit dem OST für $s \leq t$

$$E[B_t^{(\rho_n)} | \mathcal{G}_s] = E[M_{\tau_t}^{(\nu_n)} | \mathcal{F}_{\tau_s}] = M_{\tau_s}^{(\nu_n)} = B_s^{(\rho_n)},$$

d.h. $(B_t^{(\rho_n)}, \mathcal{G}_t)_{t \geq 0}$ ist ein Martingal. Da offensichtlich $\rho_n \uparrow \infty$ gilt, ist somit $(B_t, \mathcal{G}_t)_{t \geq 0}$ ein lokales Martingal. Bei $(B_t^2 - t, \mathcal{G}_t)_{t \geq 0}$ verfährt man analog. \square

Die Voraussetzung der strengen Monotonie des Klammerprozesses kann beseitigt werden, wenn man mit rechtstetigen Filtrationen arbeitet. Informell ist klar, dass selbst bei stetigem $\langle M \rangle$ der Umkehrprozess τ unstetig sein kann,

wenn $\langle M \rangle$ konstante Pfadstücke hat (wir erinnern uns in diesem Zusammenhang an den bereits in der Vorlesung Stochastik I diskutierten Übergang von Verteilungsfunktionen zu Quantilfunktionen). Andererseits ist aufgrund der Interpretation des Klammerprozesses als quadratischer Variation ebenso klar, dass auf diesen Konstantheitsintervallen M ebenfalls konstant ist, bei der Hintereinanderschaltung von τ und M also wieder ein Prozess mit stetigen Pfaden entsteht.

4.7 Stochastische Differentialgleichungen. Die in Abschnitt 4.5 eingeführte Differentialnotation liefert einen ersten Hinweis auf eines der für Anwendungen wichtigsten Teilgebiete der stochastischen Analysis: stochastische Differentialgleichungen. Besonders solche Gleichungen, die von einer Brownschen Bewegung $B = (B_t)_{t \geq 0}$ ‘angetrieben’ werden, sind hierbei wichtig. Konkret nennt man ein Semimartingal $X = (X_t)_{t \geq 0}$ eine Lösung der stochastischen Differentialgleichung

$$dX_t = a(X_t) dt + b(X_t) dB_t,$$

wenn

$$X_t = X_0 + \int_0^t a(X_s) ds + \int_0^t b(X_s) dB_s$$

für alle $t \geq 0$ gilt. Betrachtet man die Zuwächse von X über kleine Zeitintervalle, so ergibt sich bei glatten a und b näherungsweise

$$X_{t+h} - X_t \approx a(X_t)h + b(X_t)(B_{t+h} - B_t),$$

der Zuwachs über ein Intervall der Länge h ist bei kleinem h und bei $X_t = x$ also näherungsweise normalverteilt mit Mittelwerte $a(x)h$ und Varianz $b(x)^2 h$; X verhält sich lokal wie eine Brownsche Bewegung mit Drift a und Varianz b^2 .

BEISPIEL 4.35 (a) Für $X = (X_t)_{t \geq 0}$ mit $X_t = x_0 + \mu t + \sigma B_t$ für alle $t \geq 0$, $\mu, \sigma \in \mathbb{R}$ fest, gilt $X_t = x_0 + \int_0^t \mu ds + \int_0^t \sigma dB_s$, d.h. X genügt der stochastischen Differentialgleichung

$$dX_t = \mu dt + \sigma dB_t, \quad X_0 = x_0,$$

also mit konstanten Koeffizientenfunktionen $a \equiv \mu$, $b \equiv \sigma$. In dieser Situation ist die obige näherungsweise Interpretation exakt.

(b) Eine besondere Rolle, u.a. in der modernen Finanzmathematik, spielt die *exponentielle Brownsche Bewegung*,

$$X_t := x_0 \exp(\mu t + \sigma B_t) \quad \text{für alle } t \geq 0,$$

wieder mit $\mu, \sigma \in \mathbb{R}$, sowie $x_0 > 0$. Wendet man die Itô-Formel mit $f(x) = x_0 e^x$ auf das Semimartingal $(\mu t + \sigma B_t)_{t \geq 0}$ an, so sieht man, dass X der stochastischen Differentialgleichung

$$dX_t = \left(\mu + \frac{1}{2} \sigma^2 \right) X_t dt + \sigma X_t dB_t, \quad X_0 = x_0,$$

genügt. Dies ist von der obigen Form mit $a(x) = (\mu + \sigma^2/2)x$, $b(x) = \sigma x$. Bei der exponentiellen Brownschen Bewegung sind Drift und Fluktuation (Standardabweichung) des Prozesses also proportional zum aktuellen Wert.

Wie bei den klassischen Differentialgleichungen sind Fälle, in denen sich die Lösung explizit angeben lässt, eher selten. Man wird daher nach Bedingungen für die Existenz und Eindeutigkeit von Lösungen suchen (und den Rest der Numerik überlassen ...). Wir behandeln ein solches Resultat unter relativ starken Voraussetzungen an die Koeffizientenfunktionen a und b und verwenden dabei die aus der Theorie der nicht-stochastischen Differentialgleichungen bekannte Picard-Iteration.

SATZ 4.36 *Es seien $a : \mathbb{R} \rightarrow \mathbb{R}$ und $b : \mathbb{R} \rightarrow \mathbb{R}_+$ global Lipschitz-stetig, d.h. es gelte*

$$|a(x) - a(y)| \leq C_a |x - y|, \quad |b(x) - b(y)| \leq C_b |x - y| \quad \text{für alle } x, y \in \mathbb{R}$$

mit geeigneten Konstanten $C_a, C_b \in \mathbb{R}$. Weiter sei $x_0 \in \mathbb{R}$. Dann existiert zu gegebener Brownscher Bewegung $B = (B_t)_{t \geq 0}$ eine Lösung $X = (X_t)_{t \geq 0}$ der stochastischen Differentialgleichung

$$dX_t = a(X_t) dt + b(X_t) dB_t, \quad X_0 = x_0.$$

Sind darüberhinaus a und b beschränkt, so ist die Lösung bei gegebenem B bis auf Ununterscheidbarkeit eindeutig.

BEWEIS: Wir definieren induktiv eine Folge $(X^{(n)})_{n \in \mathbb{N}_0}$ von stochastischen Prozessen $X^{(n)} = (X_t^{(n)})_{t \geq 0}$ durch $X^{(0)} \equiv x_0$ und, für alle $n \in \mathbb{N}_0$,

$$X_t^{(n+1)} := x_0 + \int_0^t a(X_s^{(n)}) ds + \int_0^t b(X_s^{(n)}) dB_s. \quad (\star)$$

Offensichtlich haben alle diese Prozesse stetige Pfade und den Wert x_0 in $t = 0$. Die elementare Ungleichung $(a + b)^2 \leq 2(a^2 + b^2)$ führt auf

$$\sup_{0 \leq s \leq t} (X_s^{(n+1)} - X_s^{(n)})^2 \leq 2\Psi_n(t) + 2\Phi_n(t),$$

wobei

$$\begin{aligned}\Psi_n(t) &:= \sup_{0 \leq s \leq t} \left(\int_0^s |a(X_u^{(n)}) - a(X_u^{(n-1)})| du \right)^2, \\ \Phi_n(t) &:= \sup_{0 \leq s \leq t} \left(\int_0^s (b(X_u^{(n)}) - b(X_u^{(n-1)})) dB_u \right)^2.\end{aligned}$$

Die Ungleichung von Cauchy-Schwarz liefert

$$\Psi_n(t) = \left(\int_0^t |a(X_s^{(n)}) - a(X_s^{(n-1)})| ds \right)^2 \leq t \cdot \int_0^t (a(X_s^{(n)}) - a(X_s^{(n-1)}))^2 ds.$$

Bei Φ_n verwenden wir die Tatsache, dass die beiden stochastischen Integrale Martingale sind, und erhalten mit der L^p -Ungleichung von Doob (Satz 8.5 der Vorlesung Stochastik II in einer Variante mit stetigen Zeitparameter, bei $p = 2$) und der Itô-Isometrie

$$\begin{aligned}E\Phi_n(t) &\leq 4E \left(\int_0^t (b(X_s^{(n)}) - b(X_s^{(n-1)})) dB_s \right)^2 \\ &= 4E \int_0^t (b(X_s^{(n)}) - b(X_s^{(n-1)}))^2 ds.\end{aligned}$$

Nun können wir die Voraussetzung anwenden, dass die Koeffizientenfunktionen Lipschitz-stetig sind, und erhalten

$$\begin{aligned}\phi_n(t) &:= E \sup_{0 \leq s \leq t} (X_s^{(n+1)} - X_s^{(n)})^2 \\ &\leq 2tC_a^2 \int_0^t E(X_s^{(n)} - X_s^{(n-1)})^2 ds + 8C_b^2 \int_0^t E(X_s^{(n)} - X_s^{(n-1)})^2 ds \\ &\leq (2C_a^2t + 8C_b^2) \int_0^t \phi_{n-1}(s) ds\end{aligned}$$

für alle $t \geq 0$. Bei $n = 0$ ergibt sich

$$\begin{aligned}\phi_0(t) &= E \sup_{0 \leq s \leq t} (X_s^{(1)} - x_0)^2 \\ &= E \sup_{0 \leq s \leq t} \left(\int_0^s a(x_0) du + \int_0^s b(x_0) dB_u \right)^2 \\ &\leq 2ta(x_0)^2 + 2b(x_0)^2 E \left(\sup_{0 \leq s \leq t} B_s^2 \right) \\ &\leq 2ta(x_0)^2 + 8b(x_0)^2t,\end{aligned}$$

wobei wir im letzten Schritt wieder die Doob'sche Ungleichung verwendet haben. Sei nun $t > 0$ vorgegeben. Wir setzen

$$a := \max\{2a(x_0)^2 + 8b(x_0)^2, 2C_a^2 t + 8C_b^2\}.$$

Mit Induktion lässt sich nun leicht zeigen, dass dann für alle $n \in \mathbb{N}_0$

$$\phi_n(s) \leq \frac{(as)^{n+1}}{(n+1)!} \quad \text{auf } 0 \leq s \leq t$$

gilt. Insbesondere existiert zu jedem $t > 0$ eine Konstante $c = c(t)$ mit

$$\phi_n(t) \leq \frac{c^{n+1}}{(n+1)!} \quad \text{für alle } n \in \mathbb{N}_0.$$

Sei nun temporär \mathcal{X} der Raum aller Abbildungen $X : [0, t] \times \Omega \rightarrow \mathbb{R}$, die im ersten Argument stetig und im zweiten messbar sind; wir identifizieren wieder ununterscheidbare Abbildungen (Prozesse). Auf \mathcal{X} wird durch

$$\|X\|^2 := E\left(\sup_{0 \leq s \leq t} X_s^2\right)$$

eine Norm definiert; $(\mathcal{X}, \|\cdot\|)$ ist vollständig (dies beweisen wir nicht; siehe Lemma 4.20 für eine ähnliche Situation). Wegen

$$\|X^{(m)} - X^{(n)}\| \leq \sum_{k=n \wedge m}^{n \vee m} \left(\frac{c^k}{k!}\right)^{1/2} \rightarrow 0 \quad \text{mit } m, n \rightarrow \infty$$

ist $(X^{(n)})_{n \in \mathbb{N}}$ eine Cauchy-Folge in $(\mathcal{X}, \|\cdot\|)$. Sei X der (bis auf Ununterscheidbarkeit) eindeutige Limes. Die $\|\cdot\|$ -Konvergenz in Verbindung mit der Stetigkeit der Koeffizientenfunktionen ist stark genug, um beim Grenzübergang in (\star) bei den Integralen Limes und Integral vertauschen zu können (beim stochastischen Integral verwendet man wieder die Itô-Isometrie), also genügt X auf dem Zeitintervall $[0, t]$ der stochastischen Differentialgleichung aus der Behauptung des Satzes. Die für verschiedene t erhaltenen Grenzprozesse sind zueinander kompatibel (man beachte die Definition von $\|\cdot\|$) und können somit zu dem gesuchten Objekt $(X_t)_{t \geq 0}$ zusammengesetzt werden.

Zur Eindeutigkeit: Bezeichnet

$$\|f\|_\infty := \sup_{t \geq 0} |f(t)|$$

die Supremumsnorm einer (beschränkten) Funktion $f : \mathbb{R}_+ \rightarrow \mathbb{R}$, so erhält man mit der elementaren Ungleichung $(a + b + c)^2 \leq 4(a^2 + b^2 + c^2)$ für jede Lösung X der stochastischen Differentialgleichung die Abschätzung

$$E\left(\sup_{0 \leq s \leq t} X_s^2\right) \leq 4(x_0^2 + t\|\mu\|_\infty^2 + 8t\|\sigma\|_\infty^2)$$

(vgl. die Abschätzung für ϕ_0 im Existenzbeweis). Sind nun X und X' beide Lösungen, so erhält man wie oben für $\phi(t) := E \sup_{0 \leq s \leq t} (X_s - X'_s)^2$ die Ungleichung

$$\phi(t) \leq c \int_0^t \phi(s) ds$$

mit einer geeigneten Konstanten c . Iteration liefert wie beim Existenzbeweis eine gegen 0 konvergente Folge von Oberschranken für ϕ , also gilt $\phi(t) = 0$. \square

4.8 Ausblick: Von Itô über Malliavin zu Skorohod. Zum Abschluss skizzieren wir einen eher funktionalanalytischen Zugang zur stochastischen Integration, der zu einer Verallgemeinerung des in den vorangegangenen Unterabschnitten entwickelten Itô-Integrals führt. Wir folgen im wesentlichen der Darstellung in dem Buch *The Malliavin Calculus and Related Topics* von David Nualart.

Wir beschränken uns auf den Wahrscheinlichkeitsraum

$$(\Omega, \mathcal{A}, \mathbb{P}) = (C_0([0, 1]), \mathcal{B}_{C_0([0, 1])}, W).$$

Hierbei ist $C_0([0, 1])$ der Raum der stetigen Funktionen $f : [0, 1] \rightarrow \mathbb{R}$ mit $f(0) = 0$, $\mathcal{B}_{C_0([0, 1])}$ die hierauf von den Projektionen $\pi_t : C_0([0, 1]) \rightarrow \mathbb{R}$, $f \mapsto f(t)$, erzeugte σ -Algebra und W ist das Wiener-Maß; dies ist die in Abschnitt 3.2 eingeführte kanonische Brownsche Bewegung $B = (B_t)_{0 \leq t \leq 1}$ mit Zeitbereich $T = [0, 1]$. Es geht insbesondere um die Struktur des Hilbert-Raumes $L^2(\Omega, \mathcal{A}, \mathbb{P})$ der quadratisch integrierbaren Zufallsvariablen auf diesem Wahrscheinlichkeitsraum; wir schreiben auch kurz $L^2(\Omega)$. Dabei spielen der Hilbert-Raum

$$L^2(T) = L^2([0, 1], \mathcal{B}_{[0, 1]}, \text{unif}(0, 1))$$

und die (lineare) Abbildung $X : L^2(T) \rightarrow L^2(\Omega)$, $1_{[0, t]} \mapsto \pi_t (= B_t)$ eine Schlüsselrolle. Diese Abbildung ist zunächst nur für bestimmte Indikatorfunktionen definiert, lässt sich aber, da deren Linearkombinationen dicht in $L^2(T)$ liegen und X auf diesen eine Isometrie ist, wie üblich fortsetzen; eine typische Rechnung ist

$$\begin{aligned} \langle X(1_{[0, s]}), X(1_{[0, t]}) \rangle_{L^2(\Omega)} &= EB_s B_t = s \wedge t \\ &= \int_0^1 1_{[0, s]}(u) 1_{[0, t]}(u) du = \langle 1_{[0, t]}, 1_{[0, t]} \rangle_{L^2(T)}. \end{aligned}$$

Man erhält so das *Wiener-Itô-Integral*, für das wir unsere frühere Notation verwenden: $X(h) = \int_0^1 h(t) dB_t = \int h dB$. Man beachte, dass hierbei nur feste Funktionen integriert werden und nicht, wie in den vorangegangenen Abschnitten, stochastische Prozesse (man kann natürlich h als stochastischen Prozess ‘mit nur einem Pfad’ auffassen). Man sieht auch leicht, dass $X(h) \sim N(0, \|h\|^2)$ gilt. Als Bild von X in $L^2(\Omega)$ erhält man also ausschließlich zentriert normalverteilte Zufallsvariablen, und es gilt in Verallgemeinerung der obigen Rechnung $\langle X(h), X(g) \rangle_{L^2(\Omega)} = \langle h, g \rangle_{L^2(T)}$.

Es sei nun \mathcal{H}_0 der von den konstanten Funktionen (Zufallsvariablen) erzeugte lineare Teilraum von $L^2(\Omega)$, \mathcal{H}_1 das bereits erwähnte Bild von X . Klar: für alle $Z_0 \in \mathcal{H}_0$, $Z_1 \in \mathcal{H}_1$ gilt $\langle Z_0, Z_1 \rangle_{L^2(\Omega)} = 0$. Wenn dies der Anfang einer Orthogonalzerlegung von L^2 ist, wie sehen dann $\mathcal{H}_2, \mathcal{H}_3, \dots$ aus? Es sei

$$H_n : \mathbb{R} \rightarrow \mathbb{R}, \quad x \mapsto \frac{(-1)^n}{n!} e^{x^2/2} \frac{d^n}{dx^n} e^{-x^2/2},$$

das n -te *Hermite-Polynom* (siehe auch Aufgabe 53, in der allerdings eine andere Normierung verwendet wird), \mathcal{H}_n der von den Zufallsvariablen $H_n(X(h))$, $h \in L^2(T)$ erzeugte Unterraum von $L^2(\Omega)$. Mit dieser Konstruktion hat man nun in der Tat

$$L^2(\Omega) = \bigoplus_{n=0}^{\infty} \mathcal{H}_n,$$

ein als *Wiener-Chaos-Zerlegung* bekanntes Resultat. Der Teilraum $\mathcal{H}_0 \oplus \dots \oplus \mathcal{H}_n$ besteht aus allen Zufallsvariablen $Z \in L^2(\Omega)$, die sich in der Form $Z = p(X(h_1), \dots, X(h_k))$ mit $h_1, \dots, h_k \in L^2(T)$ und einem (multivariaten) Polynom p vom Grad kleiner oder gleich n schreiben lassen.

Im nächsten Schritt setzen wir \mathcal{H}_n in Beziehung zu

$$L^2(T^m) = L^2([0, 1]^m, \mathcal{B}_{[0,1]}^m, \text{unif}([0, 1]^m));$$

für $m = 1$ vermittelt ja bereits X (das ‘einfache’ Wiener-Itô-Integral) eine Isometrie. Es sei \mathcal{E}^m die Menge der *elementaren Funktionen* $f : T^m \rightarrow \mathbb{R}$, die sich in der Form

$$f(t_1, \dots, t_m) = \sum_{i_1, \dots, i_m=1}^n a_{i_1, \dots, i_m} 1_{A_{i_1} \times \dots \times A_{i_m}}(t_1, \dots, t_m)$$

darstellen lassen. Hierbei setzen wir voraus, dass $A_1, \dots, A_n \in \mathcal{B}_{[0,1]}$ paarweise disjunkt sind, und dass die Koeffizienten a_{i_1, \dots, i_m} nur für Indizes mit paarweise verschiedenen Komponenten von 0 verschieden sind. Für diese definieren wir das *multiple Wiener-Itô-Integral* durch

$$I_m(f) := \sum_{i_1, \dots, i_m=1}^n a_{i_1, \dots, i_m} X(A_{i_1}) \cdot \dots \cdot X(A_{i_m}).$$

Auf \mathcal{E}^m ist I_m linear. Da die Reihenfolge der Zufallsvariablen vertauscht werden kann, gilt $I_m(f) = I_m(\tilde{f})$, wobei \tilde{f} die *Symmetrisierung* zu f bezeichnet:

$$\tilde{f}(t_1, \dots, t_m) := \frac{1}{m!} \sum_{\sigma \in S_m} f(t_{\sigma(1)}, \dots, t_{\sigma(m)}),$$

wie üblich ist dabei S_m die Menge der Permutationen von $\{1, \dots, m\}$. Entscheidend ist nun

$$\langle I_m(f), I_m(g) \rangle_{L^2(\Omega)} = \begin{cases} m! \langle \tilde{f}, \tilde{g} \rangle_{L^2(T^m)}, & \text{bei } m = n, \\ 0, & \text{sonst.} \end{cases}$$

Hier werden nun auch die Forderungen an die Mengen und die Koeffizienten bei den elementaren Funktionen plausibel: $I_m(f)$ ist eine Linearkombination von Produkten von jeweils genau m unabhängigen, zentriert normalverteilten Zufallsvariablen. Bei $m \neq n$ tauchen also beim Ausmultiplizieren nur Produkte auf, bei denen sich ein vom Rest unabhängiger Faktor mit Erwartungswert 0 abspalten lässt, d.h. es gilt $E I_m(f) I_n(g) = 0$ bei $m \neq n$ (man kann annehmen, dass f und g auf der gleichen Familie A_1, \dots, A_n aufbauen). Damit wird es nun möglich, I_m auf ganz $L^2(T^m)$ fortzusetzen; auf dem Teilraum $L_s^2(T^m)$ der symmetrischen Funktionen ist diese Fortsetzung injektiv. Man kann zeigen, dass man für Elemente von $L_s^2(T^m)$ der Form $h^{\otimes m}$ mit $h \in L^2(T)$, $h^{\otimes m}(t_1, \dots, t_m) = h(t_1) \cdots h(t_m)$,

$$I_m(h^{\otimes m}) = m! H_m(X(h))$$

erhält und dies zu einem Beweis dafür ausbauen, dass \mathcal{H}_m , die m -te Komponente in der Chaos-Zerlegung, gerade das Bild von $L_s^2(T^m)$ unter I_m ist. In Analogie zu dem oben beim einfachen Wiener-Itô-Integral erfolgten Rückgriff auf eine früher verwendete Notation liegt es nahe,

$$I_m(f) = \int_T \cdots \int_T f(t_1, \dots, t_m) dB_{t_1} \cdots dB_{t_m}$$

zu schreiben. Dabei ist Vorsicht angesagt: Es geht nicht um ein iteriertes Integral. Bei $m = 2$ beispielsweise ist $\int_T f(t_1, t_2) dB_{t_1}$ für jedes t_2 eine Zufallsgröße, das einfache Wiener-Itô-Integral von $t_2 \mapsto \int_T f(t_1, t_2) dB_{t_1}$ also nicht definiert. Insbesondere würde sich bei einer iterativen Interpretation $I_m(h^{\otimes m}) = X(h)^m$ ergeben — im Widerspruch zum oben angegebenen Resultat.

Die bis hierhin durchgeführten Konstruktionen können wie folgt zusammengefasst werden:

$$\begin{array}{ccccccc} L^2(\Omega) & = & \mathcal{H}_0 & \otimes & \mathcal{H}_1 & \otimes & \cdots \otimes \mathcal{H}_m & \otimes & \cdots \\ & & \uparrow I_0 & & \uparrow I_1 & & \cdots & & \uparrow I_m & \cdots \\ & & L_s^2(T^0) & & L_s^2(T^1) & & \cdots & & L_s^2(T^m) & \cdots \end{array}$$

Alle Abbildungen in diesem Diagramm sind Isometrien: I_0 bildet die Konstante $c \in \mathbb{R}$ ($\cong L^2_s(T^0)$) auf die Zufallsgröße ab, die konstant gleich c ist; I_1 ist unsere Ausgangsabbildung X . Insbesondere gibt es zu jedem $Z \in L^2(\Omega)$ eine eindeutige Folge $(f_m)_{m \in \mathbb{N}_0}$ mit $f_m \in L^2_s(T^m)$ für alle $m \in \mathbb{N}_0$, derart dass

$$Z = \sum_{m=0}^{\infty} I_m(f_m)$$

gilt. Dies ist die *Wiener-Reihe* zu Z .

Es sei \mathcal{S} die Menge aller $Z \in L^2(\Omega)$, die in der Form $Z = f(X(h_1), \dots, X(h_n))$ mit einer ‘glatten’ Funktion $f : \mathbb{R}^n \rightarrow \mathbb{R}$ und $h_1, \dots, h_n \in L^2(T)$ dargestellt werden können. Für solche Z 's definieren wir die *Malliavin-Ableitung* (in t) durch

$$D_t Z := \sum_{i=1}^n \frac{\partial f}{\partial x_i}(X(h_1), \dots, X(h_n)) h_i(t).$$

Man kann DZ als Element von $L^2(T \times \Omega)$ betrachten; dieser Raum ist isomorph zu

$$L^2(\Omega; L^2(T)) := \{U : \Omega \rightarrow L^2(T) : E\|U\|_{L^2(T)}^2 < \infty\}$$

(bei der ersten Variante geht es um $(t, \omega) \mapsto (D_t Z)(\omega)$, bei der zweiten um $\omega \mapsto (t \mapsto (D_t Z)(\omega))$). Als Zufallsgröße mit Werten in $L^2(T)$ kann DZ als Richtungsableitung interpretiert werden:

$$\begin{aligned} \langle DZ, h \rangle_{L^2(T)} &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left(f(X(h_1) + \epsilon \langle h_1, h \rangle, \dots, X(h_n) + \epsilon \langle h_n, h \rangle) \right. \\ &\quad \left. - f(X(h_1), \dots, X(h_n)) \right). \end{aligned}$$

Der Operator D lässt sich ‘abschließen’; $\mathbb{D}^{1,2}$ bezeichne den Abschluss von \mathcal{S} bzgl.

$$\|Z\|_{1,2} := (E|Z|^2 + E\|DZ\|_{L^2(T)}^2)^{1/2}$$

in $L^2(\Omega)$. Wie interagiert dieser Operator mit der Wiener-Reihe? Ist f von der Form

$$f(t_1, \dots, t_m) = \prod_{i=1}^m 1_{A_i}(t_i),$$

so gilt

$$Z := I_m(f) = \prod_{i=1}^m X(1_{A_i}) = g(X(h_1), \dots, X(h_m))$$

mit $g(x_1, \dots, x_m) = \prod_{i=1}^m x_i$ und $h_i = 1_{A_i}$, also

$$\begin{aligned} D_t(I_m(f)) &= \sum_{i=1}^m \frac{\partial g}{\partial x_i}(X(h_1), \dots, X(h_m)) h_i(t) \\ &= \sum_{i=1}^m \left(\prod_{\substack{j=1, \dots, m \\ j \neq i}} X(1_{A_j}) \right) 1_{A_i}(t) \\ &= \sum_{i=1}^m I_{m-1}(f_i(\cdot, t)) \end{aligned}$$

mit

$$f_i(\cdot, t) : T^{m-1} \rightarrow \mathbb{R}, \quad (x_1, \dots, x_{m-1}) \mapsto f(x_1, \dots, x_{i-1}, t, x_i, \dots, x_{m-1})$$

(beim Nachweis der letzten Gleichheit in der obigen Rechnung betrachtet man die Fälle $t \in A_i$ und $t \notin A_i$ separat). Bei symmetrischen Funktionen ergibt sich eine Vereinfachung, denn es spielt dann keine Rolle, in welcher Position das Argument t eingeschoben wird, und man erhält

$$Z = \sum_{m=0}^{\infty} I_m(f_m) \in \mathbb{D}^{1,2} \implies D_t Z = \sum_{m=1}^{\infty} m I_{m-1}(f_m(\cdot, t)).$$

Formal haben wir somit

$$D : \mathbb{D}^{1,2}(\subset L^2(\Omega)) \rightarrow L^2(T \times \Omega).$$

Der zu D adjungierte Operator δ mit Definitionsbereich $\text{Def}(\delta)$ ist nun das angekündigte *Skorohod-Integral*:

$$\langle Z, \delta(U) \rangle_{L^2(\Omega)} = \langle DZ, U \rangle_{L^2(T \times \Omega)} \quad \text{für alle } Z \in \mathbb{D}^{1,2}, U \in \text{Def}(\delta).$$

Betrachtet man $U = (U_t)_{0 \leq t \leq 1}$ als stochastischen Prozess mit Zeitbereich $T = [0, 1]$, so ordnet δ also dem Prozess eine Zufallsgröße $\delta(U)$ zu, für die gilt:

$$EZ\delta(U) = E \int_0^1 D_t Z U_t dt \quad \text{für alle } Z \in \mathbb{D}^{1,2}.$$

Man beachte, dass bei dem Prozess U keine Vorhersehbarkeit gefordert wird: In der Tat tauchen Filtrationen etc. bisher gar nicht auf. In Analogie zu der obigen Aussage bei der Malliavin-Ableitung haben wir auch beim Skorohod-Integral eine transparente Interaktion mit der Wiener-Reihe: Für $U \in \text{Def}(\delta)$ gilt

$$U_t = \sum_{m=0}^{\infty} I_m(f_m(\cdot, t)) \implies \delta(U) = \sum_{m=0}^{\infty} I_{m+1}(\tilde{f}_m).$$

Hierbei ist $f_m(\cdot, t) \in L_s^2(T^m)$ für alle $t \in T$ und für die zugehörige Symmetrisierung $\tilde{f}_m \in L^2(T^{m+1})$ gilt

$$\begin{aligned} & \tilde{f}_m(t_1, \dots, t_m, t) \\ &= \frac{1}{m+1} \left(f_m(t_1, \dots, t_m, t) + \sum_{i=1}^m f(t_1, \dots, t_{i-1}, t, t_{i+1}, \dots, t_m, t_i) \right). \end{aligned}$$

Man kann zeigen, dass für U 's, die bzgl. der natürlichen Filtration vorhersehbar sind und für die das Itô-Integral existiert, $\delta(U)$ mit diesem Itô-Integral übereinstimmt.