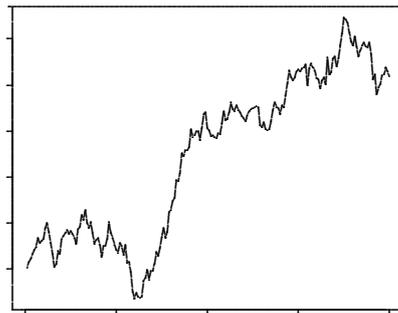


Theorie Stochastischer Prozesse

Karl Grill
Institut für Statistik und Wahrscheinlichkeitstheorie



Version 0.9, März 2011

©Karl Grill

Veröffentlicht unter der Creative Commons Attribution Sharealike License

22. März 2012

Inhaltsverzeichnis

1	Bevor es losgeht	2
2	Grundlagen	4
2.1	Definition eines stochastischen Prozesses	4
2.2	Wichtige Klassen von stochastischen Prozessen	11
2.3	Sigmaalgebren und Stoppzeiten	12
3	Markovprozesse	16
3.1	Allgemeine Definitionen	16
3.2	Markovketten in diskreter Zeit	17
3.3	Markovketten in stetiger Zeit	23
3.3.1	Grundsätzliches	23
3.3.2	Geburts- und Todesprozesse	27
3.3.3	Die Minimallösung	30
3.4	Allgemeine Theorie der Markovprozesse	31
3.4.1	Markovprozesse und Übergangsfunktionen	31
3.4.2	Die starke Markoveigenschaft	33
3.4.3	Trajektorieneigenschaften	35
3.4.4	Übergangsoperatoren	38
3.4.5	Der infinitesimale Operator und Trajektorieneigenschaften	44
4	Martingale	46
4.1	Definiton	46
5	Stochastische Differentialrechnung	56
5.1	Motivation — der Brown'sche Geschwindigkeitsprozess	56
5.2	Das Itô-Integral	58
5.3	Stochastische Differentiale und die Itô-Formel	61
5.4	Stochastische Differentialgleichungen	64

Kapitel 1

Bevor es losgeht

Die Beweise lügen nicht

aus CSI, TV-Serie

... ein paar Anmerkungen:

1. Dieses Skriptum ist “work in progress” und entsprechend mit Bedacht zu verwenden. Sie müssen mit Fehlern, Ungenauigkeiten, fehlenden Beweisen und dergleichen rechnen. Momentan ist meine Hauptsorge, endlich den ganzen Stoff der Vorlesung komplett drin zu haben, was zur Folge hat, dass Beweise und Beispiele und auch sonst einige Dinge einstweilen unausgeführt bleiben. Der Leser ist also gefordert, die Lücken eigenständig mit Hilfe anderer Quellen zu füllen. Offensichtlich gibt es mindestens einen Wikipedia-Artikel, der hierher führt, deswegen möchte ich speziell die Leser, die von dort kommen, auffordern, dieses Skriptum und generell Online-Inhalte kritisch zu betrachten (und hier ist es besonders einfach, weil man nicht denken muss sondern nur rechnen ;-).
2. Weil dieses Skriptum online sichtbar ist, habe ich jetzt auch explizit einen Copyright/Lizenzvermerk angebracht.
3. Wenn Sie mithelfen wollen, dieses Werk zu verbessern, schicken Sie Korrekturen, Vorschläge, etc. an grill@ci.tuwien.ac.at
4. Für die, die bei mir die Prüfung (aus Theorie stochastischer Prozesse) machen wollen: Prüfungsstoff ist der Inhalt dieses Skriptums 14 Tage vor der Prüfung.
5. Mein Dank geht an alle, die mitgeholfen haben, dieses Skriptum zu schreiben, vor allem an Michael Melcher, der die Vorgängerversion dieses Skriptums geschrieben hat, und an Norbert Kusolitsch, einmal für sein Buch

über Maßtheorie, das einiges vorwegnimmt, was sonst hier ausgebreitet werden müsste, und das ich auch immer wieder gern zitiere, und auch dafür, dass er mir den Anstoss gegeben hat, der nötig war, damit sich dieses Skriptum von halbfertig zu halbfertig++ weiterentwickelt.

Kapitel 2

Grundlagen

2.1 Definition eines stochastischen Prozesses

Ich glaube nicht, dass der liebe
Gott würfelt.

Einstein

Nolite cessare ludere. Fraus vobis.

Carmina Burana

Wie immer in der Wahrscheinlichkeitstheorie ist der Ausgangspunkt unserer Überlegungen ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$. In der Wahrscheinlichkeitstheorie betrachtet man Zufallsvariable auf einem solchen Wahrscheinlichkeitsraum, das sind messbare Abbildungen von (Ω, \mathcal{F}) nach $(\mathbb{R}, \mathcal{B})$. Die interessanten Ergebnisse aus der Wahrscheinlichkeitstheorie befassen sich mit Folgen (also abzählbaren Familien) von Zufallsvariablen. Wir gehen ein Stück darüber hinaus und definieren:

Definition 2.1 *T sei eine Teilmenge von \mathbb{R} . Ein stochastischer Prozess auf T ist eine Familie $(\xi(t), t \in T)$ von Zufallsvariablen mit Werten in $X \subset \mathbb{R}^d$.*

Wir werden allgemein Zufallsvariable mit griechischen Buchstaben bezeichnen; für stochastische Prozesse wird in der Literatur auch gerne die Notation ξ_t verwendet, um die Analogie zu Folgen herauszustreichen; wir verwenden die “Funktions”-Notation $\xi(t)$, weil sie praktischer ist, wenn man kompliziertere Ausdrücke als Argument einsetzt.

Die Menge T heißt der *Parameterraum* des Prozesses ξ . Beliebte Wahlen dafür sind \mathbb{N} oder \mathbb{Z} (dann ist der Prozess eine Folge von Zufallsvariablen, bzw. \mathbb{R} oder $[0, \infty)$). Im ersten Fall spricht man von einem Prozess in diskreter Zeit, im anderen von einem Prozess in stetiger Zeit. Es sind auch allgemeinere Parameterräume möglich, insbesondere mehrdimensionale — dann nennt man den Prozess ein *stochastisches Feld*.

Der Wertebereich X der Zufallsvariablen $\xi(t)$ heißt der Phasen- oder Zustandsraum des Prozesses. Neben der Unterscheidung zwischen diskreten (höchstens abzählbaren) und stetigen Zustandsräumen kann man dann natürlich noch zwischen ein- und mehrdimensionalen unterscheiden.

Als nächstes stellt sich die Frage, wie man einen solchen stochastischen Prozess (bzw. seine Verteilung) angeben kann. Dazu muss man jedenfalls alle endlichdimensionalen Verteilungen $\mathbb{P}_{\xi(t_1), \dots, \xi(t_n)}$ angeben. Diese Familie von Verteilungen muss die Bedingung erfüllen, dass die Verteilung $\mathbb{P}_{\xi(t_1), \dots, \xi(t_n)}$ mit der entsprechenden Randverteilung von $\mathbb{P}_{\xi(t_1), \dots, \xi(t_{n+1})}$ übereinstimmt (wobei die t_i nicht geordnet sein müssen) — wir sprechen von einer *konsistenten* Familie. Diese Bedingung ist aber nicht nur notwendig, sondern auch hinreichend. Es gilt der

Satz 2.1 (Existenzsatz von Kolmogorov) *Zu jeder konsistenten Familie von Verteilungen gibt es einen Prozess $\xi(t)$ mit diesen endlichdimensionalen Verteilungen. Außerdem ist die Einschränkung des zugehörigen Wahrscheinlichkeitsmaßes \mathbb{P} auf die Sigmaalgebra \mathcal{F}_∞ , die von den Zufallsvariablen $\xi(t), t \in T$ erzeugt wird (d.h., die kleinste Sigmaalgebra, bezüglich der alle $\xi(t)$ messbar sind) eindeutig bestimmt.*

Der Beweis wurde schon in der Maß- und Wahrscheinlichkeitstheorie geführt.

Ein stochastischer Prozess wird also durch seine endlichdimensionalen (Rand-) Verteilungen bestimmt. Das führt uns zu den folgenden Definitionen:

Definition 2.2 *Zwei Prozesse ξ und η heißen äquivalent, wenn sie dieselben endlichdimensionalen Verteilungen haben. Der Prozess η heißt dann eine Version des Prozesses ξ .*

Offensichtlich sind zwei Prozesse auf demselben Wahrscheinlichkeitsraum äquivalent, wenn für jedes $t \in T$ $\xi(t)$ und $\eta(t)$ fast sicher übereinstimmen. In diesem Fall gibt es für jedes t eine Nullmenge, auf deren Komplement $\xi(t)$ und $\eta(t)$ übereinstimmen. Wenn wir fordern, dass diese Nullmenge für alle t dieselbe ist, gelangen wir zur

Definition 2.3 *Die Prozesse ξ und η auf demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ heißen ununterscheidbar, wenn*

$$\mathbb{P}^*(\exists t \in T : \xi(t) \neq \eta(t)) = 0.$$

Wir haben hier das äußere Maß geschrieben, weil im allgemeinen nicht sichergestellt ist, dass dieses “Ereignis” messbar ist. Die Ununterscheidbarkeit zweier Prozesse impliziert jedenfalls ihre Äquivalenz.

Die fast sichere Übereinstimmung von $\xi(t)$ und $\eta(t)$ sichert die Ununterscheidbarkeit von ξ und η im allgemeinen nur dann, wenn der Parameterraum abzählbar ist.

Das prominenteste Beispiel eines Prozesses in stetiger Zeit ist der Wienerprozess, der uns immer wieder begegnen wird:

Definition 2.4 Der Wienerprozess β ist durch die folgenden Bedingungen charakterisiert:

1. $\beta(0) = 0$,
2. für $t_0 < t_1 < \dots < t_n$ sind die Zufallsvariablen $(\beta(t_i) - \beta(t_{i-1}), i = 1, \dots, n)$ unabhängig,
3. für $t_0 < t_1$ ist $\beta(t_1) - \beta(t_0)$ normalverteilt mit Mittel 0 und Varianz $t_1 - t_0$.

Ein d -dimensionaler Wienerprozess ist ein Vektor, dessen Komponenten unabhängige eindimensionale Wienerprozesse sind.

Man überprüft leicht, dass diese Bedingungen eine konsistente Familie von Verteilungen festlegen. Der Wienerprozess wird auch gerne als “Brownsche Bewegung” bezeichnet; allerdings gibt es außer dem Wienerprozess noch andere Prozesse, die als Modelle für dieses physikalische Phänomen angesehen werden können.

Genau genommen ist der Prozess ξ eine Funktion $\xi(t, \omega)$ von zwei Variablen, weil ja die Zufallsvariable $\xi(t)$ selbst wieder eine Abbildung von Ω in die reellen Zahlen ist. Wir können also, je nachdem welche der beiden Variablen festgehalten wird, einen stochastischen Prozess auf zwei Arten betrachten: für fixes t ist $\xi(t, \omega)$ eine Zufallsvariable, und wir sind bei der Ansicht des Prozesses als eine Familie von Zufallsvariablen; für fixes ω ist $\xi(t, \omega)$ eine Funktion von t , wir können ξ also auch als eine zufällig gewählte Funktion ansehen. Wenn wir $\xi(t)$ für fixes ω betrachten, sprechen wir von einer *Trajektorie* oder einem *Pfad* des Prozesses.

Wenn man sich nun für gewisse Eigenschaften eines Prozesses interessiert, etwa für die Stetigkeit, kann man einmal den ersten Gesichtspunkt einnehmen und einen Prozess ξ als stetig in einem gewissen Sinn bezeichnen, wenn für $s \rightarrow t$ die Zufallsvariablen $\xi(s)$ in diesem Sinn gegen $\xi(t)$ konvergieren. Wir definieren also:

Definition 2.5 Der Prozess ξ heißt

1. stetig in Wahrscheinlichkeit oder stochastisch stetig, wenn für alle $t \in T$

$$\lim_{s \rightarrow t} \xi(s) = \xi(t) \text{ in Wahrscheinlichkeit,}$$

d.h., wenn für jedes $\epsilon > 0$

$$\lim_{s \rightarrow t} \mathbb{P}(|\xi(s) - \xi(t)| > \epsilon) = 0$$

gilt.

2. stetig im Quadratmittel, wenn für alle $t \in T$

$$\lim_{s \rightarrow t} \mathbb{E}((\xi(s) - \xi(t))^2) = 0.$$

3. *fast sicher stetig, wenn für alle $t \in T$*

$$\mathbb{P}(\lim_{s \rightarrow t} \xi(s) = \xi(t)) = 1.$$

Diese Definitionen sind sehr schön, aber sie lassen etwas zu wünschen übrig. Es sei nämlich τ gleichverteilt auf $[0, 1]$, und

$$\xi(t) = \begin{cases} 1 & \text{wenn } t \geq \tau, \\ 0 & \text{sonst.} \end{cases}$$

Dann ist ξ in jedem der drei Sinne unserer letzten Definition stetig, aber offensichtlich sind die Pfade dieses Prozesses keine stetigen Funktionen.

Wir ändern also jetzt unseren Standpunkt und fragen danach, ob der Prozess ξ *stetige Pfade* besitzt. Auch diese Frage ist nicht so leicht zu beantworten, wie das folgende (zugegebenermaßen primitive) Beispiel zeigt:

Es sei wieder τ gleichverteilt auf $[0, 1]$ und

$$\begin{aligned} \xi(t) &= 0 \\ \eta(t) &= \begin{cases} 1 & \text{wenn } t = \tau, \\ 0 & \text{sonst.} \end{cases} \end{aligned}$$

ξ und η sind äquivalent, aber ξ hat stetige Trajektorien, η nicht. Es ist also allein durch die Angabe der endlichdimensionalen Verteilungen nicht festgelegt, ob die Trajektorien stetig sind (der tiefere Grund dafür ist, dass das ‘Ereignis’ ‘ $\xi(t)$ ist stetig’ nicht \mathcal{F}_∞ -messbar ist). Wir müssen deshalb unsere Frage umformulieren, und fragen, ob es eine Version des Prozesses gibt, die stetige Trajektorien besitzt. Eine einfache hinreichende Bedingung dafür liefert der folgende

Satz 2.2 (Kolmogorov’sches Stetigkeitskriterium) *Wenn es $a > 0$, $b > 1$ und $K < \infty$ gibt, sodass für alle $s, t \in T$*

$$\mathbb{E}(|\xi(s) - \xi(t)|^a) \leq K|s - t|^b,$$

dann hat ξ eine Version mit stetigen Trajektorien.

Wir beweisen den Satz für $s, t \in [0, 1]$. Der allgemeine Fall folgt daraus durch ‘Aneinanderstückeln’.

Wir definieren Ereignisse

$$A_n = [\text{es gibt } k \leq 2^n : |\xi(\frac{k}{2^n}) - \xi(\frac{k-1}{2^n})| > q^n],$$

wobei wir $2^{1-b} < q < 1$ wählen.

Nach der Ungleichung von Markov ergibt sich

$$\mathbb{P}(A_n) \leq \frac{K}{(2^{b-1}q)^n}.$$

Die Reihe

$$\sum \mathbb{P}(A_n)$$

konvergiert also, und das Lemma von Borel-Cantelli liefert, dass

$$\mathbb{P}(\limsup_{n \rightarrow \infty} A_n) = 0,$$

oder anders gesagt: mit Wahrscheinlichkeit eins können nur endlich viele der Ereignisse A_n auftreten. Es gibt also mit Wahrscheinlichkeit 1 ein n_0 , sodass für $n \geq n_0$ und $k \leq 2^n$

$$|\xi(\frac{k}{2^n}) - \xi(\frac{k-1}{2^n})| \leq q^n.$$

daraus folgt, dass für zwei dyadische Rationalzahlen s und t mit $|s - t| \leq 2^{-n}$

$$|\xi(s) - \xi(t)| \leq \frac{2q^n}{1 - q}.$$

Wir setzen jetzt für jede dyadische Rationalzahl t

$$\eta(t, \omega) = \begin{cases} 0 & \text{falls } \omega \in \limsup A_n, \\ \xi(t) & \text{sonst.} \end{cases}$$

Für alle anderen t setzen wir $\eta(t)$ gleich dem Limes von $\eta(s_n)$, wobei s_n eine Folge von dyadischen Rationalzahlen ist, die gegen t konvergiert (nach dem oben gesagten existiert dieser Grenzwert, weil $\eta(s_n)$ eine Cauchyfolge bildet). Weil außerdem (wieder als Folge aus der Markov-Ungleichung und weil $\eta(s_n)$ fast sicher mit $\xi(s_n)$ übereinstimmt) die Folge $\eta(s_n)$ in Wahrscheinlichkeit gegen $\xi(t)$ konvergiert, stimmen $\eta(t)$ und $\xi(t)$ mit Wahrscheinlichkeit eins überein, und durch ein zusätzliches Cauchyfolgenargument überzeugt man sich, dass $\eta(t)$ stetige Trajektorien hat. η ist also genau die gesuchte Version von ξ mit stetigen Pfaden.

Für eine normalverteilte Zufallsvariable ξ mit Varianz σ^2 und Mittel 0 gilt

$$\mathbb{E}(|\xi|^\alpha) = C_\alpha \sigma^\alpha.$$

Daraus können wir schließen, dass es eine Version des Wienerprozesses gibt, die stetige Trajektorien besitzt. Eine solche Version bezeichnen wir als *Standard-Wienerprozess*.

Im allgemeinen ist für die Existenz einer Version eines Prozesses, der zu ξ äquivalent ist und fast sicher stetige (rechtsstetige, beschränkte, ...) Trajektorien besitzt, notwendig und hinreichend, dass das äußere Maß \mathbb{P}^* der entsprechenden Menge von Funktionen gleich 1 ist. Das ist gleichbedeutend damit, dass jede messbare Menge, die die Menge der stetigen (rechtstetigen, etc.) Funktionen umfasst, Wahrscheinlichkeit 1 hat. Die Produktsigmaalgebra \mathcal{B}^T lässt sich in der Form

$$\mathcal{B}^T = \bigcup_{S \subset T, |S| \leq \aleph_0} \mathcal{B}^S$$

schreiben, was zur Folge hat, dass sich die messbaren Obermengen einer Menge A von Funktionen jeweils aus allen Funktionen bestehen, die auf einer abzählbaren Menge $S \subset T$ mit einer der Funktionen aus A übereinstimmen. Es bleibt

also die Aufgabe, Kriterien für Eigenschaften wie die Stetigkeit zu bestimmen, die sich an beliebigen abzählbaren Teilmengen des Parameterraums überprüfen lassen. Von einer stetigen Funktion auf einem kompakten Intervall weiß man, dass sie gleichmäßig stetig ist, und diese gleichmäßige Stetigkeit bleibt natürlich auch für eine Einschränkung des Definitionsbereichs erhalten. Deshalb gilt:

Satz 2.3 *Der Prozess ξ hat genau dann eine stetige Version, wenn auf ξ auf jeder beschränkten abzählbaren Teilmenge von T mit Wahrscheinlichkeit 1 gleichmäßig stetig ist.*

Wenn man sich nur für die Rechtsstetigkeit interessiert, gibt es das folgende Kriterium:

Zuerst definieren wir

Definition 2.6 *$a < b$ seien zwei reelle Zahlen. Die Funktion ξ hat auf der Menge S n Durchquerungen ("upcrossings"), wenn es $s_1 < t_1 < \dots < s_n < t_n \in S$ gibt, für die $\xi(s_i) < a$ und $\xi(t_i) > b$ gilt.*

Damit gilt:

Satz 2.4 *Ein Prozess ξ besitzt eine Version, deren Trajektorien mit Wahrscheinlichkeit eins in allen Punkten rechts- und linksseitige Grenzwerte haben, wenn für jede abzählbare beschränkte Menge $S \subset T$ und für alle rationalen Zahlen $a < b$ die Anzahl der Durchquerungen der des Intervalls (a, b) mit Wahrscheinlichkeit eins endlich ist. Ist die Anzahl der Durchquerungen auf beliebigen abzählbaren Teilmengen von T endlich, dann existiert auch der Grenzwert $\lim_{t \rightarrow \infty} \xi(t)$.*

Zum Beweis muss man sich nur davon überzeugen, dass die Endlichkeit der Anzahl der Durchquerungen auf beliebigen abzählbaren Teilmengen des Definitionsbereichs für deterministische Funktionen äquivalent zur Existenz der einseitigen Grenzwerte ist.

Für die Rechtsstetigkeit der Trajektorien (für eine geeignete Version des Prozesses) muss dann nur noch sichergestellt sein, dass der rechtsseitige Grenzwert mit dem Wert des Prozesses übereinstimmt, also dass der Prozess in Wahrscheinlichkeit rechtsstetig ist.

Wenn man mit den Trajektorien des Prozesses rechnen will (etwa ihr Integral berechnen), sollten sie (Borel-) messbar sein. Wir wollen natürlich dann auch sicherstellen, dass das Ergebnis dieser Rechnung eine Zufallsvariable ist, und deshalb definieren wir

Satz 2.5 *Der Prozess ξ heißt messbar, wenn $\xi(t, \omega)$ als Funktion beider Argumente $\mathcal{B} \otimes \mathcal{F}$ -messbar ist.*

Wenn man die beiden Konzepte der Adaptiertheit und der Messbarkeit kombiniert, kommt man zu

Definition 2.7 *Der Prozess ξ heißt bezüglich der Filtration (\mathcal{F}_t) progressiv messbar, wenn für jedes t $\xi(s, \omega)$ auf $[0, t] \times \Omega$ $(\mathcal{B} \cap [0, t]) \otimes \mathcal{F}_t$ -messbar ist.*

Es gilt:

Satz 2.6 *Rechtsstetige Prozesse sind (progressiv) messbar.*

Satz 2.7 *Falls für die beiden rechtsstetigen Prozesse ξ und η für alle $t \in T$ $\xi(t)$ und $\eta(t)$ fast sicher übereinstimmen, dann sind ξ und η ununterscheidbar.*

Beweise als Übung.

Die schwächste Forderung, die man an die Stetigkeit eines Prozesses stellen kann, ist die *Separabilität*:

Definition 2.8 *Ein Prozess ξ heißt separabel, wenn es eine abzählbare dichte Teilmenge S von T gibt, sodass für jedes $t \in T$ $\xi(t)$ in der Menge der Häufungspunkte von Folgen $\xi(s_n)$ mit $s_n \in S$ und $s_n \rightarrow t$ liegt. Eine Menge S , die diese Forderung erfüllt, heißt separable Menge für den Prozess ξ .*

Es gilt der folgende Satz:

Satz 2.8 *Zu jedem stochastischen Prozess gibt es eine separable Version (es kann allerdings notwendig sein, dass die separable Version die Werte $\pm\infty$ annimmt).*

Zum Beweis definieren wir für $a < b$ und $c < d$

$$\lambda(a, b, c, d) = \sup\{\mathbb{P}(\exists t \in [a, b] \cap S : \xi(t) \in [c, d]) : S \subset T, |S| \leq \aleph_0\}.$$

Es ist klar, dass dieses Supremum angenommen wird. Wir wählen eine abzählbare Menge $S(a, b, c, d)$, für die

$$\lambda(a, b, c, d) = \mathbb{P}(\exists t \in [a, b] \cap S(a, b, c, d) : \xi(t) \in [c, d])$$

und setzen

$$S = \bigcup_{a < b, c < d, a, b, c, d \in \mathbb{Q}} S(a, b, c, d).$$

Es ist leicht zu sehen, dass $\xi(t)$ mit Wahrscheinlichkeit eins in der Menge der Häufungspunkte von $\xi(s)$ für $s \in S, s \rightarrow t$ liegen muss. Andernfalls gäbe es nämlich ein rationales Intervall $[a, b] \ni t$ und ein rationales Intervall $[c, d]$, sodass mit positiver Wahrscheinlichkeit $\xi(t) \in [c, d]$ und $\xi(s) \notin [c, d]$ für alles $s \in S \cap [a, b]$. Dann ist aber für $S' = S(a, b, c, d) \cup \{t\}$

$$\mathbb{P}(\exists t \in [a, b] \cap S' : \xi(t) \in [c, d]) > \lambda(a, b, c, d),$$

im Widerspruch zur Definition von λ . Wir können also eine separable Version von ξ erhalten, indem wir dann, wenn $\xi(t)$ nicht in der Menge der Häufungspunkte enthalten ist (was ja mit Wahrscheinlichkeit 0 geschieht), $\xi(t)$ durch einen beliebigen Häufungspunkt ersetzen (etwa den Limes superior).

2.2 Wichtige Klassen von stochastischen Prozessen

I ain't looking to [...]
 simplify you, classify you [...],
 analyze you, categorize you [...],
 or confine you or define you.

Bob Dylan, All I really want to do

In gewisser Weise sind die meisten stochastischen Prozesse Verallgemeinerungen der Idee einer Folge von unabhängigen identisch verteilten Zufallsvariablen bzw. von Summen davon. Eine fundamentale Eigenschaft einer Folge von unabhängigen identisch verteilten Zufallsvariablen ist die Tatsache, dass sich an der Verteilung dieser Folge nichts ändert, wenn man endlich viele Elemente am Anfang entfernt. Diese Eigenschaft isoliert betrachtet gibt Anlass zur folgenden

Definition 2.9 *Ein stochastischer Prozess heißt stationär, wenn für alle $t_1, \dots, t_n \in T$ und alle h die Vektoren $(\xi(t_1), \dots, \xi(t_n))$ und $(\xi(t_1+h), \dots, \xi(t_n+h))$ dieselbe Verteilung haben.*

Für manche Zwecke ist es ausreichend, wenn man sich nicht um die exakten Verteilungen eines Prozesse kümmert, sondern nur die ersten beiden Momente betrachtet (deren Existenz man natürlich voraussetzen muss). Dann bezeichnet man $m(t) = \mathbb{E}(\xi(t))$ als die Mittelwertfunktion und $R(s, t) = \mathbf{Cov}(\xi(s), \xi(t))$ als die Kovarianzfunktion des Prozesses, und man kann definieren

Definition 2.10 *Ein Prozess ξ mit endlicher Varianz heißt schwach stationär, wenn für seine Mittelwerts- und Kovarianzfunktionen gilt:*

$$m(t) = m,$$

$$R(s, t) = R(|s - t|).$$

Es muss also die Mittelwertfunktion konstant sein und die Kovarianz nur von der Differenz ihrer Argumente abhängen.

Wenn man nun von einer Summe von unabhängigen Zufallsvariablen ausgeht, so ist die direkte Übertragung in die Welt der stochastischen Prozesse der Prozess mit unabhängigen Zuwächsen

Definition 2.11 *Ein stochastischer Prozess ξ heißt ein Prozess mit unabhängigen Zuwächsen, wenn für $t_1 < t_2 < \dots < t_n$ die Zufallsvariablen*

$$\xi(t_1), \xi(t_2) - \xi(t_1), \dots, \xi(t_n) - \xi(t_{n-1})$$

unabhängig sind.

Zur weiteren Abstraktion stellt man fest, dass in einer Summe von unabhängigen Zufallsvariablen die weitere Entwicklung nach dem Zeitpunkt n nicht von der Vergangenheit davor abhängt (abgesehen von dem Wert S_n zum Zeitpunkt n). Diese Eigenschaft gibt Anlass zu der

Definition 2.12 Ein stochastischer Prozess ξ heißt ein Markovprozess, wenn die Verteilung des Prozess nach dem Zeitpunkt t von der Vergangenheit vor t nur über $\xi(t)$ abhängt, formal ist für $s_1 < \dots < s_n < s < t$ und für jede Borelmenge B

$$\mathbb{P}(\xi(t) \in B | \xi(s_1), \dots, \xi(s_n), \xi(s)) = \mathbb{P}(\xi(t) \in B | \xi(s))$$

(hier und für die weitere Zukunft sei angemerkt, dass Gleichungen, die bedingte Wahrscheinlichkeiten bzw. Erwartungen enthalten, grundsätzlich nur fast sicher zu verstehen sind — vgl. dazu auch die Anmerkungen im nächsten Kapitel. Wir werden in Hinkunft auf diese Tatsache nicht gesondert hinweisen).

Prozesse mit unabhängigen Zuwächsen sind Markovprozesse, ebenso Prozesse, für die alle $\xi(t)$ unabhängig sind - die letzteren haben allerdings in stetiger Zeit einige unangenehme Eigenschaften, beispielsweise sind ihre Trajektorien mit Wahrscheinlichkeit eins nirgends stetig.

Eine zweite Abstraktion von der Vorstellung einer Summe von unabhängigen Zufallsvariablen stellen die Martingale dar. Diese Repräsentieren die Vorstellung eines fairen Spiels — ein solches Spiel ist dadurch gekennzeichnet, dass der Spieler im Mittel weder dazugewinnt noch verliert. Formal sieht das so aus:

Definition 2.13 Ein Prozess $\xi(t)$ heißt Martingal, wenn $\mathbb{E}(\xi(t_i))$ endlich ist und für $t_1 < t_2 < \dots < t_n < s < t$

$$\mathbb{E}(\xi(t) | \xi(t_1), \dots, \xi(t_n), \xi(s)) = \xi(s).$$

Prozesse mit unabhängigen Zuwächsen sind Martingale, wenn die Zuwächse Erwartungswert null haben. Die Bezeichnung “Martingal” wurde früher (in Frankreich?) für die Strategie des Verdoppelns des Einsatzes bei Verlust verwendet. Das Kapital eines Spielers, der so spielt (wenn er mit Wahrscheinlichkeit $1/2$ den Einsatz dazugewinnt) ist ein Beispiel für ein Martingal, das kein Prozess mit unabhängigen Zuwächsen ist.

Definition 2.14 ξ heißt Gaussprozess, wenn die endlichdimensionale Randverteilungen (mehrdimensionale) Normalverteilungen sind.

Gaussprozesse haben den Vorteil, dass sie durch Mittelwert- und Kovarianzfunktion eindeutig bestimmt werden; insbesondere sind für Gaussprozesse die starke und schwache Stationarität äquivalent.

Der Wienerprozess ist als Prozess mit unabhängigen (und normalverteilten) Zuwächsen ein Gaussprozess, Markovprozess und Martingal.

2.3 Sigmaalgebren und Stoppzeiten

Ich glaube, Sie haben jetzt
“Stopp” gesagt

aus CSI, TV-Serie

In der Theorie der stochastischen Prozesse wird der Parameter t oft und gerne als die Zeit interpretiert. Dabei kommt man zum Beispiel auf die folgende Frage: der Verlauf des Prozesses bis zum Zeitpunkt t_0 sei bekannt. Wie kann man den Wert $\xi(t)$ zu einem Zeitpunkt $t > t_0$ möglichst gut schätzen? Ohne jetzt über Details zu sprechen (etwa, wie die Güte einer Schätzung beurteilt werden soll), stehen wir jedenfalls vor dem Problem, dass wir definieren müssen, was wir damit meinen, wenn wir sagen, dass wir den Schätzwert für $\xi(t)$ aus den Werten $\xi(s)$ für $s \leq t_0$ ausrechnen. Im Fall eines diskreten Parameterraums ist die Antwort ziemlich einfach: unsere Schätzung wird einfach eine Funktion $f(\xi(0), \xi(1), \dots, \xi(t_0))$ der beobachteten Werte sein. Für stetige Parameterräume müssten wir hier eine Funktion von unendlich (sogar überabzählbar) vielen Argumenten schreiben. Das ist nicht nur nicht ganz leicht hinzuschreiben, sondern auch formal nicht befriedigend. Eine bessere Idee ist, alle die Schätzungen zuzulassen, die sich als Grenzwerte von Zufallsvariablen der Form $f_n(\xi(t_1), \dots, \xi(t_n))$ schreiben lassen, wobei natürlich $t_i \leq t_0, i = 1, \dots, n$ gelten soll. Das ist gleichbedeutend damit, dass die Schätzung eine Zufallsvariable ist, die messbar bezüglich der Sigmaalgebra \mathcal{F}_{t_0} ist, die von den Zufallsvariablen $\xi(s)$ für $s \leq t_0$ erzeugt wird (das ist die kleinste Sigmaalgebra, bezüglich der alle $\xi(s)$ messbar sind). Wir erhalten so zu jedem $t \in T$ eine Sigmaalgebra \mathcal{F}_t , und es gilt natürlich, dass für $s \leq t$ $\mathcal{F}_s \subseteq \mathcal{F}_t$ gilt. Wir definieren jetzt

Definition 2.15 *eine Familie $(\mathcal{F}_t, t \in T)$ von Sigmaalgebren heißt eine Filtration, wenn für $s < t$ $\mathcal{F}_s \subseteq \mathcal{F}_t$ gilt. Ist insbesondere $(\xi(t), t \in T)$ ein stochastischer Prozess und \mathcal{F}_t die Sigmaalgebra, die von den Zufallsvariablen $\xi(s)$ mit $s \leq t$ erzeugt wird, dann nennen wir $(\mathcal{F}_t, t \in T)$ die natürliche Filtration von ξ .*

Anschaulich interpretiert ist \mathcal{F}_t die Menge aller Ereignisse, über deren Eintreten wir zum Zeitpunkt t Bescheid wissen. Genauso sind die \mathcal{F}_t -messbaren Zufallsvariablen genau die, deren Wert wir zum Zeitpunkt t mit Sicherheit kennen.

Es macht Sinn, allgemeinere Filtrationen als die natürlichen Filtrationen zu betrachten — es kann ja sein, dass wir außer dem Verlauf des Prozesses noch zusätzliche Informationen zur Verfügung haben, also über mehr Ereignisse Bescheid wissen als nur die, die durch ξ bestimmt sind. Was wir allerdings sinnvollerweise fordern müssen, ist, dass wir zu jedem Zeitpunkt den Wert von $\xi(t)$ kennen. Das gibt Anlass zur folgenden

Definition 2.16 *Der Prozess $\xi(t)$ heißt adaptiert an die Filtration \mathcal{F}_t , wenn für jedes $t \in T$ die Zufallsvariable $\xi(t)$ bezüglich \mathcal{F}_t messbar ist.*

Der bedingte Erwartungswert $\mathbb{E}(\eta|\mathcal{F})$ ist, anschaulich gesprochen, das Mittel aller Werte von η , die noch auftreten können, wenn uns die Information über die Ereignisse in \mathcal{F} bekannt ist. Formal haben wir

Definition 2.17 *Die bedingte Erwartung $\mathbb{E}(\eta|\mathcal{F})$ ist eine \mathbb{F} -messbare Zufallsvariable, für die*

$$\mathbb{E}(I_A(\eta|\mathcal{F})) = \mathbb{E}(\eta I_A)$$

für alle $A \in \mathcal{F}$ gilt (eine äquivalente Definition erhält man, wenn man diese Gleichung nicht nur für Indikatorfunktionen, sondern für alle beschränkten \mathcal{F} -messbare Funktionen verlangt). Für ein Ereignis B heißt

$$\mathbb{P}(B|\mathcal{F}) = \mathbb{E}(I_B|\mathcal{F})$$

die bedingte Wahrscheinlichkeit von B unter \mathcal{F} . Ist ξ eine weitere \mathcal{F} -messbare (ein- oder mehrdimensionale) Zufallsvariable, definieren wir

$$\mathbb{E}(\eta|\xi) = \mathbb{E}(\eta|\xi^{-1}(\mathcal{B}))$$

als die bedingte Erwartung von η unter ξ ; nach einem Satz aus der Maßtheorie gibt es dann eine borelmessbare Funktion g , sodass

$$\mathbb{E}(\eta|\xi) = g(\xi)$$

gilt. Damit setzen wir

$$\mathbb{E}(\eta|\xi = x) = g(x).$$

Die bedingten Wahrscheinlichkeiten $\mathbb{P}(A|\xi)$ und $\mathbb{P}(A|\xi = x)$ werden wieder als die bedingten Erwartungen von I_A definiert.

Der Satz von Radon-Nikodym garantiert die Existenz der bedingten Erwartung, allerdings nur bis auf fast sichere Übereinstimmung; klarerweise ist jede Funktion, die sich von einer bedingten Erwartung nur auf einer Menge von Maß 0 unterscheidet, ebenfalls eine bedingte Erwartung. Für diskrete (bedingende) Zufallsvariable ξ und $\mathbb{P}(\xi = x) > 0$ ergibt sich natürlich die klassische bedingte Wahrscheinlichkeit. Mit dieser Definition lassen sich die Definitionen der stochastischen Prozesse aus dem vorigen Abschnitt etwas knapper formulieren. Etwa ist ein Prozess mit unabhängigen Inkrementen ein Prozess, für den $\xi(t) - \xi(s)$ für alle $s < t$ von \mathcal{F}_s unabhängig ist, und ein Martingal wird durch die Bedingung

$$\mathbb{E}(\xi(t)|\mathcal{F}_s) = \xi(s)$$

definiert. Dabei kann man auch allgemeinere Filtrationen als die natürliche Filtration von ξ einsetzen (man spricht dann von einem \mathcal{F}_t -Martingal, um es vom kanonischen Fall zu unterscheiden).

Der nächste wichtige Begriff ist der einer Stoppzeit. Man kann sich wieder vorstellen, dass der stochastische Prozess $\xi(t)$ den Verlauf eines Spiels beschreibt, und dass der Spieler nach jedem Spiel aufgrund des bisherigen Spielverlaufs entscheidet, ob er weiterspielt oder nicht. Es ist also zu jedem Zeitpunkt aus dem Spielverlauf ersichtlich, ob der Spieler schon aufgehört (gestoppt) hat oder nicht. Formal ergibt das

Definition 2.18 Eine Zufallsvariable τ heißt Stoppzeit bezüglich der Filtration (\mathcal{F}_t) , wenn Für jedes $t \in T$ das Ereignis $[\tau \leq t]$ bezüglich \mathcal{F}_t messbar ist.

Es gilt

Satz 2.9 $\sigma, \tau, \tau_1, \tau_2, \dots$ seien Stoppzeiten. Dann sind auch $\min(\sigma, \tau)$, $\max(\sigma, \tau)$ und $\sup_n \tau_n$ Stoppzeiten.

Zum Beweis stellen wir fest, dass wegen

$$[\min(\sigma, \tau) \leq t] = [\sigma \leq t] \cup [\tau \leq t],$$

$$[\max(\sigma, \tau) \leq t] = [\sigma \leq t] \cap [\tau \leq t],$$

und

$$[\sup_n \tau_n \leq t] = \bigcap_n [\tau_n \leq t]$$

die Ereignisse auf der linken Seite alle \mathcal{F}_t -messbar sind.

Wenn τ eine Stoppzeit ist, bezeichnen wir mit \mathcal{F}_τ die Menge aller Ereignisse, über deren Eintreffen wir im Moment des Stoppens Bescheid wissen:

Definition 2.19 τ sei eine Stoppzeit bezüglich der Filtration (\mathcal{F}_t) . Dann heißt

$$\mathcal{F}_\tau = \{A : A \cap [\tau \leq t] \in F_t \text{ für alle } t \in T\}$$

die Vergangenheit von τ .

Ein typisches Beispiel für eine Stoppzeit ist die Übergangszeit τ_C in eine Menge C . Diese ist definiert als das Infimum aller t , für die $\xi(t)$ in C liegt (im eigentlichen Sinn ist das nur dann eine Stoppzeit, wenn der Parameterraum diskret ist oder wenn C abgeschlossen ist und $\xi(t)$ rechtsstetige Trajektorien hat, man kann aber τ_C zu einer Stoppzeit machen, indem man die umfangreichere Filtration $\mathcal{F}_{t+} = \bigcap_{s>t} F_s$ betrachtet, das ist aber nicht immer wünschenswert, weil dann etwa nicht mehr garantiert ist, dass $\xi(\tau_C) \in C$ gilt).

Kapitel 3

Markovprozesse

3.1 Allgemeine Definitionen

You know I'd love to live with you
but you make me forget so very
much

Leonard Cohen, So long Marianne

Wie wir bereits wissen, ist ein Markovprozess dadurch charakterisiert, dass für $s < t$ die Verteilung von $\xi(t)$ von der Vergangenheit bis s nur über den Wert von ξ zum Zeitpunkt s abhängt. Es ist also

$$\mathbb{P}(\xi(t) \in A | \mathcal{F}_s) = \mathbb{P}(\xi(t) \in A | \xi(s)).$$

Wir teilen die Markovprozesse in Klassen ein, je nachdem, welche Werte wir für den Parameter- und den Zustandsraum zulassen. Ist der Zustandsraum diskret, sprechen wir von einer Markovkette und unterscheiden weiter in Markovketten in stetiger oder diskreter Zeit (manchmal auch kurz diskrete bzw. stetige Markovketten). Im anderen Fall, wenn also der Zustandsraum stetig ist, sprechen wir von einem eigentlichen Markovprozess (der Fall, dass der Parameterraum diskret ist und der Zustandsraum stetig, wird auch bisweilen als eine Markovkette bezeichnet. Wir werden diesen Fall nicht betrachten).

Ein Markovprozess wird durch die *Übergangswahrscheinlichkeiten* oder *Übergangsfunktion*

$$P(x, s, t, A) = \mathbb{P}(\xi(t) \in A | \xi(s) = x)$$

zusammen mit der Anfangsverteilung

$$P(0, A) = \mathbb{P}(\xi(0) \in A)$$

festgelegt. Von besonderer Bedeutung ist der Fall, dass die Übergangsfunktion nur von der Differenz $t - s$ abhängt. In diesem Fall ist der weitere Verlauf des Prozesses, wenn er zum Zeitpunkt t in x ist, derselbe wie für einen Prozess, der zum Zeitpunkt 0 in x startet. Solche Prozesse heißen *homogene Markovprozesse*.

3.2 Markovketten in diskreter Zeit

Wir beschäftigen uns jetzt also mit Markovprozessen, für die sowohl der Parameterraum oder der Zustandsraum höchstens abzählbar sind. Weil der Zustandsraum diskret ist, genügt es, die Übergangsfunktion für einelementige Mengen A anzugeben. Weiters werden wir uns auf homogene Markovketten beschränken. In diesem Fall hängt

$$p_{ij}(t) = P(i, s, s + t, \{j\}) = \mathbb{P}(\xi(s + t) = j | \xi(s) = i)$$

nicht von s ab. Wir nennen $p_{ij}(t)$ die t -stufigen Übergangswahrscheinlichkeiten. Aus dem Satz von der vollständigen Wahrscheinlichkeit und der Markoveigenschaft erhält man die *Chapman-Kolmogorov Gleichungen*

$$p_{ik}(s + t) = \sum_j p_{ij}(s)p_{jk}(t).$$

wenn man die $p_{ij}(t)$ zu der (endlichen oder unendlichen) Matrix

$$P(t) = (p_{ij}(t))_{X \times X},$$

der *Übergangsmatrix* zusammenfasst, dann erhalten die Chapman-Kolmogorov Gleichungen die Form

$$P(s + t) = P(s)P(t)$$

bzw.

$$P(t) = P^t,$$

wenn wir $P = P(1)$ setzen. Aus der Matrix P der einstufigen Übergangswahrscheinlichkeiten p_{ij} kann man also alle Übergangswahrscheinlichkeiten berechnen.

Da in der Folge bedingte Wahrscheinlichkeiten bezüglich des Startwertes sehr oft vorkommen, definieren wir für beliebige Ereignisse A und Zufallsvariable η

$$\mathbb{P}_i(A) := \mathbb{P}(A | \xi(0) = i)$$

und

$$\mathbb{E}_i(\eta) := \mathbb{E}(\eta | \xi(0) = i).$$

Wir bezeichnen den Zustand j als Nachfolger von i , wenn es ein $t \geq 0$ gibt, sodass $p_{ij}(t) > 0$, und schreiben dafür

$$i \rightarrow j.$$

Gilt sowohl $i \rightarrow j$ als auch $j \rightarrow i$, dann nennen wir i und j verbunden oder kommunizierend und schreiben

$$i \leftrightarrow j.$$

Das Kommunizieren ist eine Äquivalenzrelation, und die Äquivalenzklassen bezüglich dieser Relation werden wir einfach nur "Klassen" nennen; wenn es

Anlass zu Verwechslungen gibt, dann sprechen wir genauer von “Rekurrenzklassen”.

Eine Markovkette, in der alle Zustände miteinander kommunizieren, nennt man *irreduzibel*.

Es gibt nun eine Reihe von Eigenschaften, die entweder für alle Zustände in einer Klasse gleichzeitig gelten, oder für keinen. Solche Eigenschaften nennen wir *Klasseneigenschaften*.

Die erste dieser Eigenschaften ist die *Periode*

$$d(i) = \text{ggT} \{t \geq 0 : p_{ii}(t) > 0\}.$$

Ist die Periode einer Klasse = 1, dann nennen wir diese Klasse aperiodisch.

Um zu beweisen, dass die Periode eine Klasseneigenschaft ist, betrachten wir zwei verbundene Zustände i und j mit den Perioden $d(i)$ und $d(j)$. Wir können nun Zeiten $s, t \geq 0$ finden, sodass $p_{ij}(s)$ und $p_{ji}(t)$ positiv sind. Ist nun u eine Zahl mit $p_{jj}(u) > 0$, dann gilt wegen der Chapman-Kolmogorov Gleichungen

$$p_{ii}(s+t) \geq p_{ij}(s)p_{ji}(t) > 0$$

und

$$p_{ii}(s+u+t) \geq p_{ij}(s)p_{jj}(u)p_{ji}(t) > 0.$$

Nach der Definition der Periode sind sowohl $s+t$ als auch $s+t+u$ durch $d(i)$ teilbar, also auch u . Weil das für alle u gilt, ist $d(i)$ ein gemeinsamer Teiler aller u , für die $p_{jj}(u) > 0$. Daraus folgt $d(i)|d(j)$, und weil genauso $d(j)|d(i)$ gefolgert werden kann, gilt schließlich $d(i) = d(j)$.

Weitere Klasseneigenschaften ergeben sich aus dem Rückkehrverhalten von ξ . Wir definieren

$$\tau_i = \inf\{t > 0 : \xi(t) = i\},$$

die Übergangszeit bzw. (wenn aus i gestartet wird) Rückkehrzeit (Rekurrenzzeit),

$$\nu_i = |\{t > 0 : \xi(t) = i\}|,$$

die Anzahl der Besuche in i ,

$$f_{ij}(t) = \mathbb{P}_i(\xi(t) = j, \xi(s) \neq j, 1 \leq s < t) = \mathbb{P}_i(\tau_j = t),$$

und

$$f_{ij} = \sum_{t=1}^{\infty} f_{ij}(t) = \mathbb{P}_i(\tau_j < \infty).$$

Es gilt:

Satz 3.1 *Die folgenden Bedingungen sind äquivalent:*

1. $\mathbb{P}_i(\tau_i < \infty) = 1$,
2. $\mathbb{P}_i(\nu_i = \infty) = 1$,
3. $\mathbb{E}_i(\nu_i) = \infty$,

$$4. \sum_t p_{ii}(t) = \infty.$$

Die Zufallsvariable τ_i ist die Zeit bis zum ersten Besuch (bzw. zur ersten Rückkehr) in i , ν_i die Anzahl der Besuch in i . Bezeichnet man mit $\tau_i^{(k)}$ die Zeit bis zum k -ten Besuch in i , dann hat $\tau_i^{(k+1)} - \tau_i^{(k)}$ (wenn $\tau_i^{(k)}$ endlich ist), dieselbe Verteilung wie τ_i unter der Bedingung, dass in i gestartet wird. Ist nun τ_i mit Wahrscheinlichkeit eins endlich, dann auch $\tau_i^{(k)}$. Mit Wahrscheinlichkeit eins ist also $\nu_i \geq k$, und somit $\nu_i = \infty$, und auch $\mathbb{E}_i(\nu_i) = \infty$. Ist andererseits τ_i mit positiver Wahrscheinlichkeit p unendlich, dann ist

$$\mathbb{P}_i(\nu_i = 0) = \mathbb{P}_i(\tau_i = \infty) = p,$$

$$\mathbb{P}_i(\nu_i = 1) = \mathbb{P}_i(\tau_i < \infty, \tau_i^{(2)} = \infty) = p(1 - p),$$

$$\mathbb{P}_i(\nu_i = 2) = p(1 - p)^2, \dots$$

ν_i hat also eine geometrische Verteilung, ist daher mit Wahrscheinlichkeit eins endlich und hat eine endliche Erwartung. Für den letzten Punkt stellen wir fest, dass

$$\mathbb{E}_i(\nu_i) = \mathbb{E}_i(\sum_{t=1}^{\infty} I_{\{i\}}(\xi(t))) = \sum_{t=1}^{\infty} \mathbb{P}_i(\xi(t) = i) = \sum_{t=1}^{\infty} p_{ii}(t).$$

Definition 3.1 Erfüllt ein Zustand i die Bedingungen des vorigen Satzes, so heißt er rekurrent, andernfalls heißt er transient.

Satz 3.2 Rekurrenz und Transienz sind Klasseneigenschaften.

Es seien wieder i und j zwei verbundene Zustände, und u und v seien so gewählt, dass $p_{ij}(u)$ und $p_{ji}(v)$ positiv sind. Ist nun i rekurrent, dann ist

$$\sum_{t>0} p_{jj}(t) \geq \sum_{t>u+v} p_{jj}(t) \geq \sum_{t>u+v} p_{ji}(v)p_{ii}(t-u-v)p_{ij}(u) = \infty.$$

j ist also ebenfalls rekurrent. Umgekehrt folgt aus der Rekurrenz von j auch die von i . Die Rekurrenz von i und j sind also äquivalent, also ist die Rekurrenz eine Klasseneigenschaft.

Bei der Rekurrenz lässt sich noch weiter unterscheiden:

Definition 3.2 Ein Zustand i heißt positiv rekurrent, wenn

$$\mathbb{E}_i(\tau_i) < \infty$$

und nullrekurrent, wenn dieser Erwartungswert unendlich ist.

Satz 3.3 Positive Rekurrenz und Nullrekurrenz sind Klasseneigenschaften.

Satz 3.4 Für jede irreduzible aperiodische Markovkette gilt

$$\pi_j = \lim_{t \rightarrow \infty} p_{ij}(t) = \frac{1}{\mathbb{E}_i(\tau_i)}.$$

Falls die π_i nicht alle gleich null sind (d.h., wenn die Kette positiv rekurrent ist), dann ist π_j eine stationäre Verteilung, d.h., es gilt

$$\sum_i \pi_i p_{ij} = \pi_j$$

und

$$\sum_i \pi_i = 1.$$

Gibt es umgekehrt eine solche stationäre Verteilung, dann ist die Kette positiv rekurrent.

Im periodischen (und irreduziblen) Fall bleibt der Satz richtig, wenn man

$$\pi_j = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t p_{ij}(s)$$

setzt.

Wenn die Kette nicht irreduzibel ist, dann stellt sich das Resultat so dar:

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{s=1}^t p_{ij}(s) = f_{ij} \pi_j.$$

Für den Beweis zeigen wir zuerst die Erneuerungsgleichung

Satz 3.5 für jedes $t > 0$ gilt:

$$p_{ij}(t) = \sum_{s=1}^t f_{ij}(s) p_{jj}(t-s).$$

Diese ergibt sich einfach durch

$$p_{ij}(t) = \mathbb{P}_i(\xi(t) = j) = \sum_{s=1}^t \mathbb{P}_i(\xi(t) = j, \tau_j = s).$$

Im transienten Fall ist die Behauptung von Satz 3.4 klar, weil rechts trivialerweise 0 steht, und auf der linken Seite

$$\limsup_{t \rightarrow \infty} \mathbb{P}_i(\xi(t) = j) \leq \mathbb{P}_i(\limsup_{t \rightarrow \infty} [\xi(t) = j]) = \mathbb{P}_i(\nu_j = \infty) = 0.$$

Der Fall $i \neq j$ lässt sich leicht erledigen, wenn $i = j$ bewiesen ist, weil dann in der Erneuerungsgleichung alle $p_{jj}(t-s)$ gegen π_j konvergieren, und eine

Anwendung des Satzes von der dominierten Konvergenz zeigt, dass dann auch $p_{ij}(t)$ gegen π_j konvergiert.

Im Fall $i = j$ nehmen wir der Einfachheit halber an, dass $f_{ii}(1) > 0$ gilt. Mit etwas an zusätzlicher Arbeit funktioniert der Beweis auch für den Fall, dass der größte gemeinsame Teiler aller t , für die $f(t) > 0$ gilt, gleich eins ist, also für den Fall einer aperiodischen Kette.

Wir werden im Beweis die Indizes i unterdrücken, um die Notation einfacher zu machen. Wir schreiben also $f(t)$, $p(t)$, etc.

Wir setzen

$$\pi = \limsup_{t \rightarrow \infty} p(t).$$

Es existiert eine Folge $t_n \rightarrow \infty$, sodass

$$\pi = \lim_{n \rightarrow \infty} p(t_n).$$

Aus der Erneuerungsgleichung erhalten wir

$$f(1)p(t_n - 1) = p(t_n) - \sum_{k=2}^{t_n} f(k)p(t_n - k).$$

In der Summe auf der rechten Seite kann die Obergrenze gleich unendlich gesetzt werden, wenn man $p(t)$ für negatives t gleich null setzt.

Aus dem Lemma von Fatou folgt

$$\limsup_{n \rightarrow \infty} \sum_{k=2}^{\infty} f(k)p(t_n - k) \leq \sum_{k=2}^{\infty} f(k) \limsup_{n \rightarrow \infty} p(t_n - k) \leq \sum_{k=2}^{\infty} \pi f(k) = \pi(1 - f(1)).$$

Daraus ergibt sich

$$f(1) \liminf_{n \rightarrow \infty} p(t_n - 1) \geq \pi f(1),$$

also

$$\liminf_{n \rightarrow \infty} p(t_n - 1) \geq \pi,$$

und (weil natürlich die umgekehrte Ungleichung für den Limes superior gilt),

$$\lim_{n \rightarrow \infty} p(t_n - 1) = \pi.$$

Daraus folgt wiederum, dass für jedes $s > 0$

$$\lim_{n \rightarrow \infty} p(t_n - s) = \pi.$$

Jetzt setzen wir

$$g(t) = \sum_{s \geq t} f(s).$$

Damit lautet die Erneuerungsgleichung

$$p(t) = \sum_{s \geq 1} (g(s) - g(s+1))p(t-s),$$

umgeformt

$$p(t) + \sum_{s \geq 1} g(s+1)p(t-s) = \sum_{s \geq 1} g(s)p(t-s),$$

bzw.

$$p(t) + \sum_{s \geq 2} g(s)p(t+1-s) = \sum_{s \geq 1} g(s)p(t-s),$$

und weil $g(1) = 1$,

$$\sum_{s \geq 1} g(s)p(t+1-s) = \sum_{s \geq 1} g(s)p(t-s) = \dots = \sum_{s \geq 1} g(s)p(1-s) = 1.$$

Eine weitere Anwendung des Lemmas von Fatou ergibt

$$\pi \sum_{s \geq 1} g(s) \leq \liminf_n \sum_{s \geq 1} g(s)p(t-n-s) = 1.$$

Also ist

$$\limsup_{t \rightarrow \infty} p(t) = \pi = \frac{1}{\sum_{s \geq 1} g(s)} = \frac{1}{\sum_{s \geq 1} sf(s)}.$$

Der letzte Nenner ist genau $\mathbb{E}_i(\tau_i)$. Wenn dieser Erwartungswert unendlich ist, dann ist $\pi = 0$ und der Limes superior ein Limes. Ist der Erwartungswert endlich, dann kann man dasselbe Argument für den Limes inferior durchführen, was die Abschätzung

$$\liminf_{t \rightarrow \infty} p_{ii}(t) \geq \frac{1}{\mathbb{E}_i(\tau_i)}$$

liefert. Beides zusammen beweist den ersten Teil des Satzes.

Es bleibt noch zu zeigen, dass π_i eine stationäre Verteilung liefert. Aus

$$\sum_j p_{ij}(t) = 1$$

folgt (wieder aus dem Lemma von Fatou)

$$\sum_j \pi_j \leq 1.$$

und ebenso aus

$$p_{ij}(t+1) = \sum_k p_{ik}(t)p_{kj}$$

die Ungleichung

$$\pi_j \geq \sum_k \pi_k p_{kj},$$

und somit

$$\sum_j \pi_j \geq \sum_j \sum_k \pi_k p_{kj} = \sum_k \pi_k,$$

also sind die Ungleichungen in der Mitte Gleichungen, und also

$$\pi_j = \sum_k \pi_k p_{kj}.$$

Durch iterieren ergibt sich daraus

$$\pi_j = \sum_k \pi_k p_{kj}(t),$$

was einerseits zur Folge hat, dass jeder Nachfolger eines positiv rekurrenten Zustands wieder positiv rekurrent ist und also die positive Rekurrenz eine Klasseeigenschaft ist, und im irreduziblen Fall für $t \rightarrow \infty$ mit dem Satz von der dominierten Konvergenz

$$\pi_j = \pi_j \sum_k \pi_k,$$

also ist im positiv rekurrenten Fall auch die Summe der Grenzwerte π_j gleich eins, und der Satz ist bewiesen.

Zusammengefasst können wir sagen, dass eine irreduzible Kette genau dann rekurrent ist, wenn es eine stationäre Verteilung gibt.

3.3 Markovketten in stetiger Zeit

3.3.1 Grundsätzliches

Jetzt wollen wir uns dem Fall zuwenden, dass der Zeitparameter beliebige positive Werte annehmen kann. Einige Dinge können wir direkt aus dem vorigen Abschnitt übernehmen, etwa die Definition der Übergangsmatrizen $P(t)$ und die Chapman-Kolmogorov Gleichungen

$$P(s+t) = P(s)P(t).$$

Zusätzlich verlangen wir die Stetigkeit der Übergangswahrscheinlichkeiten bei 0:

$$\lim_{t \rightarrow 0} P(t) = P(0) = I.$$

Mit dieser Bedingung schließen wir etwa Prozesse aus, für die alle Werte $\xi(t)$ unabhängig sind.

Andere Eigenschaften können wir übernehmen, indem wir feststellen, dass für jedes $a > 0$ die Folge $\eta(n) = \xi(an)$ eine Markovkette in diskreter Zeit bildet. Damit lassen sich die Begriffe des Kommunizierens, der Rekurrenz, Transienz und positiven Rekurrenz übertragen. Periodizität kann im stetigen Fall nicht auftreten.

Leider gibt es im stetigen Fall nicht so wie im diskreten Fall ein ausgezeichnetes Element wie die 1, aus der wir alle anderen Zeiten erhalten können. Stattdessen verwenden wir ein wenig Differentialrechnung. Es folgt nämlich schon aus der Stetigkeit bei 0, die wir gerade gefordert haben:

Satz 3.6 Wenn für alle i und j

$$\lim_{t \rightarrow 0} p_{ij}(t) = p_{ij}(0)$$

gilt, dann existieren

$$q_{ij} = \lim_{t \rightarrow 0} \frac{p_{ij}(t) - p_{ij}(0)}{t},$$

und es gilt $0 \leq q_{ij} < \infty$ für alle $i \neq j$ und $-\infty \leq q_{ii} \leq 0$.

Den Beweis führen wir zuerst für den Fall $i = j$. Wir wählen ein $t > 0$. Für $h < t$ setzen wir $n = \lfloor t/h \rfloor$ und $d = t - nh$. Aus den Gleichungen von Chapman-Kolmogorov folgt

$$p_{ii}(t) = p_{ii}(nh + d) \geq p_{ii}(h)^n p_{ii}(d) \geq p_{ii}(h)^{t/h} p_{ii}(d)$$

und damit

$$\frac{\log(p_{ii}(t))}{t} \geq \frac{\log(p_{ii}(h))}{h} + \frac{\inf_{d \leq h} \log(p_{ii}(d))}{t}. \quad (3.1)$$

Jetzt sei

$$q_{ii} = \limsup_{t \rightarrow 0} \frac{\log(p_{ii}(t))}{t}.$$

Wir wählen eine Folge h_k , für die

$$q_{ii} = \lim_{k \rightarrow \infty} \frac{\log(p_{ii}(h_k))}{h_k}.$$

Wenn wir diese Folge in (3.1) einsetzen, dann konvergiert der zweite Summand auf der rechten Seite wegen der Stetigkeit von $q_{ii}(t)$ gegen null, und wir erhalten

$$\frac{\log(p_{ii}(t))}{t} \geq q_{ii}.$$

Das wiederum impliziert

$$\liminf_{t \rightarrow 0} \frac{\log(p_{ii}(t))}{t} \geq q_{ii},$$

also stimmen Limes inferior und Limes superior überein, und wir erhalten die behauptete Konvergenz.

Im Fall $i \neq j$ können wir zunächst zu jedem $\epsilon > 0$ ein t_0 wählen, sodass für alle $t < t_0$ sowohl $p_{ii}(t)$ als auch $p_{jj}(t)$ mindestens gleich $1 - \epsilon$ ist. Dann gilt auch $p_{ij}(t) \leq \epsilon$ und $p_{ji}(t) \leq \epsilon$.

Wir setzen wieder für $h < t$ wie oben $t = nh + d$ und setzen

$$p_{ij}(t) = \sum_{1 \leq k \leq n} \mathbb{P}_i(\xi((k-1)h) = i, \xi(kh) = \xi(t) = j, \xi(lh) \neq i, k < l \leq n) \geq$$

$$\sum_{1 \leq k \leq n} \mathbb{P}_i(\xi((k-1)h) = i) \mathbb{P}_i(\xi(h) = j) \mathbb{P}_j(\xi(t-kh) = j, \xi(lh) \neq i, 0 < l \leq n-k) \geq$$

$$\sum_{1 \leq k \leq n} p_{ii}((k-1)h) p_{ij}(h) (p_{jj}(t-kh) - \mathbb{P}_j(\exists l \neq n-k : \xi(lh) = i)).$$

Die letzte Wahrscheinlichkeit können wir wie folgt darstellen:

$$\mathbb{P}_j(\exists l \neq n-k : \xi(lh) = i) = \sum_{l \leq n-k} \mathbb{P}_j(\xi(lh) = i, \xi(rh) \neq i, r < l).$$

Andererseits ist

$$\begin{aligned} \epsilon &\geq \mathbb{P}_j(\xi(nh) = i) \geq \sum_{l \leq n-k} \mathbb{P}_j(\xi(lh) = i, \xi(rh) \neq i, r < l) p_{ii}((n-l)h) \\ &\geq (1-\epsilon) \sum_{l \leq n-k} \mathbb{P}_j(\xi(lh) = i, \xi(rh) \neq i, r < l). \end{aligned}$$

Also

$$\mathbb{P}_j(\exists l \neq n-k : \xi(lh) = i) \leq \frac{\epsilon}{1-\epsilon}$$

und damit

$$p_{ij}(t) \geq (1-\epsilon) n p_{ij}(h) (1-\epsilon - \frac{\epsilon}{1-\epsilon}) \geq (1-3\epsilon) n p_{ij}(h).$$

Für $\epsilon < 1/3$ ergibt sich

$$p_{ij}(h) \leq \frac{1}{(1-3\epsilon) \lfloor t/h \rfloor} p_{ij}(t).$$

Für $h \rightarrow 0$ folgt daraus

$$\limsup \frac{p_{ij}(h)}{h} \leq \frac{1}{1-3\epsilon} \frac{p_{ij}(t)}{t}.$$

Wir lassen nun t eine Folge t_k durchlaufen, für die

$$q_{ij} = \liminf_{t \rightarrow 0} \frac{p_{ij}(t)}{t} = \lim_{k \rightarrow \infty} \frac{p_{ij}(t_k)}{t_k}.$$

Weil t_k irgendwann unter jedes $t_0(\epsilon)$ fällt, gilt für jedes $\epsilon > 0$

$$\limsup \frac{p_{ij}(h)}{h} \leq \frac{1}{1-3\epsilon} q_{ij},$$

also auch

$$\limsup \frac{p_{ij}(h)}{h} \leq q_{ij},$$

und wieder stimmen Limes inferior und Limes superior überein.

Die Zahlen q_{ij} heißen die infinitesimalen Parameter der Kette; man kann sie wieder zu einer Matrix Q zusammenfassen, die wir den infinitesimalen Erzeuger der Kette nennen.

Wenn der Zustandsraum endlich ist, dann kann man zusätzlich finden, dass auch die Diagonalelemente q_{ii} endlich sind, und dass $\sum_j q_{ij} = 0$ gilt. Für unendliche Zustandsräume gilt diese Beziehung nicht mehr unbedingt. Ketten, die diese zusätzlichen Eigenschaften erfüllen, nennen wir *konservativ*.

Wir hoffen natürlich, dass wir durch die Angabe des infinitesimalen Erzeugers die zugehörige Markovkette festlegen können; dazu müssen wir daraus nur die Übergangswahrscheinlichkeiten ausrechnen können (dann kann man an den Existenzsatz von Kolmogorov appellieren).

Betrachten wir zunächst die Wahrscheinlichkeiten

$$\mathbb{P}_i(\xi(\frac{k}{n}) = i, 0 \leq k \leq nT) = p_{ii}(\frac{1}{n})^{\lfloor nT \rfloor},$$

und

$$\mathbb{P}_i(\xi(\frac{k}{n}) = j | \xi(\frac{k}{n}) \neq i, \xi(\frac{l}{n}) = i, 0 < l < k) = \frac{p_{ij}(1/n)}{1 - p_{ii}(1/n)}.$$

Für $n \rightarrow \infty$ konvergiert die erste Wahrscheinlichkeit gegen $e^{q_{ii}T}$, die zweite gegen $\frac{q_{ij}}{-q_{ii}}$. Daraus kommt man zu der naiven Interpretation (die man auch streng machen kann, wie wir in Kürze sehen werden): ist der (konservative) Prozess im Zustand i mit $q_{ii} \neq 0$, dann bleibt er dort noch eine Zeit, die eine Exponentialverteilung mit Parameter $-q_{ii}$ besitzt, dann springt er mit Wahrscheinlichkeit

$$r_{ij} = -\frac{q_{ij}}{q_{ii}}$$

in den Punkt $j \neq i$ (ein Zustand mit $q_{ii} = 0$ ist natürlich absorbierend).

Die Matrix $R = (r_{ij})$ (mit $r_{ij} = 0$) ist wieder die Übergangsmatrix einer diskreten Markovkette, die wir die eingebettete Markovkette nennen (im Fall $q_{ii} = 0$ setzen wir $r_{ii} = 1$, damit R wirklich eine Übergangsmatrix ist; eigentlich fällt das aber nicht ins Gewicht, weil die entsprechenden Sprünge nicht auftreten).

Wir können uns jetzt überlegen, was bei nicht konservativen Ketten passiert: der Fall $q_{ii} = -\infty$ bedeutet, dass die Zeit, die der Prozess in i verbringt, fast sicher gleich null ist. Ein solcher Zustand, den man beim Betreten sofort wieder verlässt, heißt *instabiler Zustand*. Wenn die Zeilensumme von Q nicht null ist, dann ist die Zeilensumme der Matrix R kleiner als eins, und das bedeutet, dass der Prozess mit positiver Wahrscheinlichkeit im Nichts (der Unendlichkeit, Nirvana, ...) verschwindet — von dort muss er natürlich sofort wieder zurückkehren, damit die endlichdimensionalen Verteilungen brav bleiben, aber wie das konkret passiert, lässt einigen Spielraum, ist jedenfalls durch den infinitesimalen Erzeuger nicht festgelegt.

Durch formalen Grenzübergang erhält man aus den Chapman-Kolmogorov Gleichungen die Kolmogorovschen Differentialgleichungen:

Die Rückwärtsgleichung

$$\dot{P}(t) = QP(t)$$

und die Vorwärtsgleichung

$$\dot{P}(t) = P(t)Q.$$

Da diese Ableitung nur formal ist, erhebt sich die Frage nach ihrer Gültigkeit. Es zeigt sich, dass die Rückwärtsgleichung freundlicher ist, die Bedingungen für die Gültigkeit der Vorwärtsgleichung sind restriktiver. Es gilt:

Satz 3.7 *Wenn ξ konservativ ist, dann gilt die Rückwärtsgleichung.*

Wir setzen

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \frac{1 - p_{ii}(h)}{h} \left(-p_{ij}(t) + \sum_{k \neq i} \frac{p_{ik}(h)}{1 - p_{ii}(h)} p_{kj}(t) \right).$$

Der erste Faktor auf der rechten Seite konvergiert gegen $-q_{ii}$, die Brüche in der Summe sind nichtnegativ und haben Summe 1, bilden also eine Wahrscheinlichkeitsverteilung, und konvergieren gegen die Wahrscheinlichkeitsverteilung r_{ik} . Diese punktweise Konvergenz ist bekanntlich äquivalent zur schwachen Konvergenz und damit zur Konvergenz der Integrale von beschränkten Funktionen, die hier eine Summe sind. Also ergibt sich für $h \rightarrow 0$

$$p'_{ij}(t) = -q_{ii}(-p_{ij}(t) + \sum_{k \neq i} r_{ik} p_{kj}(t)) = \sum_k q_{ik} p_{kj}(t).$$

Satz 3.8 *Wenn die q_{ij} beschränkt sind, dann gilt die Vorwärtsgleichung.*

Aus dem Beweis der Existenz von q_{ii} folgt

$$p_{ii}(t) \geq e^{q_{ii}t} \geq 1 + q_{ii}(t)t.$$

Daher ist

$$\left| \frac{p_{ij}(t) - p_{ij}(0)}{t} \right| \leq \frac{1 - p_{ii}(t)}{t}$$

beschränkt, und man kann auf

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \sum_k p_{ik} \frac{p_{kj}(h) - p_{kj}(0)}{h}$$

den Satz von der dominierten Konvergenz anwenden.

3.3.2 Geburts- und Todesprozesse

Die einfachsten Beispiele für Markovketten in stetiger Zeit sind diejenigen, die nur Übergänge zu den Nachbarzuständen zulassen. Das bedeutet, dass der infinitesimale Erzeuger die folgende Form hat:

$$q_{ii} = -(\lambda_i + \mu_i),$$

$$q_{i,i+1} = \lambda_i,$$

$$q_{i,i-1} = \mu_i.$$

Um den Zustand $i = 0$ nicht gesondert behandeln zu müssen, setzen wir $\mu_0 = 0$.

Die Übergänge von einem Zustand in den nächsthöheren nennen wir die “Geburten”, die umgekehrten Übergänge die “Tode”, λ_i heißt die Geburtsrate, μ_i die Sterberate (im Zustand i).

Wenn alle μ_i gleich null sind, dann sprechen wir von einem reinen Geburtsprozess, sind alle λ_i gleich null, von einem reinen Todesprozess.

Für reine Geburtsprozesse mit $\xi(0) = 0$ ist die Übergangszeit τ_n bis zum Zustand n die Summe von n unabhängigen Zufallsvariablen mit Parametern $\lambda_0, \dots, \lambda_{n-1}$. τ_n hat daher den Erwartungswert

$$\frac{1}{\lambda_0} + \dots + \frac{1}{\lambda_{n-1}}.$$

Wenn die Reihe

$$\sum_n \frac{1}{\lambda_n}$$

konvergiert, dann hat

$$\tau = \sup_n \tau_n$$

endlichen Erwartungswert und ist daher selbst endlich. Das bedeutet aber, dass zu dem endlichen Zeitpunkt τ der Prozess schon alle natürlichen Zahlen durchlaufen hat und wieder einmal im (bitte persönliche Präferenz einsetzen) verschwunden ist. Wir sprechen auch von der “Explosion” des Prozesses. In unserem Fall können wir nach der Explosion fortsetzen, indem einfach in irgendeinen Zustand gesprungen wird. Es ist also durch den infinitesimalen Erzeuger der Prozess nur bis zur ersten Explosion eindeutig festgelegt. Diese Überlegungen gelten sinngemäß für alle Markovketten. Am angenehmsten sind natürlich jene Prozesse, die nicht explodieren, für die also die Explosionszeit mit Wahrscheinlichkeit eins unendlich ist. Solche Prozesse heißen regulär:

Definition 3.3 *Eine Markovkette $\xi(t)$ in stetiger Zeit heißt regulär, wenn sie in endlicher Zeit mit Wahrscheinlichkeit 1 nur endlich viele Sprünge macht.*

Die Zeit bis zur Explosion τ ist die Summe der Zeiten, die in den einzelnen Zuständen verbracht werden, die die Kette besucht. Bezeichnet man die eingebettete Markovkette mit η , so kann man feststellen, dass

$$\tau = \sum_i \tau_i,$$

dabei ist τ_i exponentialverteilt mit Parameter $\lambda_{\eta(i)}$, wobei wir zur Vereinfachung $\lambda(k) = -q_{kk}$ gesetzt haben. Daraus ergibt sich

Satz 3.9 *ξ ist genau dann regulär, wenn fast sicher*

$$\sum 1/\lambda_{\eta(i)} = \infty.$$

Insbesondere ist ξ regulär, wenn die eingebettete Kette rekurrent ist.

Mit dem infinitesimalen Erzeuger allein funktioniert das folgende Kriterium:

Satz 3.10 ξ ist genau dann regulär, wenn für ein $z > 0$ (und dann zwingend für alle) $x = 0$ die einzige nichtnegative und beschränkte Lösung der Gleichung

$$Qx = zx$$

ist.

Mit der Notation von oben setzen wir

$$\sigma_k = \sum_{i=1}^k \tau_i$$

und setzen

$$x_i^{(k)} = \mathbb{E}_i(e^{-z\sigma_k}).$$

Es ergibt sich die Gleichung

$$x_i^{(k+1)} = \frac{-q_{ii}}{-q_{ii} + z} \sum_{j \neq i} r_{ij} x_j^{(k)}.$$

Für $k \rightarrow \infty$ ergibt sich

$$x_i = \lim_{k \rightarrow \infty} x_i^{(k)} = \mathbb{E}_i(e^{-z\tau})$$

und

$$x_i = \frac{-q_{ii}}{-q_{ii} + z} \sum_{j \neq i} r_{ij} x_j.$$

x erfüllt also die Gleichung $Qx = zx$.

Wenn ξ singulär ist, dann ist τ mit positiver Wahrscheinlichkeit endlich, und daher ist mindestens ein x_i positiv. Wenn umgekehrt y nichtnegativ, von 0 verschieden und beschränkt ist und die Gleichung $Qy = zy$ erfüllt, dann können wir ohne Beschränkung der Allgemeinheit $y \leq 1$ annehmen, also

$$y \leq x^{(0)}$$

und durch Induktion

$$y \leq x^{(n)}$$

und für $n \rightarrow \infty$

$$y \leq x.$$

Damit ist auch x nicht identisch 0 und daher τ mit positiver Wahrscheinlichkeit endlich.

Geburts- und Todesprozesse sind jedenfalls regulär, wenn $\sum 1/\lambda_i = \infty$. Etwas mehr Arbeit liefert das Ergebnis:

Satz 3.11 *Ein Geburts- und Todesprozess ist genau dann regulär, wenn*

$$\sum_{i=0}^{\infty} \sum_{j=0}^i \frac{1}{\lambda_j} \prod_{k=j+1}^i \frac{\mu_k}{\lambda_k} = \infty.$$

Wenn alle λ_i und μ_i positiv sind, dann gilt:

Satz 3.12 *Ein regulärer Geburts- und Todesprozess mit positiven Geburts- und Sterberaten ist genau dann rekurrent, wenn*

$$\sum_n \prod_{i=1}^n \frac{\mu_i}{\lambda_i} = \infty,$$

und positiv rekurrent, wenn

$$\sum_n \prod_{i=0}^n \frac{\lambda_i}{\mu_{i+1}} < \infty.$$

3.3.3 Die Minimallösung

Die Konstruktion einer stetigen Markovkette über die eingebettete Markovkette führt zu der sogenannten Minimallösung einer Kette. Dazu führen wir zusätzlich die Zufallsvariable $\nu(t)$ ein, die gleich der Anzahl der Sprünge bis zum Zeitpunkt t ist.

Die Minimallösung ist dann

$$p_{ij}^{(0)}(t) = \mathbb{P}_i(\xi(t) = j, \nu(t) < \infty)$$

und kann als

$$\lim_{n \rightarrow \infty} p_{ij}^{(n)}(t)$$

mit

$$p_{ij}^{(n)}(t) = \mathbb{P}_i(\xi(t) = j, \nu(t) < n)$$

erhalten werden.

Das führt zu

$$p_{ij}^{(1)} = e^{q_{ii}t} \delta_{ij} \tag{3.2}$$

und

$$p_{ij}^{(n+1)}(t) = p_{ij}^{(1)}(t) + \int_0^t \sum_{k \neq i} q_{ik} e^{q_{ii}s} p_{kj}^{(n)}(t-s) ds. \tag{3.3}$$

Um diese Überlegungen streng zu machen, verwenden wir (3.2) und (3.3) als Definition und nennen

$$p_{ij}^{(0)}(t) = \lim_{n \rightarrow \infty} p_{ij}^{(n)}(t)$$

die Minimallösung. Der Grenzwert existiert wegen $p_{ij}^{(n+1)} \geq p_{ij}^{(n)}$.

Für die Minimallösung gilt

Satz 3.13 Die Minimallösung $p_{ij}^{(0)}(t)$ erfüllt die Vorwärts- und die Rückwärts-
gleichung, und für jede Lösung $p_{ij}(t)$ der Rückwärts-
gleichung mit $p_{ij}(0) = \delta_{ij}$ gilt

$$p_{ij}(t) \geq p_{ij}^{(0)}(t).$$

Die Rückwärts-
gleichung ist äquivalent zu

$$p_{ij}(t) = p_{ij}^{(1)}(t) + \int_0^t \sum_{k \neq i} q_{ik} e^{q_{ii}s} p_{kj}(t-s) ds.$$

Das ist genau, was sich aus (3.3) durch den Grenzübergang $n \rightarrow \infty$ ergibt.
Außerdem folgt daraus für jede nichtnegative Lösung der Rückwärts-
gleichung die Ungleichung

$$p_{ij}(t) \geq p_{ij}^{(1)}(t)$$

und durch Induktion

$$p_{ij}(t) \geq p_{ij}^{(n)}(t),$$

und für $n \rightarrow \infty$

$$p_{ij}(t) \geq p_{ij}^{(0)}(t).$$

Ebenfalls durch Induktion zeigt man

$$p_{ij}^{n+1}(t) = p_{ij}^{(1)}(t) + \int_0^t \sum_{k \neq j} q_{kj} e^{q_{jj}s} p_{ik}^{(n)}(t-s) ds,$$

woraus sich für $n \rightarrow \infty$ die Vorwärts-
gleichung ergibt.

Die Kette ξ ist natürlich genau dann regulär, wenn die Minimallösung eine
Wahrscheinlichkeitsverteilung bildet (d.h. $\sum_j p_{ij}^{(0)}(t) = 1$), und dann ist die Mi-
nimallösung die einzige mögliche Übergangsfunktion zu diesem infinitesimalen
Erzeuger.

3.4 Allgemeine Theorie der Markovprozesse

3.4.1 Markovprozesse und Übergangsfunktionen

Wir betrachten nun Markovprozesse, für die sowohl der Parameterraum als auch
der Zustandsraum stetig sind, d.h., wir wählen $T = [0, \infty)$ und X als eine
Teilmenge von \mathbb{R}^d . In den meisten Fällen werden wir uns auf den Fall $X = \mathbb{R}$
beschränken, hauptsächlich deswegen, weil dann die Beweise einfacher zu
führen sind, aber die Ergebnisse gelten auch für den allgemeinen Fall. Man kann
auch noch allgemeinere Zustandsräume betrachten (etwa separable metrische
Räume), aber bei zu allgemeinen Zustandsräumen kann es Probleme mit einigen
Existenzaussagen geben.

Der Ausgangspunkt unserer Untersuchungen ist die Markoveigenschaft

$$\mathbb{P}(\xi(t) \in A | \mathcal{F}_s) = \mathbb{P}(\xi(t) \in A | \xi(s))$$

für alle $s \leq t$ und alle Borelmengen A .

Das gibt uns Anlass dazu, die folgende Funktion zu betrachten:

$$P(s, t, x, A) = \mathbb{P}(\xi(t) \in A | \xi(s) = x).$$

Diese Funktion nennen wir die Übergangsfunktion des Markovprozesses. Sie soll die folgenden Eigenschaften erfüllen:

Definition 3.4 Die Funktion P heißt Übergangsfunktion, wenn

1. $P(s, t, x, A)$ ist messbar (als Funktion von x).
2. $P(s, t, x, A)$ ist ein Wahrscheinlichkeitsmaß (als Funktion von A),
3. $P(s, s, x, A) = I_A(x)$,
4. für $s \leq t \leq u$ gilt

$$P(s, u, x, A) = \int P(s, t, x, dy)P(t, u, y, A).$$

Hier tauchen naturgemäß zwei Fragen auf: einerseits die, ob es zu jedem Markovprozess eine Übergangsfunktion gibt, andererseits die, ob man zu einer gegebenen Übergangsfunktion einen Markovprozess mit dieser Übergangsfunktion finden kann.

Die Antwort auf die zweite Frage ist leichter zu geben. Zu jedem Anfangswert x und für $t_1 < t_2 < \dots < t_n$ und Borelmengen A_1, \dots, A_n kann man durch

$$\mathbb{P}_x(\xi(t_1) \in A_1, \dots, \xi(t_n) \in A_n) = \int_{A_n} \dots \int_{A_1} P(0, t_1, x, dx_1)P(t_1, t_2, x_1, dx_2) \dots P(t_{n-1}, t_n, x_{n-1}, dx_n)$$

eine Familie von Wahrscheinlichkeitsverteilungen definieren, die wegen Eigenschaft 4 der Übergangsfunktion konsistent ist, und der Existenzsatz von Kolmogorov stellt sicher, dass es einen Prozess mit diesen endlichdimensionalen Verteilungen gibt. Es gibt also nicht nur einen solchen Markovprozess, sondern sogar zu jedem Anfangswert $\xi(0) = x$ einen eigenen Markovprozess $\xi^{(x)}(t)$. Wir sprechen daher von der Familie von Markovprozessen (oder kurz von der Markov-Familie) $\xi^{(x)}$ und stellen fest, dass durch eine Übergangsfunktion die zugehörige Markov-Familie eindeutig festgelegt wird.

Die zweite Frage führt uns in die Abgründe der Maßtheorie. Man kann nämlich die bedingte Wahrscheinlichkeit $P(s, t, x, A)$ für fixes A als Radon-Nikodym-Dichte erhalten. Damit ist die Messbarkeit als Funktion von x automatisch gewährleistet. Nicht ganz so einfach ist es mit der zweiten Eigenschaft: für eine fixe Folge von disjunkten Mengen gilt zwar mit Wahrscheinlichkeit eins, dass die bedingte Wahrscheinlichkeit für die Vereinigung gleich der Summe der einzelnen bedingten Wahrscheinlichkeiten ist, aber weil es überabzählbar viele solche Folgen gibt, kann man nicht unbedingt folgern, dass die Sigmaadditivität

mit Wahrscheinlichkeit eins gleichzeitig für alle solchen Folgen gilt. Man kann in der Tat (ein berühmtes Beispiel von Dirichlet) Räume angeben, auf denen man keine bedingten Wahrscheinlichkeiten angeben kann, die Übergangsfunktionen sind. In unserem Fall (für reelle Zustandsräume) gibt es aber immer eine Übergangsfunktion. Im allgemeinen legt allerdings ein einzelner Markovprozess die Übergangsfunktion nicht in eindeutiger Weise fest — man betrachte etwa den degenerierten Fall $\xi(t) = \xi(0)$. Die Markovfamilie aber legt natürlich die Übergangsfunktion eindeutig fest.

Von besonderer Bedeutung sind Übergangsfunktionen, die nur von der Differenz $t - s$ abhängen. In diesem Fall schreiben wir

$$P(t, x, A) = P(s, s + t, x, A)$$

und nennen den zugehörigen Prozess homogen. Wir werden in der Folge nur homogene Prozesse betrachten. Das bedeutet keine wesentliche Einschränkung, weil man aus jedem n -dimensionalen Markovprozess einen homogenen $n + 1$ -dimensionalen Prozess erhalten kann, indem man $\xi_{n+1}(t) = t$ setzt.

3.4.2 Die starke Markoveigenschaft

Die Markoveigenschaft bedeutet, anschaulich gesprochen, dass das Verhalten des Prozesses ξ nach dem Zeitpunkt t mit dem Verhalten eines Prozesses übereinstimmt, der zum Zeitpunkt 0 in $\xi(t)$ startet. Für gewisse Anwendungen möchte man diese Tatsache auch dann ausnutzen, wenn man den fixen Zeitpunkt t durch einen zufälligen Zeitpunkt τ ersetzt. Es ist klar, dass solche Zeitpunkt nur Stoppzeiten sein können — wenn τ von der Zukunft des Prozesses abhängt, kann man nicht erwarten, dass das Verhalten von ξ in der Zukunft nur vom Wert $\xi(\tau)$ abhängt.

Wir definieren:

Definition 3.5 Ein Markovprozess ξ heißt starker Markovprozess, wenn für jede endliche Stoppzeit τ

$$(\xi(\tau + t) \in A | \mathfrak{F}_\tau) = P(t, \xi(\tau), A).$$

Diese Forderung stellt eine echte Verschärfung der Markoveigenschaft dar. Es gibt also Prozesse, die zwar die einfache Markoveigenschaft erfüllen, nicht aber die starke. Als Beispiel betrachten wir den Prozess mit der Übergangsfunktion

$$P(t, x, A) = \begin{cases} \int_A \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/2t} dy & \text{wenn } x \neq 0, \\ I_A(0) & \text{sonst.} \end{cases}$$

Diese Funktion erfüllt alle Forderungen an eine Übergangsfunktion, und die zugehörige Markovfamilie ist gegeben durch

$$\xi(t) = x + \beta(t),$$

wenn $\xi(0) = x \neq 0$, und

$$\xi(t) = 0,$$

wenn $\xi(0) = 0$.

Setzt man nun $\tau = \inf\{t : \xi(t) = 0\}$, dann ist τ eine Stoppzeit, und wenn $\xi(0) = x \neq 0$, dann ist $\xi(\tau + t)$ ein Wienerprozess (wir setzen voraus, dass der Wienerprozess die starke Markoveigenschaft erfüllt, das wird in Kürze bewiesen), und nicht konstant gleich null, was der Fall wäre, wenn unsere Markov-Familie die starke Markoveigenschaft erfüllt.

Wir wollen nun überlegen, welche Bedingungen wir an eine Familie von Markovprozessen stellen müssen, damit die starke Markoveigenschaft erfüllt ist. Wir wollen also erreichen, dass für eine Stoppzeit τ

$$\mathbb{P}(\xi(\tau + t) \in A | \mathfrak{F}_\tau) = P(t, \xi(\tau), A)$$

gilt. Das ist äquivalent zu der Forderung, dass für jedes beschränkte und stetige f

$$\mathbb{E}(f(\xi(\tau + t)) | \mathfrak{F}_\tau) = \int P(t, \xi(\tau), dy) f(y).$$

Wir setzen

$$g(x) = \int P(t, x, dy) f(y),$$

und nach der Definition der bedingten Erwartung ist die obige Forderung äquivalent dazu, dass für jedes $A \in \mathfrak{F}_\tau$

$$\mathbb{E}(I_A f(\xi(\tau + t))) = \mathbb{E}(I_A g(\xi(\tau))).$$

Nun sei $\tau_n = ([\tau n] + 1)/n$. Für jedes n ist τ_n eine Stoppzeit, für $n \rightarrow \infty$ gilt $\tau_n \rightarrow \tau$, und wegen

$$A \cap [\tau_n \leq t] = A \cap [\tau < [nt]/n] \in \mathfrak{F}_{[nt]/n} \subseteq \mathfrak{F}_t$$

ist $A \in \mathfrak{F}_{\tau_n}$. Weil τ_n nur abzählbar viele Werte annimmt, gilt für τ_n die starke Markoveigenschaft, also

$$\mathbb{E}(I_A f(\xi(\tau_n + t))) = \mathbb{E}(I_A g(\xi(\tau_n))).$$

In dieser Gleichung wollen wir $n \rightarrow \infty$ gehen lassen. Dabei sollen beide Seiten gegen die Formeln konvergieren, die man erhält, wenn man jeweils τ_n durch τ ersetzt. Für die linke Seite funktioniert das jedenfalls, wenn $\xi(\tau_n + t)$ gegen $\xi(\tau + t)$ konvergiert, und das ist sicher der Fall, wenn die Trajektorien von ξ rechtsstetig sind. Auf der rechten Seite sollte dann auch noch $g(\xi(\tau_n))$ gegen $g(\xi(\tau))$ konvergieren. Dafür ist hinreichend, dass g stetig ist. Diese Eigenschaft ist uns eine eigene Definition wert:

Definition 3.6 Die Markovfamilie ξ mit der Übergangsfunktion P heißt *Fellersch*, wenn für jede beschränkte und stetige Funktion f und jedes $t > 0$ die Funktion

$$g(x) = \mathbb{E}_x(f(\xi(t))) = \int f(y) P(t, x, dy)$$

stetig ist.

Unsere Überlegungen von vorhin können wir jetzt so zusammenfassen:

Satz 3.14 *Wenn ξ eine Fellersche Markov-Familie mit rechtsstetigen Trajektorien ist, dann gilt die strenge Markoveigenschaft.*

Als Beispiel für die Anwendung der starken Markoveigenschaft dient das *Spiegelungsprinzip* für den Wienerprozess. Wir wollen die Wahrscheinlichkeit

$$\mathbb{P}\left(\sup_{0 \leq t \leq 1} \beta(t) \geq x\right)$$

berechnen. Wir setzen $\tau = \inf\{t : \beta(t) \geq x\}$. Wegen der Stetigkeit der Trajektorien gilt $\beta(\tau) = x$ wenn τ endlich ist, und die gesuchte Wahrscheinlichkeit ist gleichbedeutend mit

$$\mathbb{P}(\tau \leq 1).$$

Wir zerlegen diese Wahrscheinlichkeit:

$$\begin{aligned} \mathbb{P}(\tau \leq 1) &= \mathbb{P}(\tau \leq 1, \beta(1) > x) + \mathbb{P}(\tau \leq 1, \beta(1) \leq x) = \\ &= \mathbb{P}(\tau \leq 1, \beta(1) - \beta(\tau) > 0) + \mathbb{P}(\tau \leq 1, \beta(1) - \beta(\tau) \leq 0). \end{aligned}$$

Wegen der starken Markoveigenschaft ist $\beta(s + \tau) - \beta(\tau)$ ein Wienerprozess, der unabhängig von \mathcal{F}_τ ist; insbesondere ist er symmetrisch um 0, also können wir in der letzten Wahrscheinlichkeit die Ungleichung umdrehen und erhalten

$$\begin{aligned} \mathbb{P}(\tau \leq 1) &= \mathbb{P}(\tau \leq 1, \beta(1) - \beta(\tau) > 0) + \mathbb{P}(\tau \leq 1, \beta(1) - \beta(\tau) \geq 0) = \\ &= \mathbb{P}(\tau \leq 1, \beta(1) > x) + \mathbb{P}(\tau \leq 1, \beta(1) \geq x) = 2\mathbb{P}(\beta(1) > x) = 2(1 - \Phi(x)). \end{aligned}$$

3.4.3 Trajektorieneigenschaften

Der letzte Satz führt uns zu der Frage, unter welchen Bedingungen an die Übergangsfunktion die Trajektorien der zugehörigen Markovfamilie gewisse Stetigkeitseigenschaften erfüllen. Dazu definieren wir

$$U_\epsilon(x) = \{y : |y - x| < \epsilon\}$$

und

$$\alpha_\epsilon(h) = \sup_{t \leq h} \sup_{x \in X} P(t, x, U_\epsilon^C).$$

Das erste und schwächste Ergebnis in diesem Kapitel ist

Satz 3.15 *Wenn für jedes $\epsilon > 0$*

$$\lim_{h \rightarrow 0} \alpha_\epsilon(h) < 1/2,$$

dann kann man eine Markovfamilie mit der Übergangsfunktion P finden, für die in allen $t \geq 0$ die Grenzwerte

$$\lim_{s \downarrow t} \xi(s)$$

und

$$\lim_{s \uparrow t} \xi(s)$$

existieren.

Wir zeigen zunächst, dass für $S \subset [0, h]$ höchstens abzählbar die Beziehung

$$\mathbb{P}(\sup_{s \in S} |\xi(s) - \xi(0)| > \epsilon) \leq \frac{\alpha_{\epsilon/2}(h)}{1 - \alpha_{\epsilon/2}(h)}. \quad (3.4)$$

Es genügt, dieses Resultat für endliches S zu zeigen, der allgemeine Fall folgt daraus durch Grenzübergang. Dafür definieren wir Stopzeiten

$$\tau = \inf\{s \in S : |\xi(s) - \xi(0)| > \epsilon\}$$

und

$$\sigma = \inf\{s \in S, s > \tau : |\xi(s) - \xi(\tau)| > \epsilon/2\}.$$

Damit erhalten wir

$$\alpha_{\epsilon/2}(h) \geq \mathbb{P}(|\xi(t) - \xi(0)| > \epsilon/2) \geq \mathbb{P}(\tau \leq h, \sigma > h) \geq \mathbb{P}(\tau \leq h, \sigma - \tau > h) \geq \mathbb{P}(\tau \leq h)(1 - \alpha_{\epsilon/2}(h)).$$

Wir greifen auf Satz 2.4 zurück. Wir fixieren rationale Zahlen $a < b$ und setzen $\epsilon = b - a$. Dazu gibt es ein $h > 0$ mit

$$\alpha_{\epsilon/2}(h) < 1/2.$$

Die Anzahl der Durchquerungen von $[a, b]$ auf dem Intervall $[0, t_0]$ ist genau dann endlich, wenn dasselbe auf jedem Teilintervall der Länge h gilt, wir können uns also o.B.d.A. auf das Intervall $[0, h]$ beschränken. Es sei also S eine abzählbare Teilmenge dieses Intervalls.

Für eine endliche Teilmenge S' von S definieren wir

$$\tau_0 = 0,$$

$$\sigma_i = \inf\{t \geq \tau_{i-1}, t \in S' : \xi(t) < a\},$$

$$\tau_i = \inf\{t \geq \sigma_i, t \in S' : \xi(t) > b\}.$$

Die Wahrscheinlichkeit dafür, dass es auf S' mindestens n Durchquerungen gibt, lässt sich folgendermaßen abschätzen

$$\mathbb{P}(N \geq n) = \mathbb{P}(\tau_n \leq h) \leq \mathbb{P}(\tau_i - \sigma_i \leq h, i = 1, \dots, n) \leq \left(\frac{\alpha_{\epsilon/2}(h)}{1 - \alpha_{\epsilon/2}(h)} \right)^n.$$

Diese Abschätzung ist unabhängig von der Größe von S' und bleibt daher erhalten, wenn wir S' gegen S steigen lassen. Sie strebt für $n \rightarrow \infty$ gegen 0, also ist die Anzahl der Durchquerungen von $[a, b]$ auf S mit Wahrscheinlichkeit eins endlich.

Aus diesem Satz ergibt sich unmittelbar eine hinreichende Bedingung für die Existenz von rechtsstetigen Trajektorien. Dazu muss nur der rechtsseitige Grenzwert fast sicher mit $\xi(t)$ übereinstimmen. Weil die fast sichere Konvergenz die Konvergenz in Wahrscheinlichkeit impliziert, gibt es also rechtsstetige Trajektorien, wenn für $s \downarrow t$ $\xi(s)$ in Wahrscheinlichkeit gegen $\xi(t)$ konvergiert, und wir erhalten

Satz 3.16 Wenn für jedes $\epsilon > 0$

$$\lim_{h \rightarrow 0} \alpha_\epsilon(h) = 0,$$

dann gibt es eine Markovfamilie mit der Übergangsfunktion P die in allen Punkten rechtsstetige Trajektorien mit linksseitigen Grenzwerten hat.

Für die Existenz von stetigen Trajektorien gilt

Satz 3.17 (Dynkin-Kinney) Wenn für jedes $\epsilon > 0$ und $h \rightarrow 0$

$$\alpha_\epsilon(h) = o(h),$$

dann kann die zugehörige Markovfamilie mit stetigen Trajektorien gewählt werden.

Wir wählen ein abzählbares $S \subset [0, t]$, eine Folge $\epsilon_n \rightarrow 0$ und h_n so, dass

$$\sum_n \frac{\alpha_{\epsilon_n/2}(h_n)}{h_n} < \infty.$$

Dazu definieren wir Ereignisse

$$A_n = \{\exists i < t/h_n, d \leq h_n : ih_n + d \in S, |\xi(ih_n + d) - \xi(ih_n)| > \epsilon\}.$$

Mit den Argumenten von vorhin kommen wir zu der Abschätzung

$$\mathbb{P}(A_n) \leq \frac{t}{h_n} \frac{\alpha_{\epsilon/2}(h)}{1 - \alpha_{\epsilon/2}(h)}.$$

Die Summe dieser Wahrscheinlichkeiten ist endlich, deshalb treten nach dem Lemma von Borel-Cantelli mit Wahrscheinlichkeit 1 nur endlich viele davon auf, und das impliziert die gleichmäßige Stetigkeit von ξ auf S , und Satz 2.3 garantiert die Existenz einer Version von ξ mit stetigen Trajektorien.

Wir kehren jetzt kurz zu Satz 3.16 zurück. Dessen Aussage bleibt richtig, wenn wir den absoluten Abstand $|y-x|$ durch eine allgemeinere Distanz ersetzen. Besonders interessant ist die Distanz, die durch die Metrik

$$d(x, y) = \begin{cases} 0 & \text{wenn } x = y, \\ 1 & \text{sonst} \end{cases}$$

gegeben ist. Die Funktionen, die bezüglich dieser Distanz rechtsstetig sind und in allen Punkten linksseitige Grenzwerte haben, sind genau die stückweise konstanten Funktionen, die auf jedem endlichen Intervall nur endlich viele Sprungstellen haben. Es ergibt sich also

Satz 3.18 wenn für $h \rightarrow 0$ gleichmäßig in x

$$\lim_{t \rightarrow 0} P(t, x, \{x\}) = 1,$$

dann gibt es eine Markovfamilie mit der Übergangsfunktion P , deren Trajektorien stückweise konstant sind und auf jedem endlichen Intervall nur endlich viele Sprünge haben.

3.4.4 Übergangsoperatoren

In diesem Abschnitt wollen wir ein Analog zu den Übergangsmatrizen von Markovketten entwickeln. Dazu sei $\xi(t)$ eine Markovkette mit den Übergangsmatrizen

$$P^t = (p_{ij}(t))_{X \times X}.$$

Wir überlegen und als erstes Interpretationen für das Produkt der Übergangsmatrizen mit einem Vektor. Dazu sei

$$p_i = \mathbb{P}(\xi(s) = i)$$

die Verteilung von $\xi(s)$. Nach dem Satz von der vollständigen Wahrscheinlichkeit ergibt sich für die Verteilung von $\xi(s+t)$

$$\mathbb{P}(\xi(s+t) = j) = \sum_i p_i p_{ij}(t) = (pP^t)_j.$$

Andererseits ist für eine beschränkte Folge f_i

$$(P^t f)_i = \sum_j p_{ij}(t) f_j = \mathbb{E}_i(f(\xi(t))).$$

Für einen allgemeinen Markovprozess ist die Verteilung von $\xi(s)$ ein Wahrscheinlichkeitsmaß ν auf dem Zustandsraum X , und für die Verteilung von $\xi(s+t)$ ergibt sich

$$\mathbb{P}(\xi(t+s) \in A) = \int P(t, x, A) \nu(dx).$$

Das rechte Integral ergibt nicht nur für Wahrscheinlichkeitsmaße ν einen Sinn, sondern für alle endlichen signierten Maße, und wir definieren

$$\nu P^t(A) = \int P(t, x, A) \nu(dx),$$

und nennen P^t die Übergangsoperatoren.

Auch für die Multiplikation von rechts kann man ein Analog finden. Dazu sei f eine beschränkte messbare Funktion und

$$P^t f(x) = \mathbb{E}_x(f(\xi(t))) = \int f(y) P(t, x, dy).$$

Auf dem Raum \mathbf{V} der endlichen signierten Maße ist durch die Totalvariation

$$\|\nu\| = |\nu|(X)$$

eine Norm definiert, auf dem Raum \mathbf{B} der beschränkten Funktionen verwenden wir die Supremumsnorm

$$\|f\| = \sup_{x \in X} |f(x)|.$$

Beide Räume sind mit den jeweiligen Normen vollständig, also Banachräume.

Für die Übergangsoperatoren gilt

Satz 3.19 *Eigenschaften von Übergangsoperatoren:*

1. P^t ist eine Kontraktion, d.h.,

$$\|P^t f\| \leq \|f\| \text{ bzw. } \|\nu P^t\| \leq \|\nu\|,$$

2. P^t ist positiv, d.h., wenn $f \geq 0$ (bzw. $\nu \geq 0$), dann gilt auch $P^t f \geq 0$ (bzw. $\nu P^t \geq 0$).

3. $P^t \mathbf{1} = \mathbf{1}$ bzw. $\nu P^t(\Omega) = \nu(\Omega)$,

4. die Halbgruppeneigenschaft (vulgo Chapman-Kolmogorov):

$$P^{t+s} = P^t P^s.$$

5. Die beiden Operatoren auf \mathbf{V} und \mathbf{B} sind konjugiert:

$$\int f d(\nu P^t) = \int P^t f d\nu.$$

Wir sprechen von der Halbgruppe von Kontraktionen P^t .

Es eröffnet sich die Frage, ob jede Halbgruppe von Operatoren auf \mathbf{B} (bzw. auf \mathbf{V}), die die Bedingungen von Satz 3.19 erfüllt, auch eine Markovfamilie (bzw. die Übergangsfunktion) festlegt (es sollte klar sein, dass die zugehörige Übergangsfunktion eindeutig festgelegt ist, wenn sie denn existiert). Wenn der Operator P^t auf \mathbf{B} definiert ist, kann der zugehörige Operator auf \mathbf{V} als dualer Operator erhalten werden, der allerdings nicht auf \mathbf{V} , sondern auf dem (außer für endliches X) umfangreicheren Raum \mathbf{B}^* , das heißt, dass νP^t zwar (als Element von \mathbf{B}^*) bestimmt werden kann, aber kein Maß zu sein braucht (beginnt man umgekehrt mit Operatoren auf \mathbf{V} , dann gibt es dieselben Probleme). Einen Weg aus diesem Dilemma zeigt der Darstellungssatz von Riesz: \mathbf{V} ist nämlich der Dualraum zum Raum \mathbf{C} der stetigen Funktionen. Wenn P^t also eine Halbgruppe ist, die auf \mathbf{C} operiert (also Fellersch), dann legt sie in jedem Fall eine Übergangsfunktion fest.

Unser Ziel ist es natürlich, ein Äquivalent zum infinitesimalen Erzeuger der Markovketten zu finden. Wir definieren also den infinitesimalen Operator durch

$$Qf = \lim_{t \rightarrow 0} \frac{P^t f - f}{t}. \quad (3.5)$$

Dieser Grenzwert muss nicht existieren — die Menge der Funktionen, für die er (im Sinne der Normkonvergenz) existiert, nennen wir den Definitionsbereich D_Q des infinitesimalen Operators. Für diese Funktionen folgt aus Gleichung (3.5) die Beziehung

$$\lim_{t \rightarrow 0} \|P^t f - f\| = 0.$$

Das bringt uns dazu, den Raum der starken Stetigkeit

$$\mathbf{B}_0 = \{f \in \mathbf{B} : \lim_{t \rightarrow 0} \|P^t f - f\| = 0\}$$

zu definieren.

Als Beispiel betrachten wir den Wienerprozess: $P^t f$ ist für jede beliebige beschränkte Funktion f und $t > 0$ eine beschränkte Ableitung hat und daher gleichmäßig stetig ist. Für $f \in \mathbf{B}_0$ konvergiert $P^t f$ gegen f , das als gleichmäßiger Limes von gleichmäßig stetigen Funktionen selbst gleichmäßig stetig sein muss. Umgekehrt stellt man leicht fest, dass eine gleichmäßig stetige Funktion in \mathbf{B}_0 liegt. Der Raum der starken Stetigkeit für den Wienerprozess stimmt also mit dem Raum der beschränkten gleichmäßig stetigen Funktionen überein.

Wir haben schon die Inklusion $D_Q \subseteq \mathbf{B}_0$ gefunden. Außerdem gilt:

Satz 3.20 \mathbf{B}_0 ist ein abgeschlossener Teilraum von \mathbf{B} . Weiters gelten die Inklusionen $P^t \mathbf{B}_0 \subseteq \mathbf{B}_0$ und $QD_Q \subseteq \mathbf{B}_0$.

Es sein f_n eine Folge von Funktionen aus \mathbf{B}_0 , die gleichmäßig gegen f konvergiert. Dann ist wegen der Kontraktionseigenschaft von P^t

$$\|P^t f - f\| \leq \|P^t f - P^t f_n\| + \|P^t f_n - f_n\| + \|f_n - f\| \leq \|P^t f_n - f_n\| + 2\|f_n - f\|.$$

Wir können n so wählen, dass der zweite Term auf der rechten Seite kleiner als ϵ wird, und für dieses n wird der erste Term kleiner als ϵ , wenn nur t klein genug ist. f ist also in B_0 enthalten.

Wegen

$$\|P^s P^t f - P^t f\| = \|P^t P^s f - P^t f\| = \|P^t(P^s f - f)\| \leq \|P^s f - f\|$$

ist $P^t f$ in \mathbf{B}_0 , wenn f in \mathbf{B}_0 ist.

Wegen der Abgeschlossenheit von \mathbf{B}_0 ist natürlich für $f \in D_Q$ auch Qf als gleichmäßiger Grenzwert von Funktionen aus \mathbf{B}_0 auch in \mathbf{B}_0 enthalten.

Unser Ziel ist natürlich, die Halbgruppe P^t und damit die Übergangsfunktion durch Q festzulegen. Damit das möglich ist, muss die Verteilung von $\xi(t)$ durch die Angabe der Erwartungswerte $P^t f = \mathbb{E}f(\xi(t))$ für $f \in \mathbf{b}_0$ eindeutig bestimmt sein. \mathbf{B}_0 muss also reichhaltig genug sein, um Punkte zu trennen.

Ohne weitere Probleme erhalten wir jedenfalls die Kolmogorov'schen Differentialgleichungen:

Satz 3.21 Für $f \in D_Q$ ist auch $P^t f \in D_Q$ und es gelten die Vorwärts- und Rückwärtsgleichungen:

$$\frac{\partial}{\partial t}(P^t f(x)) = P^t Qf(x) = QP^t f(x).$$

Für $s > 0$ und $f \in D_Q$ gilt (wieder wegen der Kontraktionseigenschaft)

$$\|s^{-1}(P^{t+s} f - P^t f) - P^t Qf\| = \|s^{-1}(P^s P^t f - P^t f) - P^t Qf\| =$$

$$\|P^t(s^{-1}(P^s f - f) - Qf)\| \leq \|s^{-1}(P^s f - f) - Qf\|.$$

Der letzte Term geht für $s \rightarrow 0$ nach der Definition von D_Q gegen 0, also auch die anderen. Für den zweiten Teil bedeutet das, dass auch $P^t f \in D_Q$ gilt und

dass $QP^t f = P^t Qf$ gilt. Der erste Term zeigt, dass die rechtsseitige Ableitung von $P^t f$ nach t gleich $P^t Qf$ ist. Dass das auch für die linksseitige Ableitung gilt, folgt aus Ungleichung

$$\|s^{-1}(P^t f - P^{t-s} f) - P^t Qf\| \|s^{-1}(P^t f - P^{t-s} f) - P^{t-s} Qf\| + \|P^{t-s} Qf - P^t Qf\|.$$

Der erste Term auf der rechten Seite geht für $s \rightarrow 0$ gegen 0, wie wir gerade gesehen haben, der zweite wegen $Qf \in \mathbf{B}_0$.

Eine wichtige Eigenschaft des infinitesimalen Operators ist das Maximumprinzip:

Satz 3.22 Falls $f \in D_Q$ in x ein absolutes Maximum besitzt, dann gilt

$$Qf(x) \leq 0.$$

Das ist klar, weil $P^t f(x) = \mathbb{E}_x f(\xi(t)) \leq f(x)$ und deswegen

$$t^{-1}(P^t f(x) - f(x)) \leq 0,$$

und für $t \rightarrow 0$ ergibt sich die Behauptung.

Zum weiteren Studium des infinitesimalen Operators definieren wir zunächst die *Resolvente*:

Definition 3.7 Für $\lambda > 0$ heißen die Operatoren

$$R_\lambda f = \int_0^\infty e^{-\lambda t} P^t f dt$$

die *Resolvente der Halbgruppe P^t* .

Für $f \in \mathbf{B}_0$ ist die Resolvente wohldefiniert und legt (nach dem Umkehrungssatz für Laplacetransformationen) die Übergangoperatoren eindeutig fest, weil ja $P^t f(x)$ als Funktion von t stetig ist.

Satz 3.23 Für die Resolvente gelten die folgenden Eigenschaften:

$$\|R_\lambda\| = 1/\lambda, \tag{3.6}$$

$$\lim_{\lambda \rightarrow \infty} \lambda R_\lambda f = f, \tag{3.7}$$

$$R_\lambda R_\mu = \frac{R_\mu - R_\lambda}{\mu - \lambda} \quad (\lambda \neq \mu), \tag{3.8}$$

$$P^t R_\lambda f(x) = R_\lambda P^t f(x) = e^{\lambda t} \left(R_\lambda f(x) - \int_0^t e^{-\lambda s} P^s f(x) ds \right). \tag{3.9}$$

Die Gleichungen (3.9) und (3.8) ergeben sich durch einfache Umformungen der entsprechenden Integrale. Für (3.6) stellen wir fest, dass

$$\left\| \int e^{-\lambda t} P^t f dt \right\| \leq \int e^{-\lambda t} \|P^t f\| dt \leq \int e^{-\lambda t} \|f\| dt = \frac{1}{\lambda} \|f\|,$$

mit Gleichheit für $f = \text{const.}$

Für (3.7) berechnen wir

$$\|\lambda R_\lambda f - f\| = \left\| \int \lambda e^{-\lambda t} (P^t f - f) dt \right\| \leq \int \lambda e^{-\lambda t} \|P^t f - f\| dt = \int e^u \|P^{u/\lambda} f - f\| dt.$$

Im letzten Integral geht der Integrand für $\lambda \rightarrow \infty$ punktweise gegen 0, und der Satz von der dominierten Konvergenz ergibt, dass auch das Integral gegen 0 geht.

Die Gleichung (3.8) zeigt, dass $R_\lambda f$ für jedes $\mu \neq \lambda$ auch in $R_\mu \mathbf{B}_0$ liegt, es gilt also $R_\lambda \mathbf{B}_0 \subseteq R_\mu \mathbf{B}_0$, und weil die Rollen von λ und μ vertauschbar sind, ist der Wertebereich $R_\lambda \mathbf{B}_0$ für alle λ gleich. Wegen (3.7) ist dieser Wertebereich dicht in \mathbf{B}_0 . Wir werden sehen, dass dieser Wertebereich mit D_Q übereinstimmt:

Satz 3.24 Die Resolvente R_λ ist die Umkehrung des Operators $\lambda \text{id} - Q$, genauer:

1. Für $f \in D_Q$ gilt

$$R_\lambda(\lambda f - Qf) = f$$

und

2. für $f \in \mathbf{B}_0$ ist $R_\lambda f \in D_Q$ und

$$\lambda R_\lambda f - QR_\lambda f = f.$$

Nach (3.9) gilt

$$\begin{aligned} t^{-1}(P^t R_\lambda f - R_\lambda f) &= t^{-1} R_\lambda (P^t f - f) = \frac{e^{\lambda t} - 1}{t} R_\lambda f - \frac{1}{t} \int_0^t e^{\lambda(t-s)} P^s f ds = \\ &= \frac{e^{\lambda t} - 1}{t} R_\lambda f - \int_0^1 e^{\lambda t(1-u)} P^{tu} f du. \end{aligned}$$

Die rechte Seite konvergiert für $f \in \mathbf{B}_0$ gegen $\lambda R_\lambda f - f$. Der erste Ausdruck auf der linken Seite zeigt, dass $R_\lambda f \in D_Q$ liegt und dass

$$QR_\lambda f = \lambda R_\lambda f - f$$

gilt; für $f \in D_Q$ konvergiert der zweite Ausdruck auf der linken Seite gegen $R_\lambda Qf$, und es ergibt sich

$$R_\lambda Qf = \lambda R_\lambda f - f.$$

Im allgemeinen ist Q kein beschränkter (also stetiger) Operator. Er ist aber immerhin abgeschlossen:

Definition 3.8 Ein Operator A , der auf dem Teilraum D des Banachraums \mathbf{B} definiert ist, heißt abgeschlossen, wenn für $f_n \in D$ aus $f_n \rightarrow f$ und $Af_n \rightarrow g$ folgt, dass $f \in D$ und $Af = g$ gilt.

Satz 3.25 *Q ist abgeschlossen.*

Wir setzen $g_n = Qf_n$. Der vorige Satz besagt, dass

$$f_n = \lambda R_\lambda f_n - R_\lambda g_n.$$

Für $n \rightarrow \infty$ konvergiert f_n gegen f und g_n gegen g , also

$$f = \lambda R_\lambda f - R_\lambda g.$$

Daher ist f in $R_\lambda \mathbf{B}_0 = D_Q$ und Anwendung von $\lambda \text{id} - Q$ auf beiden Seiten ergibt $\lambda f - Qf = \lambda f - g$, also $Qf = g$.

Wir haben also folgende Eigenschaften des infinitesimalen Operators Q gefunden:

1. Q ist abgeschlossen.
2. Der Definitionsbereich D_Q liegt dicht in \mathbf{B}_0 .
3. Für jedes $\lambda > 0$ ist $R_\lambda = (\lambda \text{id} - Q)^{-1}$ ein beschränkter Operator auf ganz \mathbf{B}_0 mit $\|R_\lambda\| = 1/\lambda$.
4. Q erfüllt das Maximumprinzip.

Insbesondere lässt sich aus Q die Resolvente eindeutig bestimmen und damit die Übergangsooperatoren auf \mathbf{B}_0 ; wenn \mathbf{B}_0 reichhaltig genug ist, wird damit auch die Übergangsfunktion eindeutig bestimmt. Es bleibt die Frage, ob jeder Operator, der die genannten Eigenschaften erfüllt, auch infinitesimaler Operator einer Halbgruppe von Kontraktionen ist. Das sagt der Satz von Hille-Yosida:

Satz 3.26 *Q ist genau dann der infinitesimale Operator einer stark stetigen Halbgruppe auf dem Banachraum \mathbf{B} , wenn*

1. Q ist abgeschlossen
2. Der Definitionsbereich D_Q liegt dicht in \mathbf{B}_0 .
3. Für jedes $\lambda > 0$ ist $R_\lambda = (\lambda \text{id} - Q)^{-1}$ ein beschränkter Operator auf ganz \mathbf{B}_0 mit $\|R_\lambda\| \leq 1/\lambda$.

Wenn der Banachraum B der Raum $C(X)$ der stetigen Funktionen auf dem Kompaktum X ist, kann Punkt 3. ersetzt werden durch

- 3a. Für jedes $f \in \mathbf{B}$ und $\lambda > 0$ hat die Gleichung $\lambda g - Qg = f$ eine Lösung in D_Q .
- 3b. Es gilt das Maximumprinzip.

(Ohne Beweis).

3.4.5 Der infinitesimale Operator und Trajektorieneigenschaften

Wir wenden uns nun der Frage zu, wie sich die Bedingungen für die (Rechts-) Stetigkeit der Trajektorien am infinitesimalen Operator erkennen lassen. Wir erinnern uns, dass die Rechtsstetigkeit der Trajektorien aus der gleichmäßigen stochastischen Stetigkeit des Prozesses folgt. Für den infinitesimalen Operator haben wir:

Satz 3.27 *Der Zustandsraum X sei kompakt und zu jedem $x \in X$ und jedem $\epsilon > 0$ existiere ein $f \in D_Q$ mit $0 \leq f \leq 1$, $f(y) = 0$ für $|y - x| < \epsilon/3$ und $f(x) = 1$ für $|y - x| > 2\epsilon/3$. Dann hat ξ eine Version mit rechtstetigen Trajektorien (und linksseitigen Grenzwerten).*

Wegen der Kompaktheit können wir X mit endlich vielen Umgebungen $U(x_i, \epsilon/3)$, $i = 1, \dots, n$ überdecken. Die zugehörigen Funktionen f_i sind in \mathbf{B}_0 , also konvergiert für $t \rightarrow 0$ $P^t f_i$ gleichmäßig gegen f_i . Wir können also für jedes $\delta > 0$ ein t_0 finden, sodass für $t < t_0$ und $i = 1, \dots, n$ $\|P^t f_i - f_i\| < \delta$ gilt. Für beliebiges $x \in X$ gibt es ein i mit $|x_i - x| < \epsilon/3$; dann ist $f_i \geq I_{V(x, \epsilon)}$ und daher

$$\mathbb{P}_x(\xi(t) \in V(x, \epsilon) < P^t f_i(x) < \delta).$$

Diese Wahrscheinlichkeit geht daher gleichmäßig gegen 0, was zu zeigen war.

Die Dynkin-Kinney Bedingung für die Stetigkeit ist verbunden mit der *Lokalität* von Q :

Definition 3.9 *Der Operator A heißt lokal, wenn wenn der Wert $Af(x)$ nur vom Verhalten von f in einer Umgebung von x abhängt. Genauer: wenn f und g zwei Funktionen aus dem Definitionsbereich von Q in einer Umgebung von x übereinstimmen, dann gilt*

$$Af(x) = Ag(x).$$

Typische Beispiele für lokale Operatoren sind Differentialoperatoren, Integraloperatoren sind nicht lokal. Es gilt:

Satz 3.28 *Wenn der Prozess ξ die Dynkin-Kinney Bedingung erfüllt, dann ist Q lokal.*

In der umgekehrten Richtung braucht man etwas mehr:

Satz 3.29 *Der Zustandsraum X sei kompakt, Q sei lokal, und zu jedem $x \in X$ und jedem $\epsilon > 0$ existiere ein $f \in D_Q$ mit $0 \leq f \leq 1$, $f(y) = 0$ für $|y - x| < \epsilon/3$ und $f(x) = 1$ für $|y - x| > 2\epsilon/3$. Dann hat ξ eine Version mit stetigen Trajektorien.*

Es funktioniert dieselbe Methode wie beim vorigen Satz, nur muss man jetzt ausnutzen, dass $f_i \in D_Q$ ist. Wegen der Lokalität von Q stimmt für $|x - x_i| < \epsilon/3$ $Qf(x)$ mit $Q0(x) = 0$ überein, und deshalb konvergiert jetzt $P^t f_i(x)/t$ gleichmäßig gegen 0, was genau der Dynkin-Kinney Bedingung entspricht.

Für Treppenförmige Trajektorien haben wir

Satz 3.30 *Wenn Q auf ganz \mathbf{B} definiert und beschränkt ist, dann besitzt ξ stückweise konstante Trajektorien mit einer endlichen Anzahl von Sprüngen auf jedem endlichen Intervall.*

Man kann in diesem Fall die Form von Q genauer angeben. Für eine Borelmenge A mit $x \notin A$ setzen wir

$$\nu_{h,x}(A) = P^h I_A(x)/h.$$

Die $\nu_{h,x}$ sind endliche Maße auf dem Sigmaring $\mathcal{B} \cap \mathbb{R} \setminus \{x\}$, und konvergieren für $h \rightarrow 0$ punktweise gegen eine Funktion ν_x , die ebenso ein Maß sein muss, und wegen der Beschränktheit von Q ist ν_x endlich. Insgesamt ergibt sich

$$Qf(x) = \int_X f(y)\nu_x(dy) - \nu_x(X)f(x).$$

Q ist also ein Integraloperator.

Die Interpretation dieses Ergebnisse ist natürlich wieder, dass der Prozess in x eine Zeit verbringt, die mit Parameter $\nu_x(X)$ exponentialverteilt ist, und dann in einen Zustand springt, der nach $\nu_x/\nu_x(X)$ verteilt ist.

Kapitel 4

Martingale

4.1 Definiton

Die Definition eines Martingales ist bereits erwähnt worden. Wir präzisieren jetzt

Definition 4.1 $(\mathcal{F}_t, t \in T)$ sei eine Filtration. Ein an (\mathcal{F}_t) adaptierter integrierbarer Prozess ξ heißt

- (\mathcal{F}_t) -Martingal, wenn für $s < t$

$$\mathbb{E}(\xi(t)|\mathcal{F}_s) = \xi(s),$$

- (\mathcal{F}_t) -Submartingal, wenn für $s < t$

$$\mathbb{E}(\xi(t)|\mathcal{F}_s) \geq \xi(s),$$

- (\mathcal{F}_t) -Supermartingal, wenn für $s < t$

$$\mathbb{E}(\xi(t)|\mathcal{F}_s) \leq \xi(s).$$

Sub- und Supermartingale werden gemeinsam als *Semimartingale* bezeichnet. Wenn klar ist, um welche Filtration es sich handelt, kann auf ihre explizite Erwähnung verzichtet werden (insbesondere, wenn es sich um die natürliche Filtration des Prozesses handelt). Wir werden in der Folge alle Martingale und Semimartingale bezüglich einer bestimmten Filtration verstehen und diese nicht besonders erwähnen (insbesondere bedeutet die Tatsache, dass wir die Filtration nicht erwähnen, nicht, dass wir die Prozesse bezüglich ihrer natürlichen Filtration verstehen).

Die bekanntesten Beispiele für Martingale in diskreter Zeit sind Summen von unabhängigen Zufallsvariablen mit verschwindendem Erwartungswert. In stetiger Zeit entsprechen dem Prozesse mit unabhängigen Zuwächsen (ebenfalls mit

verschwindendem Erwartungswert); insbesondere ist der Wienerprozess ein Martingal. Wenn die Zuwächse positive bzw. negative Erwartungswerte haben, dann ergeben sich Sub- bzw. Supermartingale. Ein anderes Beispiel sind Doob'sche Martingale, bei denen η eine beliebige integrierbare Zufallsvariable ist und

$$\xi(t) = \mathbb{E}(\eta | \mathcal{F}_t).$$

In der deterministischen Welt entsprechen den Semimartingalen die monotonen Funktionen, und in gewisser Weise stellen die Martingale eine stochastische Verallgemeinerung der monotonen Funktionen dar, insbesondere, was ihr Grenzwertverhalten betrifft.

Durch Transformation lassen sich aus Martingalen bzw. Submartingalen weitere Submartingale bilden:

Satz 4.1 *Falls*

- ξ ein Martingal und f konvex oder
- ξ ein Submartingal und f konvex und monoton nichtfallend

und wenn zusätzlich $f(\xi(t))$ für jedes t integrierbar ist, dann ist $f \circ \xi$ ein Submartingal.

Die Jensen-Ungleichung impliziert

$$\mathbb{E}(f(\xi(t)) | \mathcal{F}_s) \geq f(\mathbb{E}(\xi(t) | \mathcal{F}_s)).$$

Für den Fall, dass ξ ein Martingal ist, ist damit der Satz bewiesen, für ein Submartingal muss man noch die Monotonie von f verwenden.

Eine direkte Folgerung daraus ist:

Satz 4.2 ξ sei ein Martingal oder ein nichtnegatives Submartingal. Dann ist $(\xi(t), t \leq t_0)$ gleichmäßig integrierbar.

Bekanntlich bedeutet die gleichmäßige Integrierbarkeit, dass es zu jedem $\epsilon > 0$ ein M gibt, sodass für alle $t \leq t_0$

$$\mathbb{E}(|\xi(t)| I_{|\xi(t)| > M}) < \epsilon.$$

Wegen $2(x - M/2)_+ \leq x I_{x > M} \leq (x - M)_+$ ist das gleichbedeutend damit, dass es ein M gibt, sodass für alle $t \leq t_0$

$$\mathbb{E}((|\xi(t)| - M)_+) < \epsilon.$$

Die Funktion $(|x| - M)_+$ ist konvex und auf der positiven Halbachse monoton nichtfallend, daher ist $(|\xi(t)| - M)_+$ ein Submartingal und daher

$$\mathbb{E}((|\xi(t)| - M)_+) \leq \mathbb{E}((|\xi(t_0)| - M)_+).$$

Die rechte Seite kann beliebig klein gemacht werden, und das beweist die gleichmäßige Integrierbarkeit.

Insbesondere kann in diesen Fällen die Bildung des Erwartungswerts mit dem Grenzübergang vertauscht werden.

Dieser Satz impliziert außerdem, dass Doob'sche Martingale gleichmäßig integrierbar sind; wir werden in Kürze sehen, dass umgekehrt alle gleichmäßig integrierbare Martingale als Doob'sche Martingale geschrieben werden können.

Wenn man Martingale als Modelle für ein faires Spiel betrachtet, kann man die Frage stellen, ob nicht durch geschicktes Aussteigen und wieder Einsteigen doch ein Reingewinn zu erzielen sein könnte. Die nächsten beiden Sätze zeigen, dass das nicht so ist:

Satz 4.3 (optional stopping) ξ sei ein Martingal (Submartingal) in diskreter Zeit oder in stetiger Zeit mit rechtsstetigen Trajektorien. Weiters sei τ eine Stoppzeit. Dann ist auch

$$\eta(t) = \xi(\min(t, \tau))$$

ein Martingal (Submartingal).

Satz 4.4 (optional selection) ξ sei ein Martingal (Submartingal) in diskreter Zeit oder in stetiger Zeit mit rechtsstetigen Trajektorien. Weiters seien $\sigma_i, \tau_i, i = 1, \dots, n$ Stoppzeiten mit $\sigma_i \leq \tau_i \leq \sigma_{i+1}$. Dann sind auch

$$\eta(t) = \sum_{i=1}^n (\xi(\min(t, \tau_i)) - \xi(\min(t, \sigma_i)))$$

und

$$\xi(t) - \eta(t)$$

Martingale (Submartingale).

Für den Fall der diskreten Zeit sind die Ergebnisse aus der Maßtheorie bekannt. Der Fall der stetigen Zeit wird darauf zurückgeführt, indem τ durch $\tau^{(n)} = \lceil n\tau \rceil / n$ ersetzt wird. Wegen der Rechtsstetigkeit konvergieren die Funktionswerte und wegen der vorhin bewiesenen gleichmäßigen Integrierbarkeit die bedingten Erwartungen gegen die entsprechenden Werte für τ .

Entsprechend lassen sich auch die folgenden Sätze aus der Maßtheorie für den Fall der stetigen Zeit übertragen:

Die Doob'schen Extremalungleichungen:

Satz 4.5 $\xi(t)$ sei ein Submartingal mit rechtsstetigen Trajektorien. Dann gilt:

$$\mathbb{P}(\sup_{s \leq t} \xi(s) \geq \lambda) \leq \frac{\mathbb{E}(\xi(t)_+)}{\lambda}.$$

$$\mathbb{P}(\inf_{s \leq t} \xi(s) \leq -\lambda) \leq \frac{\mathbb{E}(\xi(t)_+) - \mathbb{E}(\xi(0))}{\lambda}.$$

und die upcrossing Ungleichung:

Satz 4.6 ξ sei ein Submartingal, $S \subseteq T$ abzählbar, $a < b$ beliebig. Dann gilt für die Anzahl N der Durchquerungen von $[a, b]$ auf S die Ungleichung:

$$\mathbb{E}(N) \leq \sup_{t \in S} \frac{\mathbb{E}((\xi(s) - a)_+)}{b - a}.$$

Damit überträgt sich auch der Konvergenzsatz für Submartingale in der folgenden Form:

Satz 4.7 ξ sei ein Submartingal, $S \subseteq T$ abzählbar und

$$\sup_{t \in S} \mathbb{E}((\xi(t))_+) < \infty.$$

dann existieren mit Wahrscheinlichkeit eins in allen Häufungspunkten t von S die links- und rechtsseitigen Grenzwerte von $\xi(s)$ für $s \in S, s \rightarrow t$.

Im Lichte von Satz 2.4 bedeutet das, dass wir zu jedem Submartingal eine Version finden können, deren Trajektorien in allen Punkten rechtsseitige und linksseitige Grenzwerte besitzen. Davon zur Existenz einer Version mit rechtsstetigen Trajektorien fehlt nur noch, dass die rechtsseitigen Grenzwerte mit den Funktionswerten übereinstimmen, wozu die Rechtsstetigkeit in Wahrscheinlichkeit genügt. Der einzige Schönheitsfehler dabei ist, dass diese Version nicht mehr an die ursprüngliche Filtration adaptiert sein muss. Deshalb werden von Filtrationen von Martingalen gerne die “üblichen Bedingungen” (“usual conditions”) gefordert:

Definition 4.2 Die Filtration \mathcal{F}_t auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ erfüllt die “üblichen Bedingungen”, wenn

$$\mathcal{F}_{t+} = \bigcap_{s>t} \mathcal{F}_s = \mathcal{F}_t$$

und

$$\{N \in \mathcal{F} : \mathbb{P}(N) = 0\} \subset F_0.$$

Eine weitere Maximalungleichung von Doob:

Satz 4.8 ξ sei ein Martingal oder ein nichtnegatives Submartingal mit rechtsstetigen Trajektorien. Dann gilt:

$$\left\| \sup_{s \leq t} |\xi(s)| \right\|_p \leq \frac{p}{p-1} \|\xi(t)\|_p,$$

falls $p > 1$. Für $p = 1$ gilt

$$\left\| \sup_{s \leq t} |\xi(s)| \right\|_1 \leq \frac{e}{e-1} (1 + \|\xi(t) \log \xi(t)\|_1).$$

Es genügt, den Beweis für ein Submartingal zu führen. Wir setzen $\eta = \sup_{s \leq t} \xi(s)$ und erhalten durch partielle Integration

$$\|\eta\|_p^p = \mathbb{E}(\eta^p) = p \int_0^\infty x^{p-1} \mathbb{P}(\eta \geq x) dx.$$

Mit $\tau_x = \inf\{s \leq t : \xi(s) \geq x\} \cup \{t\}$ und dem optional stopping theorem ergibt sich

$$\begin{aligned} \int_0^\infty x^{p-1} \mathbb{P}(\xi(\tau) \geq x) dx &\leq \int_0^\infty x^{p-1} \frac{1}{x} \mathbb{E}(\xi(\tau) I_{\eta \geq x}) dx \leq \\ \int_0^\infty x^{p-1} \frac{1}{x} \mathbb{E}(\xi(t) I_{\eta \geq x}) dx &= \mathbb{E}(\xi(t) \int_0^\infty I_{x \leq \eta} x^{p-2} dx) = \\ \frac{1}{p-1} \mathbb{E}(\xi(t) \eta^{p-1}) &\leq \frac{1}{p-1} \|\xi(t)\|_p \|\eta\|_p^{p-1}. \end{aligned}$$

Die letzte Ungleichung ergibt sich durch eine Anwendung der Hölderschen Ungleichung.

Für $p = 1$ haben wir

$$\mathbb{E}\eta \leq 1 + \int_1^\infty \mathbb{P}(\eta \geq x) dx.$$

Dieselben Umformungen wie im Fall $p > 1$ führen zu

$$\mathbb{E}\eta \leq 1 + \mathbb{E}(\xi(t) \log(\eta(t)) I_{\eta > 1}) = 1 + \mathbb{E}(\xi(t) \log(\xi(t)) I_{\eta > 1}) + \mathbb{E}(\xi(t) \log\left(\frac{\eta}{\xi(t)}\right) I_{\eta > 1}).$$

Wenn nun auf den letzten Term die Ungleichung $\log x \leq x/e$ angewendet wird, folgt nach einigen trivialen Abschätzungen

$$\mathbb{E}\eta \leq \mathbb{E}(|\xi(t) \log \xi(t)|) + \frac{1}{e} \mathbb{E}\eta.$$

Zu guter letzt befassen wir uns mit der Doob-Meyer Zerlegung. Bekanntlich lässt sich ja ein Submartingal $\xi(t)$ in diskreter Zeit eindeutig als Summe eines Martingals $\eta(t)$ und eines vorhersagbaren nichtfallenden Prozesses $\alpha(t)$ darstellen.

(bekanntlich bedeutet die Vorhersagbarkeit, dass $\alpha(t)$ bezüglich \mathcal{F}_{t-1} messbar ist, und ein nichtfallender Prozess erfüllt $\alpha(0) = 0$ und $\alpha(t) \leq \alpha(t+1)$). Man überzeugt sich leicht, dass

$$\alpha(t) = \sum_{s < t} (\mathbb{E}(\xi_{s+1} | \mathcal{F}_s) - \xi(s))$$

den gesuchten nichtfallenden Prozess darstellt.)

Damit das analoge Ergebnis für Martingale in stetiger Zeit gilt, muss eine zusätzliche Regularitätsbedingung erfüllt sein:

Definition 4.3 Das Submartingal ξ ist von der Klasse (D), wenn die Familie

$$\{\xi(\tau) : \tau < \infty \text{ Stoppzeit}\}$$

gleichmäßig integrierbar ist. Das Submartingal ξ ist von der Klasse (DL), wenn für jedes $t > 0$ die Familie

$$\{\xi(\tau) : \tau \leq t \text{ Stoppzeit}\}$$

gleichmäßig integrierbar ist.

Da sowohl das Martingal η (wegen des optional stopping theorems) und der nichtfallende Prozess α (wegen $\alpha(\tau) \leq \alpha(t)$) von Klasse (LD) sind, muss dasselbe auch für ξ gelten. Diese Bedingung ist auch hinreichend; bevor wir aber den entsprechenden Satz formulieren, benötigen wir noch ein Analog zur Vorhersagbarkeit, um die Zerlegung eindeutig zu machen. Dazu betrachten wir (für den Fall der diskreten Zeit) ein beliebiges Martingal ζ (bezüglich derselben Filtration wie ξ) und das Produkt $\zeta(n)\alpha(n)$ für einen adaptierten Prozess α . Wegen der Bedingung $\alpha(0) = 0$ gilt

$$\begin{aligned} \alpha(n)\zeta(n) &= \sum_{i=0}^{n-1} (\alpha(i+1)\zeta(i+1) - \alpha(i)\zeta(i)) = \\ &= \sum_{i=0}^{n-1} \alpha(i)(\zeta(i+1) - \zeta(i)) + \sum_{i=0}^{n-1} \zeta(i+1)(\alpha(i+1) - \alpha(i)). \end{aligned}$$

Wenn man hier Erwartungswerte nimmt, dann verschwindet die erste Summe auf der rechten Seite, und wir erhalten

$$\mathbb{E}(\alpha(n)\zeta(n)) = \mathbb{E}\left(\sum_{i=0}^{n-1} \zeta(i+1)(\alpha(i+1) - \alpha(i))\right). \quad (4.1)$$

Wenn α noch dazu vorhersagbar ist, dann gilt

$$\mathbb{E}(\alpha(n)\zeta(n)) = \mathbb{E}\left(\sum_{i=0}^{n-1} \zeta(i)(\alpha(i+1) - \alpha(i))\right). \quad (4.2)$$

Wir definieren

Definition 4.4 Der Prozess ξ (in stetiger Zeit) heißt natürlich, wenn für jedes beschränkte Martingal ζ

$$\mathbb{E}(\alpha(n)\zeta(n)) = \mathbb{E}\left(\sum_{i=0}^{n-1} \zeta(i)(\alpha(i+1) - \alpha(i))\right)$$

gilt.

Die Natürlichkeit von α ist äquivalent zur Gleichheit von (4.1) und (4.2) für alle beschränkten Martingale ζ und alle n , also zu

$$\mathbb{E}((\zeta(n) - \zeta(n-1))(\alpha(n) - \alpha(n-1))) = 0.$$

Im zweiten Faktor können nach Belieben \mathcal{F}_{n-1} -messbare Zufallsvariable addiert werden, deswegen gilt auch

$$\mathbb{E}((\zeta(n) - \zeta(n-1))(\alpha(n) - \mathbb{E}(\alpha(n)|\mathcal{F}_{n-1}))) = 0,$$

und, wegen der Definition der bedingten Erwartung

$$\mathbb{E}(\zeta(n)(\alpha(n) - \mathbb{E}(\alpha(n)|\mathcal{F}_{n-1}))) = 0.$$

für eine beliebige \mathcal{F}_n -messbare beschränkte Zufallsvariable η können wir für ζ das von η erzeugte Doob'sche Martingal einsetzen, und mit $\eta = \text{sig}(\alpha(n) - \mathbb{E}(\alpha(n)|\mathcal{F}_{n-1}))$ erhalten wir

$$\mathbb{E}(|\alpha(n) - \mathbb{E}(\alpha(n)|\mathcal{F}_{n-1})|) = 0,$$

also ist die Differenz fast sicher gleich 0. Damit ist die Natürlichkeit äquivalent zur Vorhersagbarkeit.

Um diese Ergebnisse auf den Fall der stetigen Zeit zu übertragen, nehmen wir an, dass die üblichen Bedingungen gelten, und definieren zuerst

Definition 4.5 *Der Prozess α heißt nichtfallend, wenn er integrierbar ist, mit Wahrscheinlichkeit 1 monoton nichtfallende Trajektorien hat und die Bedingung $\alpha(0) = 0$ erfüllt.*

Für einen nichtfallenden Prozess α und einen beschränkten messbaren Prozess ζ kann man das Integral

$$\int_0^t \zeta(s) d\alpha(s)$$

pfadweise (als Lebesgue-Stieltjes Integral) definieren. Falls ζ ein rechtsstetiges Martingal ist, dann gilt

$$\mathbb{E}\left(\int_0^t \zeta(s) d\alpha(s)\right) = \mathbb{E}(\zeta(t)\alpha(t)). \quad (4.3)$$

Dazu wählen wir eine Zerlegung $0 = t_0 < t_1 < \dots < t_n = t$ und setzen

$$\zeta_n(t) = \zeta(t_i) \text{ für } t_{i-1} < t \leq t_n.$$

Aus (4.1) folgt, dass (4.3) für die Prozesse ζ_n erfüllt ist, und wegen der Rechtsstetigkeit konvergiert ζ_n punktweise gegen ζ , und aus dem Satz von der dominierten Konvergenz folgt (4.3) im allgemeinen Fall. Jetzt können wir definieren:

Definition 4.6 Der nichtfallende Prozess α heißt natürlich, wenn für jedes rechtsstetige beschränkte Martingal ζ

$$\mathbb{E}\left(\int_0^t \zeta(s) d\alpha(s)\right) = \mathbb{E}(\zeta(t-0)\alpha(t)) \quad (4.4)$$

gilt.

Damit gilt

Satz 4.9 ξ sei ein rechtsstetiges Submartingal der Klassen (DL). Dann gibt es eine bis auf Ununterscheidbarkeit eindeutig bestimmte Zerlegung

$$\xi = \alpha + \eta$$

von ξ in einen rechtsstetigen nichtfallenden natürlichen Prozess α und ein rechtsstetiges Martingal η .

Zunächst wird die Eindeutigkeit bewiesen. Dazu seien $\alpha + \eta$ und $\alpha' + \eta'$ zwei Zerlegungen von ξ . Wir setzen

$$\gamma = \alpha - \alpha' = \eta' - \eta,$$

wählen ein beliebiges beschränktes rechtsstetiges Martingal ζ und eine Zerlegungsfolge t_{ni} und erhalten

$$\begin{aligned} \mathbb{E}(\zeta(t)(\alpha(t) - \alpha'(t))) &= \mathbb{E}\left(\int_0^t \zeta(s-0) d\alpha(s)\right) - \mathbb{E}\left(\int_0^t \zeta(s-0) d\alpha'(s)\right) = \\ &= \lim_n \mathbb{E}\left(\sum_i \zeta(t_{ni})(\alpha(t_{n,i+1}) - \alpha(t_{ni})) - \sum_i \zeta(t_{ni})(\alpha'(t_{n,i+1}) - \alpha'(t_{ni}))\right) = \\ &= \lim_n \mathbb{E}\left(\sum_i \zeta(t_{ni})(\gamma(t_{n,i+1}) - \gamma(t_{ni}))\right) = 0, \end{aligned}$$

weil γ ein Martingal ist. Wieder setzen wir für ζ eine rechtsstetige Version des Doob'schen Martingals ein, das von $\text{sig}(\alpha(t) - \alpha'(t))$ erzeugt wird, und erhalten somit, dass $\alpha(t)$ und $\alpha'(t)$ mit Wahrscheinlichkeit 1 übereinstimmen, und zusammen mit der Rechtsstetigkeit impliziert das die Ununterscheidbarkeit von α und α' .

Für die Existenz gehen wir von der Doob-Meyer Zerlegung einer Diskretisierung

$$\xi\left(\frac{it_0}{2^n}\right) = \alpha_n\left(\frac{it_0}{2^n}\right) + \eta_n\left(\frac{it_0}{2^n}\right)$$

aus und hoffen, dass die Folge $\alpha_n(t_0)$ für $n \rightarrow \infty$ (zumindest entlang einer Teilfolge) in einem geeigneten Sinn gegen ein $\alpha(t_0)$ konvergiert. Dann können wir η als das Doob'sche Martingal zu $\xi(t_0) - \alpha(t_0)$ ansetzen. Dieser geeignete Sinn wird durch die schwache Konvergenz geboten:

Definition 4.7 Die Folge ξ_n konvergiert schwach gegen ξ wenn für alle beschränkten Zufallsvariablen η die Beziehung

$$\lim_{n \rightarrow \infty} \mathbb{E}(\eta \xi_n) = \mathbb{E}(\eta \xi)$$

gilt.

Es gilt

Satz 4.10 (Dunford-Pettis) Wenn die Folge ξ_n gleichmäßig integrierbar ist, dann gibt es eine Teilfolge, die schwach konvergiert.

(Beweisskizze: \mathcal{G} sei die Sigmaalgebra, die von $(\xi_n, n \in \mathbb{N})$ erzeugt wird. Wegen

$$\mathbb{E}(\eta \xi_n) = \mathbb{E}((\eta | \mathcal{G}) \xi_n)$$

genügt es, die Konvergenz für \mathcal{G} -messbares η zu zeigen. \mathcal{G} wird von den abzählbar vielen Mengen $[\xi_n < q], n \in \mathbb{N}, q \in \mathbb{Q}$ erzeugt. Der von diesen Mengen erzeugte Ring ist ebenfalls abzählbar. Für jede der abzählbaren Mengen A aus dem Ring gibt es eine Teilfolge von ξ_n , für die $\mathbb{E}(\xi_n I_A)$ konvergiert. Mit Cantors Diagonalverfahren kann man eine Teilfolge finden, für die alle diese Erwartungswerte konvergieren. Wir werden o.B.d.A. annehmen, dass diese Teilfolge ξ_n selbst ist. Die gleichmäßige Integrierbarkeit impliziert, dass $\mathbb{E}(I_A \xi_n)$ für beliebiges $A \in \mathcal{G}$ gleichmäßig durch Mengen aus dem Ring approximiert werden kann. Daher existiert für jedes $A \in \mathcal{G}$ der Grenzwert $\mu(A) = \lim \mathbb{E}(I_A \xi_n)$ und die gleichmäßige Integrierbarkeit impliziert, dass μ absolutstetig bezüglich \mathbb{P} ist, und die Radon-Nikodym Dichte ist der gesuchte schwache Grenzwert.)

Wir wollen also zeigen, dass $\alpha_n(t_0)$ eine gleichmäßig integrierbare Folge ist. Dazu wählen wir $M > 0$ und setzen

$$\tau_n = \frac{t_0}{2^n} \inf \{i : \alpha_n(\frac{(i+1)t_0}{2^n}) > M\}$$

(Falls unter dem Infimum die leere Menge steht, setzen wir $\tau_n = t_0$). Wegen der Vorhersagbarkeit von α_n ist τ_n eine Stoppzeit, und wegen der Monotonie von α_n gilt $[\alpha_n(t_0) > M] = [\tau_n < t_0]$. Also

$$\begin{aligned} \mathbb{E}((\alpha_n(t_0) - M)_+) &= \mathbb{E}((\alpha_n(t_0) - M) I_{\tau_n < t_0}) \leq \mathbb{E}((\alpha_n(t_0) - \alpha_n(\tau_n)) I_{\tau_n < t_0}) = \\ \mathbb{E}(\alpha_n(t_0) - \alpha_n(\tau_n)) &= \mathbb{E}(\xi(t_0) - \xi(\tau_n)) - \mathbb{E}(\eta_n(t_0) - \eta_n(\tau_n)) = \mathbb{E}(\xi(t_0) - \xi(\tau_n)) = \\ &= \mathbb{E}((\xi(t_0) - \xi(\tau_n)) I_{\alpha_n(t_0) > M}). \end{aligned}$$

Wegen $\mathbb{E}(\alpha_n(t_0)) = \mathbb{E}(\xi(t_0))$ und der Ungleichung von Markov geht $\mathbb{P}(\alpha_n(t_0) > M)$ für $M \rightarrow \infty$ gleichmäßig (in n) gegen 0, und weil ξ von Klasse (DL) ist, geht auch die rechte Seite der Ungleichungskette gleichmäßig gegen 0, und somit ist die Folge $\alpha_n(t_0)$ gleichmäßig integrierbar.

Wir können jetzt also $\alpha(t_0)$ als den schwachen Grenzwert einer Teilfolge wählen, die wir o.B.d.A. mit $\alpha_n(t_0)$ identifizieren. Wir setzen $\eta(t_0) = \xi(t_0) -$

$\alpha(t_0)$ und für $0 \leq t \leq t_0$ $\eta(t)$ als eine rechtsstetige Version von $\mathbb{E}(\eta(t_0)|\mathcal{F}_t)$. $\eta_n(t_0)$ konvergiert schwach gegen $\eta(t_0)$, und für t von der Form $2^{-N}kt_0$ und beschränktes ζ gilt

$$\begin{aligned} \lim_n \mathbb{E}(\zeta \eta_n(t)) &= \lim_n \mathbb{E}(\zeta \mathbb{E}(\eta_n(t_0)|\mathcal{F}_t)) = \lim_n \mathbb{E}(\eta_n(t_0) \mathbb{E}(\zeta|\mathcal{F}_t)) = \\ &= \mathbb{E}(\eta(t_0) \mathbb{E}(\zeta|\mathcal{F}_t)) = \mathbb{E}(\zeta \mathbb{E}(\eta(t_0)|\mathcal{F}_t)) = \mathbb{E}(\zeta \eta(t)), \end{aligned}$$

also konvergiert $\eta_n(t)$ schwach gegen $\eta(t)$, und somit auch $\alpha_n(t)$ gegen $\alpha(t) = \xi(t) - \eta(t)$. Das gibt die Monotonie von α auf den dyadischen Vielfachen von t_0 ; durch die Rechtsstetigkeit erweitert sich die Monotonie auf ganz $[0, t_0]$.

Es bleibt die Natürlichkeit von α zu beweisen. Dazu gehen wir von der Vorhersagbarkeit von α_n aus, wählen ein rechtsstetiges Martingal ζ und erhalten aus der Martingaleigenschaft von $\eta_n = \xi - \alpha_n$ und $\eta = \xi - \alpha$:

$$\begin{aligned} \mathbb{E}(\zeta(t_0)\alpha(t_0)) &= \mathbb{E}\left(\sum_{i=0}^{2^n-1} \zeta\left(\frac{it_0}{2^n}\right)\left(\alpha_n\left(\frac{(i+1)t_0}{2^n}\right) - \alpha_n\left(\frac{it_0}{2^n}\right)\right)\right) = \\ &= \mathbb{E}\left(\sum_{i=0}^{2^n-1} \zeta\left(\frac{it_0}{2^n}\right)\left(\xi\left(\frac{(i+1)t_0}{2^n}\right) - \xi\left(\frac{it_0}{2^n}\right)\right)\right) = \\ &= \mathbb{E}\left(\sum_{i=0}^{2^n-1} \zeta\left(\frac{it_0}{2^n}\right)\left(\alpha\left(\frac{(i+1)t_0}{2^n}\right) - \alpha\left(\frac{it_0}{2^n}\right)\right)\right) = \\ &= \mathbb{E}\left(\int_0^{t_0} \zeta_n(s) d\alpha(s)\right), \end{aligned}$$

wobei wir

$$\zeta_n(t) = \zeta(2^{-n}t_0 \lfloor 2^n t/t_0 \rfloor)$$

gesetzt haben. Für $n \rightarrow \infty$ liefert der Satz der dominierten Konvergenz

$$\mathbb{E}(\zeta(t_0)\alpha(t_0)) = \mathbb{E}\left(\int_0^{t_0} \zeta(s-0) d\alpha(s)\right),$$

und α ist tatsächlich natürlich.

Kapitel 5

Stochastische Differentialrechnung

5.1 Motivation — der Brown'sche Geschwindigkeitsprozess

Wir betrachten die Geschwindigkeit eines Partikels in einem zähen Medium, das einer äußeren Kraft f unterliegt:

$$v'(t) = -kv(t) + f(t). \quad (5.1)$$

Die Brown'sche Bewegung sehr kleiner Partikel erklärt sich durch die zufälligen Stöße einzelner Moleküle des Mediums, die sich durch ihre thermische Energie bewegen. Es ist plausibel, die einzelnen Stöße als unabhängig voneinander anzunehmen; es sollten daher die einzelnen Funktionswerte $f(t)$ voneinander unabhängig und identisch verteilt sein; weiters sollte keine Richtung des Raumes bevorzugt sein, was bedeutet, dass wir $\mathbb{E}f(t) = 0$ fordern. Die klassische Differentialrechnung liefert uns für (5.1) die Lösung

$$v(t) = e^{-kt}v(0) + \int_0^t e^{k(s-t)} f(s) ds.$$

Mit unseren Annahmen lässt sich dieses Integral leider nicht berechnen, weil die Trajektorien eines Prozesses mit lauter unabhängigen Werten Trajektorien hat, die nicht messbar sind und daher auch nicht integrierbar. Wir können uns allerdings überlegen, was geschehen würde, wenn wir f doch integrieren könnten, und wir eine Stammfunktion $F(t) = \int_0^t f(s) ds$ dafür hätten. Dann könnten wir in der letzten Gleichung statt $f(t) dt$ $dF(t)$ schreiben. Die Stammfunktion F sollte unabhängige und stationäre Zuwächse haben (weil die einzelnen Zuwächse als Integrale über disjunkte Intervalle geschrieben werden können und die Werte der Integranden unabhängig und identisch verteilt sind); außerdem sollte F

stetige Pfade besitzen, weil es natürlich ist, anzunehmen, dass Integrale stetige Funktionen ihrer Obergrenze sind. Diese Eigenschaften — stetige Pfade und unabhängige stationäre Zuwächse — genügen aber schon, $F(t)$ festzulegen:

Satz 5.1 *Die einzigen Prozesse mit stationären unabhängigen Zuwächsen und stetigen Trajektorien sind von der Form*

$$a + bt + c\beta(t),$$

wobei β der Standard-Wienerprozess ist.

Damit wäre der Weg frei, die Differentialgleichung (5.1) in in der Form

$$dv(t) = -kv(t)dt + cd\beta(t)$$

auszudrücken und unsere Lösung von vorhin so zu schreiben:

$$v(t) = e^{-kt}v(0) + \int_0^t ce^{k(s-t)}d\beta(t),$$

Wenn wir so wie hier deterministische Funktionen als Integranden haben, kommen wir hier mit dem Verständnis des Integrals im Riemann-Stieltjes Sinn einigermaßen durch, aber wenn die zugrundeliegende Differentialgleichung nur um ein bisschen komplizierter ist (etwa $dv = kvdt + cvd\beta$, die in der Finanzmathematik beliebt ist), dann müssen wir auch stochastische Prozesse als Integranden zulassen. Betrachten wir also etwa den Fall, dass der Integrand der Wienerprozess β ist, d.h., wir wollen

$$\int_0^t \beta(t)d\beta(t)$$

bestimmen. Dazu wählen wir eine Zerlegung $0 = t_0^{(n)} < t_1^{(n)} < \dots < t_n^{(n)} = t$ und die beiden Riemannsummen

$$S_{1n} = \sum_{i=1}^n \beta(t_{i-1}^{(n)})(\beta(t_i^{(n)}) - \beta(t_{i-1}^{(n)}))$$

und

$$S_{2n} = \sum_{i=1}^n \beta(t_i^{(n)})(\beta(t_i^{(n)}) - \beta(t_{i-1}^{(n)})).$$

Für die Existenz des Riemann-Integrals müssten beide zum selben Grenzwert konvergieren, wenn $\delta_n \max_{1 \leq i \leq n} (t_i^{(n)} - t_{i-1}^{(n)})$ gegen 0 geht. Es gilt aber

$$S_{2n} - S_{1n} = \sum_{i=1}^n ((\beta(t_i^{(n)}) - \beta(t_{i-1}^{(n)}))$$

und damit

$$\mathbb{E}(S_{2n} - S_{1n}) = \sum_{i=1}^n (t_i^{(n)} - t_{i-1}^{(n)}) = t$$

und

$$\mathbb{V}(S_{2n} - S_{1n}) = \sum_{i=1}^n 2(t_i^{(n)} - t^{(n)})_i - 1)^2 \leq 2t\delta_n.$$

Die Differenz unserer beiden Riemannsummen konvergiert also im Quadratmittel gegen $t > 0$, also kann das Integral nicht im Riemannschem Sinn existieren.

Offensichtlich hängt das Ergebnis unserer Integration von der Wahl der Stützstellen ab, und deshalb müssen wir uns auf eine solche Wahl festlegen. Im wesentlichen gibt es dafür zwei Möglichkeiten, die Sinn machen: am linken Rand des jeweiligen Intervalls oder in der Mitte. Die zweite Möglichkeit, zum sogenannten Stratonovich-Integral führt, hat den Vorteil, dass viele Sätze der klassischen Infinitesimalrechnung unverändert gültig bleiben; die erste, die zum Itô-Integral führt, braucht in vielen Fällen zusätzliche Korrekturterme zu den klassischen Formeln, hat aber sehr viel angenehmere stochastische Struktur, weshalb wir unsere Untersuchungen darauf beschränken werden.

5.2 Das Itô-Integral

Zur Definition des stochastischen Integrals gehen wir einen Weg, der zur Einführung des Lebesgue-Integrals analog ist: Zuerst wird das Integral für eine Klasse von Funktionen definiert, für die es sich auf eine Summe reduziert, dann wird die Definition auf eine größere Klasse mittels Approximation durch einfache Funktionen erweitert. Obwohl die Prozesse, die wir im ersten Schritt betrachten, sehr viel besser der Idee einer Treppenfunktion entsprechen als die "Treppenfunktionen" der Maßtheorie, heißen Sie nicht so, sondern "einfache Prozesse":

Definition 5.1 Gegeben sei der Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbb{P})$ mit der Filtration (\mathcal{F}_t) . Der Prozess $(\xi(s), 0 \leq s \leq t)$ heißt einfacher Prozess, wenn er in der Form

$$\xi(t) = \sum_{i=1}^{n-1} \eta_i I_{[t_{i-1}, t_i)}(t) + \eta_n I_{[t_{n-1}, t_n]}(t),$$

mit $0 = t_0 < t_1 < \dots < t_n = t$ und η_i für alle $i \leq n$ bezüglich $\mathcal{F}_{t_{i-1}}$ messbar und quadratintegrierbar, darstellbar ist.

Für einfache Prozesse definieren wir das stochastische Integral wenig überraschend so:

Definition 5.2 Für den einfachen Prozess

$$\xi(t) = \sum_{i=1}^{n-1} \eta_i I_{[t_{i-1}, t_i)}(t) + \eta_n I_{[t_{n-1}, t_n]}(t)$$

und einen (\mathcal{F}_t) -Wienerprozess β (d.h., ein Wienerprozess, der bezüglich \mathcal{F}_t unabhängige Zuwächse besitzt) ist das Itô-Integral definiert als

$$\int_0^t \xi(s) d\beta(s) = \sum_{i=1}^n \eta_i (\beta(t_i) - \beta(t_{i-1})).$$

Für das Integral von einfachen Prozessen gilt

Satz 5.2 *Das Integral von einfachen Prozessen hat folgende Eigenschaften:*

1. *Linearität:*

$$\int_0^t (a\xi_1(s) + b\xi_2(s))d\beta(s) = a \int_0^t \xi_1(s)d\beta(s) + b \int_0^t \xi_2(s)d\beta(s).$$

2. $J(t) = \int_0^t \xi(s)d\beta(s)$ ist eine stetige Funktion von t .

3. $J(t)$ ist ein Martingal.

4. $J(t)$ ist bezüglich \mathcal{F}_t progressiv messbar.

5. *Die fundamentale Isometrie:*

$$\mathbb{E}(J(t)^2) = \mathbb{E}\left(\int_0^t \xi(s)^2 ds\right).$$

Die Linearität folgt direkt aus der Definition, die Stetigkeit aus der Stetigkeit des Wienerprozesses. Die Martingaleigenschaft muss nur für eine einzelne Stufe gezeigt werden (weil eine Summe von Martingalen wieder ein Martingal ist). Für $t_{i-1} \leq s_1 < s_2 < t_i$ ist

$$\mathbb{E}(J(s_2) - J(s_1) | \mathcal{F}_{s_1}) = \mathbb{E}(\eta_i(\beta(s_2) - \beta(s_1)) | \mathcal{F}_{s_1}) = \eta_i \mathbb{E}(\beta(s_2) - \beta(s_1) | \mathcal{F}_{s_1}) = 0,$$

weil η_i bezüglich \mathcal{F}_{s_1} -messbar ist und der Wienerprozess unabhängige Zuwächse hat.

Die fundamentale Isometrie folgt aus

$$J(t)^2 = \sum_{1 \leq i \leq n} \eta_i^2 (\beta(t_i) - \beta(t_{i-1}))^2 + 2 \sum_{1 \leq i < j \leq n} \eta_i \eta_j (\beta(t_i) - \beta(t_{i-1})) (\beta(t_j) - \beta(t_{j-1})).$$

In der letzten Summe ist das Produkt der ersten drei Faktoren bezüglich \mathcal{F}_{j-1} messbar, der letzte ist davon unabhängig, also hat das Produkt den Erwartungswert 0. In der ersten Summe sind ebenfalls beide Faktoren unabhängig, und so ergibt sich

$$\mathbb{E}(J(t)^2) = \sum_{i=1}^n \mathbb{E}(\eta_i^2) (t_i - t_{i-1}),$$

aber das ist nichts anderes als

$$\mathbb{E}\left(\int_0^t \xi(t)^2 dt\right).$$

Die fundamentale Isometrie legt nahe, die Approximation im Sinne des Quadratischen Mittels zur Definition des Integrals im allgemeinen Fall zu verwenden.

Wir definieren also:

Definition 5.3 $\xi(t)$ sei bezüglich \mathcal{F}_t progressiv messbar und

$$\|\xi\|_2^2 = \int_0^t \mathbb{E}(\xi(s)^2) ds < \infty.$$

Weiters sei ξ_n eine Folge von einfachen Prozessen, die gegen ξ im Sinne dieser Norm konvergieren. Dann setzen wir

$$\int_0^t \xi(s) d\beta(s) = \lim \int_0^t \xi_n(s) d\beta(s).$$

Dabei ist der Limes im quadratischen Mittel zu verstehen.

Die progressiv messbaren Prozesse mit endlicher Norm sind genau diejenigen, die im Sinne der Norm durch einfache Prozesse approximiert werden können. Die einfachen Prozesse erfüllen ja diese beiden Bedingungen, daher müssen Sie auch für ihren Limes gelten. Wenn umgekehrt ξ progressiv messbar ist und zusätzlich beschränkt und stetig, dann konvergiert

$$\xi_n(s) = \xi(\lfloor ns \rfloor / n)$$

punktweise gegen ξ , und wegen des Satzes von der dominierten Konvergenz auch im Sinne der Norm.

Ein beschränktes ξ lässt sich ebenso durch die stetigen und beschränkten Prozesse

$$\xi_n(s) = n \int_{s-1/n}^s \xi(u) du$$

approximieren.

Schließlich kann ein allgemeines ξ durch die beschränkten

$$x_n(s) = \min(\max(\xi(s), n) - n)$$

approximieren.

Für das Itô-Integral eines progressiv messbaren Prozesses gelten dieselben Eigenschaften wie für einfache Funktionen:

Satz 5.3 Für das Itô-Integral gelten die Eigenschaften 1.–5. aus Satz 5.2.

Bis auf die Stetigkeit der Trajektorien folgen alle diese Eigenschaften durch Grenzübergang aus den entsprechenden Eigenschaften für die elementaren Prozesse.

Für die Stetigkeit wählen wir eine Folge ξ_n von einfachen Funktionen mit

$$\|\xi_n - \xi\|_2 < 2^{-2n-1}.$$

Daraus folgt

$$\|\xi_n - \xi_{n+1}\|_2 < 2^{-2n}.$$

Die Prozesse

$$J_n(t) = \int_0^t \xi_n(s) ds$$

sind Martingale, und deshalb auch $J_n - J_{n+1}$.

Die Maximalungleichung von Doob ergibt

$$\mathbb{P}\left(\sup_{0 \leq s \leq t} |J_n(s) - J_{n+1}(s)| > 2^{-n}\right) \leq 2^{2n} \mathbb{E}((J_n(t) - J_{n+1}(t))^2) < 2^{-2n},$$

und das Borel-Cantelli Lemma liefert, dass mit Wahrscheinlichkeit eins für hinreichend großes n

$$\|J_n - J_{n+1}\|_\infty = \sup_{0 \leq s \leq t} |J_n(s) - J_{n+1}(s)| \leq 2^{-n}.$$

Mit Wahrscheinlichkeit 1 ist also J_n eine Cauchyfolge bezüglich der gleichmäßigen Konvergenz, und weil die Prozesse J_n stetig sind, gilt das auch für die Grenzfunktion.

Als Beispiel berechnen wir das Integral des Wienerprozesses $\int_0^t \beta(s) d\beta(s)$:

Wir setzen

$$\beta_n(s) = \beta\left(\frac{t(i-1)}{2^n}\right) \text{ für } \frac{(i-1)t}{2^n} \leq s \leq \frac{it}{2^n}.$$

Diese Prozesse konvergieren im Quadratmittel gegen β und sind einfach. Das Integral von β_n ist

$$\begin{aligned} & \sum_{i=1}^{2^n} \beta\left(\frac{(i-1)t}{2^n}\right) (\beta\left(\frac{it}{2^n}\right) - \beta\left(\frac{(i-1)t}{2^n}\right)) = \\ & \frac{1}{2} \left(\sum_{i=1}^{2^n} (\beta\left(\frac{it}{2^n}\right))^2 - \beta\left(\frac{(i-1)t}{2^n}\right)^2 \right) - \sum_{i=1}^{2^n} (\beta\left(\frac{it}{2^n}\right) - \beta\left(\frac{(i-1)t}{2^n}\right))^2 = \\ & \frac{1}{2} (\beta(t)^2 - \sum_{i=1}^{2^n} (\beta\left(\frac{it}{2^n}\right) - \beta\left(\frac{(i-1)t}{2^n}\right))^2). \end{aligned}$$

Die letzte Summe haben wir schon vorher betrachtet, sie konvergiert gegen t . Wir haben also

$$\int_0^t \beta(t) d\beta(t) = \frac{\beta(t)^2 - t}{2}.$$

5.3 Stochastische Differentiale und die Itô-Formel

Wir können jetzt über stochastische Differentiale sprechen:

Definition 5.4 Der Prozess ξ mit Werten in \mathbb{R}^d hat das stochastische Differential

$$d\xi(t) = \mu(t)dt + \sigma(t)d\beta(t),$$

wenn ξ in der Form

$$\xi_i(t) = \xi_i(0) + \int_0^t \mu_i(s) ds + \sum_{j=1}^k \sigma_{ij}(s) d\beta_j(s)$$

darstellen lässt. Dabei ist μ ein d -dimensionaler Vektor und σ eine $(d \times k)$ -Matrix von quadratintegrierbaren Prozessen und β ein k -dimensionaler Wienerprozess (also ein Vektor, dessen Komponenten unabhängige Wienerprozesse sind).

Die Itô-Formel besagt, dass Funktionen eines Prozesses mit einem stochastischen Integral ebenfalls ein stochastisches Differential besitzen:

Satz 5.4 Der (d -dimensionale) Prozess ξ besitze das stochastische Differential

$$d\xi(t) = \mu(t)dt + \sigma(t)td\beta(t),$$

die Funktion $f(t, x) : \mathbb{R}^{1+d} \rightarrow \mathbb{R}$ habe stetige und beschränkte Ableitungen bis zur zweiten Ordnung. Dann hat auch der Prozess

$$\eta(t) = f(t, \xi(t))$$

ein stochastisches Differential, und es gilt

$$d\eta(t) = \frac{\partial}{\partial t} f(t, \xi(t))dt + \sum_{i=1}^d \frac{\partial}{\partial x_i} f(t, \xi(t))d\xi_i(t) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \frac{\partial^2}{\partial x_i \partial x_j} f(t, \xi(t))d\xi_i(t)d\xi_j(t).$$

Dabei werden die Produkte wie üblich ausmultipliziert und nach den Regeln

$$dt^2 = dt d\beta_i(t) = d\beta_i(t) d\beta_j(t) = 0 \quad (i \neq j),$$

$$(d\beta_i(t))^2 = dt$$

vereinfacht.

Wir beweisen den Satz für den Fall $d = 1$. Es genügt, den Fall einfacher Prozesse μ und σ zu betrachten, weil für

$$\eta_i(t) = f(t, \xi_i(t)) \quad (i = 1, 2)$$

und

$$\left| \frac{\partial f}{\partial x} \right| \leq M$$

die Ungleichung

$$|\eta_1(t) - \eta_2(t)| \leq M |\xi_1(t) - \xi_2(t)|$$

folgt. Einen einfachen Prozess kann man in die einzelnen Stufen zerlegen, also genügt es, Prozesse der Form

$$\xi(t) = \xi(0) + \mu t + \sigma \beta(t)$$

mit \mathcal{F}_0 -messbaren Zufallsvariablen ξ_0 , μ und σ .

Dafür gilt:

$$f(t, \xi(t)) - f(0, \xi(0)) = \sum_{i=1}^{2^n} (f(\frac{ti}{2^n}, \xi(\frac{ti}{2^n})) - f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-1)}{2^n}))) =$$

$$\sum_{i=1}^{2^n} (f(\frac{ti}{2^n}, \xi(\frac{ti}{2^n})) - f(\frac{t(i-1)}{2^n}, \xi(\frac{ti}{2^n}))) + \sum_{i=1}^{2^n} (f(\frac{t(i-1)}{2^n}, \xi(\frac{ti}{2^n})) - f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-1)}{2^n}))).$$

Die erste Summe kann nach dem Zwischenwertsatz als

$$\sum_{i=1}^{2^n} \frac{\partial}{\partial t} f(\frac{t(i-\theta_i)}{2^n}, \xi(\frac{ti}{2^n})) \frac{t}{2^n}$$

geschrieben werden, und diese konvergiert im Quadratmittel gegen das Integral

$$\int_0^t \frac{\partial}{\partial t} f(s, \xi(s)) ds.$$

Den zweiten Teil schreiben wir als

$$\sum_{i=1}^{2^n} \frac{\partial}{\partial x} f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-1)}{2^n})) (\xi(\frac{ti}{2^n}) - \xi(\frac{t(i-1)}{2^n})) +$$

$$\frac{1}{2} \sum_{i=1}^{2^n} \frac{\partial^2}{\partial x^2} f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-\theta_i)}{2^n})) (\xi(\frac{ti}{2^n}) - \xi(\frac{t(i-1)}{2^n}))^2 = S_1 + S_2.$$

Für den ersten Teil erhalten wir

$$S_1 = \sum_{i=1}^{2^n} \frac{\partial}{\partial x} f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-1)}{2^n})) \frac{\mu i}{2^n} +$$

$$\sum_{i=1}^{2^n} \frac{\partial}{\partial x} f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-1)}{2^n})) \sigma (\beta(\frac{ti}{2^n}) - \beta(\frac{t(i-1)}{2^n})),$$

und das konvergiert gegen

$$\int_0^t \frac{\partial}{\partial x} f(s, \xi(s)) \mu ds + \int_0^t \frac{\partial}{\partial x} f(s, \xi(s)) \sigma d\beta(s) = \int_0^t \frac{\partial}{\partial x} f(s, \xi(s)) d\xi(s).$$

Für S_2 ergibt sich:

$$S_2 = \frac{1}{2} \sum_{i=1}^{2^n} \frac{\partial^2}{\partial x^2} f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-\theta_i)}{2^n})) \frac{\mu^2}{2^{2n}} +$$

$$\sum_{i=1}^{2^n} \frac{\partial^2}{\partial x^2} f(\frac{t(i-1)}{2^n}, \xi(\frac{t(i-\theta_i)}{2^n})) \frac{\mu \sigma}{2^n} (\beta(\frac{ti}{2^n}) - \beta(\frac{t(i-1)}{2^n})) +$$

$$\frac{1}{2} \sum_{i=1}^{2^n} \frac{\partial^2}{\partial x^2} f\left(\frac{t(i-1)}{2^n}, \xi\left(\frac{t(i-\theta_i)}{2^n}\right)\right) \sigma^2\left(\beta\left(\frac{ti}{2^n}\right) - \beta\left(\frac{t(i-1)}{2^n}\right)\right)^2.$$

Die ersten beiden Summen konvergieren gegen 0, die letzte können wir für $m < n$ so umformen:

$$\begin{aligned} & \frac{1}{2} \sum_{j=1}^{2^m} \sum_{i=(j-1)2^{n-m}+1}^{j2^{n-m}} \frac{\partial^2}{\partial x^2} f\left(\frac{t(i-1)}{2^n}, \xi\left(\frac{t(i-\theta_i)}{2^n}\right)\right) \sigma^2\left(\beta\left(\frac{ti}{2^n}\right) - \beta\left(\frac{t(i-1)}{2^n}\right)\right)^2 = \\ & \frac{1}{2} \sum_{j=1}^{2^m} \sum_{i=(j-1)2^{n-m}+1}^{j2^{n-m}} \frac{\partial^2}{\partial x^2} f\left(\frac{t(j-1)}{2^m}, \xi\left(\frac{t(j-1)}{2^m}\right)\right) \sigma^2\left(\beta\left(\frac{ti}{2^n}\right) - \beta\left(\frac{t(i-1)}{2^n}\right)\right)^2 + \\ & \frac{1}{2} \sum_{j=1}^{2^m} \sum_{i=(j-1)2^{n-m}+1}^{j2^{n-m}} \left(\frac{\partial^2}{\partial x^2} f\left(\frac{t(i-1)}{2^n}, \xi\left(\frac{t(i-\theta_i)}{2^n}\right)\right) - \frac{\partial^2}{\partial x^2} f\left(\frac{t(j-1)}{2^m}, \xi\left(\frac{t(j-1)}{2^m}\right)\right) \right) \sigma^2\left(\beta\left(\frac{ti}{2^n}\right) - \beta\left(\frac{t(i-1)}{2^n}\right)\right)^2. \end{aligned}$$

Wenn m groß genug gewählt wird, kann die Norm der zweiten Summe beliebig klein gemacht werden, die erste konvergiert für $n \rightarrow \infty$ gegen

$$\frac{1}{2} \sum_{j=1}^{2^m} \sum_{i=(j-1)2^{n-m}+1}^{j2^{n-m}} \frac{\partial^2}{\partial x^2} f\left(\frac{t(j-1)}{2^m}, \xi\left(\frac{t(j-1)}{2^m}\right)\right) \sigma^2 \frac{1}{2^m}.$$

Dieser Ausdruck unterscheidet sich wieder für hinreichend großes m um beliebig wenig von

$$\frac{1}{2} \int_0^t \frac{\partial^2}{\partial x^2} f(s, \xi(s)) \sigma^2 ds = \frac{1}{2} \int_0^t \frac{\partial^2}{\partial x^2} f(s, \xi(s)) (d\xi(s))^2.$$

5.4 Stochastische Differentialgleichungen

Wir kehren zurück zu der Frage, die wir Eingangs zur Motivation gestellt haben, in etwas allgemeinerer Form:

Definition 5.5 *Der adaptierte Prozess $\xi \in \mathbb{R}^d$ erfüllt in $[0, t_0]$ die stochastische Differentialgleichung*

$$d\xi(t) = m(t, \xi(t))dt + s(t, \xi(t))d\beta(t), \quad (5.2)$$

mit den Koeffizienten $m : \mathbb{R}^{1+d} \rightarrow \mathbb{R}^d$ und $s : \mathbb{R}^{1+d} \rightarrow \mathbb{R}^{d \times k}$, dem k -dimensionalen Wienerprozess β und dem Anfangswert $\xi(0)$, wenn für alle $t \leq t_0$ die Integralgleichung

$$xi(t) = \int_0^t m(u, \xi(u))du + s(u, \xi(u))d\beta(u)$$

erfüllt ist. μ heißt die infinitesimale Drift, ss^T die infinitesimale Kovarianzmatrix des Prozesses.

Wie für die gewöhnlichen Differentialgleichungen gibt es einen Existenz- und Eindeigkeitssatz:

Satz 5.5 ξ_0 sei quadratintegrierbar, und die Koeffizienten mögen die folgenden Bedingungen erfüllen:

- Die Lipschitzbedingung: für alle $t \leq t_0$ und alle $x, y \in \mathbb{R}^d$ gelte

$$\|m(t, x) - m(t, y)\| + \|s(t, x) - s(t, y)\| \leq K\|x - y\|$$

und

- Die Wachstumsbedingung: für alle $t \leq t_0$ und alle $x \in \mathbb{R}^d$ gelte

$$\|m(t, x)\| + \|s(t, x)\| \leq K(1 + \|x\|).$$

Dann hat die stochastische Differentialgleichung (5.2) im Intervall $[0, t_0]$ eine (bis auf Ununterscheidbarkeit) eindeutig bestimmte Lösung.

Wir betrachten der Einfachheit halber den Fall $d = k = 1$. Zum Beweis der Existenz beginnen wir mit

$$\xi_0(t) = \xi(0)$$

und setzen für $n > 0$

$$\xi_n(t) = \xi(0) + \int_0^t m(u, \xi_{n-1}(u))du + \int_0^t s(u, \xi_{n-1}(u))d\beta(u).$$

Mit Hilfe der Cauchy-Schwarz Ungleichung und der fundamentalen Isometrie stellen wir fest, dass

$$\begin{aligned} \mathbb{E}((\xi_1(t) - \xi_0(t))^2) &= \mathbb{E}\left(\left(\int_0^t m(u, \xi(0))du + \int_0^t s(u, \xi(0))d\beta(u)\right)^2\right) \leq \\ &2\left(\mathbb{E}\left(\left(\int_0^t m(u, \xi(0))du\right)^2\right) + \mathbb{E}\left(\left(\int_0^t s(u, \xi(0))d\beta(u)\right)^2\right)\right) \leq \\ &2\left(\mathbb{E}\left(t\int_0^t m(u, \xi(0))^2du\right) + \mathbb{E}\left(\int_0^t s(u, \xi(0))^2du\right)\right) \leq 2(t^2 + t)\mathbb{E}((1 + |\xi(0)|)^2) \leq \\ &2(t_0^2 + t_0)\mathbb{E}((1 + |\xi(0)|)^2) = M. \end{aligned}$$

Wir setzen

$$g_n(t) = \mathbb{E}((\xi_{n+1}(t) - \xi_n(t))^2),$$

und erhalten, wieder mit Cauchy-Schwarz und fundamentaler Isometrie:

$$\begin{aligned} g_{n+1}(t) &= \mathbb{E}\left(\left(\int_0^t (m(u, \xi_{n+1}(u)) - m(u, \xi_n(u)))du + \int_0^t (s(u, \xi_{n+1}(u)) - s(u, \xi_n(u)))d\beta(u)\right)^2\right) \leq \\ &2(t_0 + 1)K^2 \int_0^t g_n(u)du. \end{aligned}$$

Aus $g_0(t) \leq M$ erhalten wir durch vollständige Induktion

$$g_n(t) \leq \frac{M(2t(t_0 + 1))^n}{n!}.$$

Daraus folgt, dass ξ_n eine Cauchyfolge bezüglich $\|\cdot\|_2$ bildet, also gibt es den Grenzwert ξ . In der Rekursionsgleichung können wir jetzt n gegen unendlich gehen lassen und erhalten so, dass ξ die Gleichung (5.2) erfüllt.

Für die Eindeutigkeit wählen wir zwei quadratintegrierbare Lösungen ξ und η mit unterschiedlichen Anfangsbedingungen $\xi(0)$ und $\eta(0)$. Wir setzen

$$g(t) = \mathbb{E}((\eta(t) - \xi(t))^2)$$

und erhalten

$$\begin{aligned} g(t) &= \mathbb{E}((\eta(0) - \xi(0) + \int_0^t (m(u, \eta(u)) - m(u, \xi(u))) du + \int_0^t (s(u, \eta(u)) - s(u, \xi(u))) d\beta(u))^2) \leq \\ &3(\mathbb{E}((\eta(0) - \xi(0))^2) + (1 + t_0)K^2 \int_0^t g(u) du). \end{aligned}$$

Wegen der Quadratintegrierbarkeit von ξ und η ist $g(t)$ auf $[0, t_0]$ beschränkt, sagen wir $g(t) \leq B$.

Daraus erhalten wir per Induktionem mit $L = 3(1 + t_0)K^2$ die Ungleichungen

$$g(t) \leq 3\mathbb{E}((\eta(0) - \xi(0))^2) \sum_{i=0}^{n-1} \frac{(Lt)^i}{i!} + B \frac{(Lt)^n}{n!},$$

und für $n \rightarrow \infty$

$$g(t) \leq 3\mathbb{E}((\eta(0) - \xi(0))^2) e^{Lt}.$$

Die Funktionswerte hängen also stetig von den Anfangswerten ab, und speziell für $\eta(0) = \xi(0)$ erhalten wir $g(t) = 0$, also $\eta(t) = \xi(t)$ mit Wahrscheinlichkeit 1.

Satz 5.6 *Die Lösung ξ der stochastischen Differentialgleichung (5.2) ist ein Markovprozess. Falls die Koeffizienten m und s nicht von t abhängen, ist ξ ein homogener Markovprozess.*

Offensichtlich erfüllt unter der Bedingung $\xi(t_0) = x$ der Prozess $\xi(t)$ für $t \geq t_0$ die Gleichung (5.2) mit dem Anfangswert x . Unter dieser Bedingung ist die Lösung der Gleichung unabhängig von \mathcal{F}_{t_0} (das folgt aus der Tatsache, dass der Wienerprozess unabhängige Zuwächse hat, etwa indem man die einzelnen Approximationsfunktionen aus dem Existenzbeweis betrachtet). Also ist ξ ein Markovprozess. Falls die Koeffizienten nicht von t abhängen, ist die Differentialgleichung für jedes t_0 dieselbe, und wegen der Eindeutigkeit der Lösung hat ist die bedingte Verteilung zum Zeitpunkt $t_0 + h$ von t_0 unabhängig von t_0 , also der Markovprozess homogen.

Wir betrachten jetzt den homogenen Fall — unsere früheren Betrachtungen verlocken uns dazu, nach dem infinitesimalen Operator dieses Prozesses zu fragen. f sei eine Funktion mit kompaktem Träger und beschränkten und gleichmäßig stetigen Ableitungen bis zur zweiten Ordnung und einen Prozess ξ , der die Differentialgleichung (5.2) mit der Anfangsbedingung $\xi(0) = x$ erfüllt. dann ist

$$P^t f(x) = \mathbb{E}(f(\xi(t))) = f(x) + \mathbb{E}\left(\int_0^t \sum_{i=1}^d \frac{\partial}{\partial x_i} f(\xi(u)) (m_i(\xi(u))) du + \sum_{j=1}^k s_{ij} d\beta_j(s)\right) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^k \sum_{l=1}^d \frac{\partial^2}{\partial x_i \partial x_k} s_{ij}(\xi(t)) s_{kj}(\xi(t)) du.$$

Die stochastischen Integrale haben als Martingale Erwartungswert 0; die restlichen Integrale konvergieren, wenn sie durch t dividiert werden, gegen den Wert des Integranden bei 0. Das führt zu

$$Qf(x) = \sum_{i=1}^d m_i(x) \frac{\partial}{\partial x_i} f(x) + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d b_{ij}(x) \frac{\partial^2}{\partial x_i \partial x_j} f(x).$$

Hier haben wir (in Matrixform)

$$b(x) = s(x)s(x)^T$$

gesetzt.

Wir stellen also fest, dass der infinitesimale Operator ein Differentialoperator ist (genauer der Abschluss eines solchen Operators).

Index

- abgeschlossener Operator, 42
- adaptiert, 13
- Äquivalenz von Prozessen, 5
- aperiodisch, 18

- bedingte Erwartung, 13
- bedingte Wahrscheinlichkeit, 13

- Chapman, 17, 23

- Definitionsbereich des infinitesimalen Erzeugers, 39
- Differentialgleichung
 - stochastische, 64
- Differentialgleichungen
 - Kolmogorovsche, 26, 40
- Drift
 - infinitesimale, 65
- Durchquerung, 9

- einfacher Prozess, 58
- eingebettete Markovkette, 26
- Erneuerungsgleichung, 20
- Erwartung
 - bedingte, 13
- Erzeuger
 - infinitesimaler, 25
- Existenzsatz von Kolmogorov, 5
- Explosion, 28

- Famile
 - konsistente, 5
- Feld
 - stochastisches, 4
- Filtration, 13
 - natürliche, 13
- fundamentale Isometrie, 59

- Gaussprozess, 12
- Geburts- und Todesprozess, 27
- Geburtsprozess, 27
- Gleichung
 - Chapman-Kolmogorov, 17, 23

- Halbgruppe, 39
- Hille, 43

- infinitesimale Drift, 65
- infinitesimale Kovarianzmatrix, 65
- infinitesimale Parameter, 25
- infinitesimaler Erzeuger, 25
 - Definitionsbereich, 39
- infinitesimaler Operator, 39
- instabiler Zustand, 26
- irreduzibel, 18
- Isometrie
 - fundamentale, 59
- Itô-Formel, 62
- Itô-Integral, 60

- Klassen von Zuständen, 18
- Klasseneigenschaften, 18, 19
- Kolmogorov
 - Existenzsatz, 5
 - Stetigkeitskriterium, 7
- Kolmogorovsche Differentialgleichungen, 26, 40
- kommunizieren, 17, 23
- konservativ, 26
- konsistente Familie, 5
- kontrahierender Operator, 39
- Kontraktion, 39
- Kovarianzfunktion, 11
- Kovarianzmatrix
 - infinitesimale, 65

- lokaler Operator, 44
- Lokalität, 44
- Markovkette, 16
 - eingebettete, 26
 - in diskreter Zeit, 16
 - in stetiger Zeit, 16
 - irreduzible, 18
 - konservative, 26
 - reguläre, 28
- Markovketten
 - in stetiger Zeit, 23
- Markovprozess, 11, 16
 - homogener, 16
- Martingal, 12, 46
- Matrix
 - infinitesimale, 25
- Maximumprinzip, 41
- Menge
 - separable, 10
- Minimallösung, 30
- Mittelwertfunktion, 11
- Nachfolger, 17
- natürlicher Prozess, 51
- Nullrekurrenz, 19
- Operator
 - abgeschlossener, 42
 - infinitesimaler, 39
 - kontrahierender, 39
 - lokaler, 44
- Parameter
 - infinitesimale, 25
- Parameterraum, 4
- Periode, 18, 23
- Pfad, 6
- Phasenraum, 5
- positive Rekurrenz, 19
- Prozess
 - adaptierter, 13
 - einfacher, 58
 - mit unabhängigen Zuwächsen, 11
 - natürlicher, 51
 - stationärer, 11
 - stochastischer, 4
 - Vergangenheit eines, 13
- Raum der starken Stetigkeit, 39
- reguläre Markovkette, 28
- Rekurrenz, 18, 23
 - positive, 19
- Rekurrenzklassen, 18
- Rekurrenzzeit, 18
- Resolvente, 41
- Rückkehrzeit, 18
- Rückwärtsgleichung, 26, 40
- Satz
 - Hille-Yosida, 43
- schwach stationär, 11
- Semimartingal, 46
- Separabilität, 10
- separable Menge, 10
- Spiegelungsprinzip, 35
- Standard-Wienerprozess, 8
- stark stationär, 11
- starke Stetigkeit, 39
- stationär, 11
 - schwach, 11
 - stark, 11
- stationäre Verteilung, 20
- Stetigkeit
 - der Trajektorien, 7
 - im Quadratmittel, 6
 - stochastische, 6
- Stetigkeitskriterium
 - von Kolmogorov, 7
- stochastische Differentialgleichung, 64
- stochastischer Prozess, 4
- stochastisches Differential, 62
- stochastisches Feld, 4
- Stopzeit, 14
- Submartingal, 46
- Supermartingal, 46
- Todesprozess, 27
- Trajektorie, 6
- Trajektoreigenschaften, 7
- Transienz, 18, 23
- Übergangsfunktion, 16

Übergangsmatrix, 17, 23
Übergangsoperatoren, 39
Übergangswahrscheinlichkeiten, 16
Übergangszeit, 18
übliche Bedingungen, 49
ununterscheidbar, 5
upcrossing, 9
usual conditions, 49

verbunden, 17
Vergangenheit
 bezüglich einer Stoppzeit, 15
Version, 5
Verteilung
 stationäre, 20
Vorwärtsgleichung, 26, 40

Wahrscheinlichkeit
 bedingte, 13
Wienerprozess, 6
 Standard-, 8

Yosida, 43

Zustände
 kommunizierende, 17
Zustand
 instabiler, 26
Zustandsraum, 5