

Hans Daduna
Institut für
Mathematische Stochastik
der Universität Hamburg

Wintersemester 2004/05

Stochastische Methoden des Operations Research

— Markov Ketten und Erneuerungstheorie —

Dieser Text diente als Arbeitsunterlage für eine 2-stündige Vorlesung:
Grundzüge der Stochastik – Stochastische Methoden des Operations Research,
gehalten zuletzt im Wintersemester 1996/97
am Fachbereich Mathematik der Universität Hamburg.
Die ursprünglich handgeschriebenen Unterlagen wurden von
Simona Dittrich und Daniel Mahnke
als $\text{T}_{\text{E}}\text{X}/\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ -Dokument erstellt.

Inhaltsverzeichnis

| | | |
|----------|---|-----------|
| 1 | Einführende Beispiele | 2 |
| 2 | Grundbegriffe für stochastische Prozesse | 4 |
| 3 | Markov–Ketten | 6 |
| 4 | Asymptotisches Verhalten, Stabilität und stationäre Verteilungen von Markov–Ketten | 37 |
| 5 | Erneuerungstheorie | 57 |
| | Literatur | 82 |

1 Einführende Beispiele

1.1 Beispiel (Erneuerungs- und Ersatzprobleme)

- a) Ein bestimmter Geräteteil, der von Zeit zu Zeit ausfällt, sei für die Systemfunktion derart wichtig, daß er sofort durch ein identisches Exemplar ersetzt werde.

Interessierende Kenngrößen sind dann z. B. die Verteilung der Anzahl der Erneuerungen in einem gewissen Zeitraum oder die Verteilung der Zeit, für die eine vorgegebene Zahl von Geräteteilen ausreicht, um das System arbeiten zu lassen.

- b) Dabei wird fast ausnahmslos davon ausgegangen, daß die Funktionszeitverteilungen der Geräteteile nicht nur identisch, sondern auch unabhängig sind, d. h. nach jeder „Erneuerung“ beginnt der Teil, der untersucht wird, sich stochastisch identisch zu seinen Vorgängern und Nachfolgern zu entwickeln. Durch diese Annahme und die daraus entstehenden Folgerungen werden die Ergebnisse einer „Erneuerungstheorie“ direkte Grundlage einer Theorie rekurrenter Ereignisse sein: dabei handelt es sich um Ereignisse im Ablauf eines komplexen Systems, nach denen die stochastische Entwicklung des Systems sich unabhängig von der Vorgeschichte fortsetzt bis zum nächsten Eintritt desselben Ereignisses etc. Die Zwischenzeiten zwischen dem Eintreten des „rekurrenten Ereignisses“ sind dann gerade das Analogon zu den Funktionszeiten der obigen Geräteteile.
- c) Die Einführung von Kosten für Ausfallzeiten, die mögliche Inspektion arbeitender Geräte, mögliche vorzeitige Erneuerung lange arbeitender Geräteteile führen zu komplizierten, aber praktikablen Ergebnissen für optimale Inspektions- und Ersatzstrategien. □

1.2 Beispiel (Lagerhaltung)

- a) In einem Lager werde ein Vorrat eines einzigen Gutes (in Stücken) gelagert. Jeden Tag treffe eine nichtnegative zufällige Bestellung ein. Diese wird soweit möglich erfüllt, und danach besteht die Möglichkeit, eine Entscheidung für eine Lageraufstockung bis zum nächsten Tag zu treffen (Bestellung).

An Kosten entstehen z. B.: Lagerkosten, Fehlmengenkosten, Bestellkosten; die Einnahmen erhält man durch den Verkauf der gelagerten Güter.

Wie soll man Bestellentscheidungen treffen?

- b) Verallgemeinerungen: mehrere Güter, Zwischenlager, Fehlmengen werden für die nächste Auslieferung vorgemerkt. □

1.3 Beispiel (Warteschlangen, Bedienungssysteme)

- a) An einer Reparaturstation kommen defekte Stücke einzeln in zufälligen Abständen an und reihen sich entsprechend ihrer Ankunftsfolge in eine Warteschlange ein, die maximal N Plätze bereitstellt. Das Stück an der Spitze der Schlange wird repariert, die Reparaturzeit sei zufällig mit gleicher Verteilung für alle Stücke. Eintreffende Stücke, die alle Warteplätze besetzt vorfinden, gehen verloren (werden abgewiesen).

Von Interesse sind die stochastische Entwicklung der Schlangenlänge über die Zeit, der Durchsatz durch das System (System-orientierte Leistungsmaße) bzw. Verweilzeit im System (Kunden-orientierte Leistungsmaße).

- b) Verallgemeinerungen sind: unterschiedliche Kundentypen mit unterschiedlichen Bedienungszeiten, Gruppenankünfte, mehrere Bedienungsgeräte, andere Bedienungsdisziplinen, unendliche Wartezimmer.
- c) Flexible Manufaktur-Systeme (FMS)

Eine Fabrik bestehe aus Teileinheiten, die einzelne Fertigungsabschnitte für gewisse Aufträge übernehmen. Ein eintreffender Auftrag hat mit gewissen Wahrscheinlichkeiten eine Reihe dieser Teileinheiten zu durchlaufen, wobei die Ablaufsteuerung sowohl zentral als auch dezentral möglich ist.

Werden die Teileinheiten durch Bedienungssysteme nach a) oder b) modelliert, hat man als Gesamtmodell für die Fabrik ein Netzwerk an Warteschlangen.

- d) Rechnernetze werden analog modelliert, wobei häufig speziell für die Bedürfnisse der Informatik entwickelte Modelle (Bedienungsdisziplinen) für die Netzknoten eingesetzt werden. \square

Die angeführten Beispiele aus den Bereichen Wirtschaft und Technik sind typische Systeme, bei deren Untersuchung (Modellierung, Analyse, Planung) stochastische Methoden eingesetzt werden müssen. Alle diese Systeme sind zufallsbeeinflusst, und erfahrungsgemäß sind diese Zufallseinflüsse von nicht vernachlässigbarer Bedeutung für die zeitliche Entwicklung. Ein allgemeines mathematisches Modell für zeitlich ablaufende zufallsbeeinflusste Systeme ist ein stochastischer Prozeß. Die Theorie stochastischer Prozesse und ihre Anwendung auf Probleme des Operations Research sind das Thema dieser Vorlesung.

Ausgeschlossen bleibt dabei die „Statistik stochastischer Prozesse“, d. h. Test- und Schätztheorie in diesen Modellen. Ebenfalls nicht geliefert wird eine allgemeinste mathematische Theorie; vielmehr sollen wichtige Teilklassen stochastischer Prozesse vorgestellt werden, deren Anwendungen meist offensichtlich sind.

Innerhalb des Operations Research haben sich unter der Bezeichnung „Angewandte stochastische Prozesse“ einige Teilklassen von Problemen zu eigenen Forschungsgebieten entwickelt:

- Ersatz- und Erneuerungstheorie,
- Netzplantheorie,
- Lagerhaltung,
- Bedienungstheorie (Warteschlangen),
- Zuverlässigkeitstheorie,
- Stochastische dynamische Optimierung.

Fast alle Klassen stochastischer Prozesse finden in einer oder der anderen dieser Problemkategorien Anwendungen. Einige wichtige Klassen:

- Erneuerungstheorie und regenerative Prozesse,
- Markov-Prozesse,
- Zeitreihen,
- Punktprozesse,
- Martingale,
- Stationäre Prozesse.

Als weitere wichtige Problemkreise sind zu nennen:

- Simulation stochastischer Prozesse,
- Statistik stochastischer Prozesse.

Anmerkung : Dieser Text, verwendet als Arbeitsunterlage zu einer Vorlesung „Stochastische Methoden des Operations Research“ am Fachbereich Mathematik der Universität Hamburg, enthält im wesentlichen klassischen Stoff aus der Theorie stochastischer Prozesse. Wie im Verlauf des Textes zu erkennen, habe ich mich auf einige, wie ich finde, auch sehr zum Weiterstudium zu empfehlende Werke gestützt: Es sind dies die Bücher von Asmussen [Asm87], Chung [Chu67], Cinlar [Çin75], Feller [Fel68] und [Fel71] und Hinderer [Hin72].

2 Grundbegriffe für stochastische Prozesse

2.1 Definition

Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, (E, \mathcal{S}) ein Meßraum, $T \neq \emptyset$ eine Menge. Ein stochastischer Prozeß mit Indexmenge T und Zustandsraum (E, \mathcal{S}) ist eine Familie $X = (X_t : t \in T)$ von Zufallsvariablen (\mathcal{A} - \mathcal{S} -meßbaren Abbildungen)

$$X_t : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{S}), \quad \omega \mapsto X_t(\omega).$$

2.2 Anmerkung

- a) Meist wird T als Zeitparameter interpretiert, insbesondere in klassischen Anwendungen. Man unterscheidet
- diskrete Zeit, z. B.: $\mathbb{N} = \{0, 1, 2, \dots\}$, \mathbb{Z} oder Teilmengen davon;
 - stetige Zeit, z. B.: $\mathbb{R}_+ = [0, \infty)$, \mathbb{R} oder Teilmengen davon.
- b) Zustandsräume repräsentieren die möglichen Ereignisse und Stichprobenräume. Man unterscheidet
- diskrete Räume, z. B.: $(\mathbb{N}^n, \mathcal{P}(\mathbb{N}^n))$, $(\{0, 1, \dots, n\}, \mathcal{P}(\{0, 1, \dots, n\}))$ (In diesem Fall werden wir die Potenzmengen- σ -Algebren nicht mitschreiben.)
 - stetige Räume, z. B.: $(\mathbb{R}^n, \mathbb{B}^n)$, $(\{0, 1\}^{\mathbb{N}}, \sigma(\{\{0, 1\}^r \times \{0\} \times \{0, 1\}^{\mathbb{N}} : r \in \mathbb{N}\}))$
- c) Andere Schreibweisen: $X = (X(t) : t \in T)$, $X_t(\omega) = X(t)(\omega) = X(t, \omega)$. □

2.3 Anmerkung

- a) Der unterliegende Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) eines stochastischen Prozesses $X = (X_t : t \in T)$ wird als die steuernde oder verursachende Wirklichkeit (oder Gesamtsystem) angesehen, aus der durch die Familie der Abbildungen $(X_t : t \in T)$ interessierende Systemdetails (Meßergebnisse, Beobachtungen) ausgesondert oder abgeleitet werden.
- b) Tritt $\omega \in \Omega$ ein, so ist damit eine Folge $(X_t(\omega) : T \in T)$ als durch ω bestimmter Pfad (Realisierung, Trajektorie) eingetreten (zu beobachten). Für festes $\omega \in \Omega$ ist also genau eine Abbildung

$$X_{(\bullet)}(\omega) = X(\bullet, \omega) : T \rightarrow E \in E^T$$

bestimmt.

- c) Ist T endlich, so hat man die stochastische Entwicklung des Prozesses $X = (X_t : t \in T)$ über die gemeinsame Verteilung des Vektors $(X_t : t \in T)$ vollständig ohne Probleme im Griff, d. h. mit Messungen zu endlich vielen Zeiten kann man z. B. die auftretenden Verteilungen schätzen.
- d) Ist T nicht endlich (insbesondere nicht abzählbar), so ist es im allgemeinen ebenfalls nur möglich, den Prozeß zu endlich vielen Zeiten innerhalb eines Experiments zu beobachten, d. h. man kann allenfalls gemeinsame Verteilungen der Art $\mathcal{L}(X_{t_1}, \dots, X_{t_n})$, $t_1, \dots, t_n \in T$, beobachten und schätzen. Es ergibt sich also ganz pragmatisch das folgende Problem:

Sei $T \neq \emptyset$ eine Indexmenge und

$$\mathcal{H}(T) = \{(t_1, \dots, t_n) : t_i \in T, t_i \neq t_j \text{ für } i \neq j, i, j = 1, \dots, n, n \in \mathbb{N}_+\}$$

die Menge der endlichen Folgen aus T ohne Wiederholung und

$$\mathcal{P} = \{P_{t_1, t_2, \dots, t_n} : (t_1, \dots, t_n) \in \mathcal{H}(T)\}$$

eine Familie von Wahrscheinlichkeitsmaßen auf den jeweiligen Produktmeßräumen $(E, \mathcal{S})^n$. Existiert dann ein stochastischer Prozeß $X = (X_t : t \in T)$ auf einem geeigneten Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit Zustandsraum (E, \mathcal{S}) und endlichdimensionalen Randverteilungen

$$\mathcal{L}(X_{t_1}, X_{t_2}, \dots, X_{t_n}) = P_{t_1, t_2, \dots, t_n}?$$

Die Antwort:

- i) Im allgemeinen ist die Existenz nicht gesichert.
 ii) Für die bei fast allen Anwendungen auftretenden Probleme sichert ein Satz von Kolmogorov die Existenz eines stochastischen Prozesses zu gegebenen Familien endlichdimensionaler Randverteilungen. Leicht einzusehen ist, daß \mathcal{P} die folgenden Konsistenzbedingungen erfüllen muß (siehe [GS80], S. 41 ff):

- 1) Sei π eine Permutation der Zahlen $(1, 2, \dots, n)$ und

$$f_\pi : E^n \rightarrow E^n, \quad (x_1, \dots, x_n) \mapsto (x_{\pi_1}, \dots, x_{\pi_n})$$

die durch π bestimmte Koordinatenpermutation auf E^n . Dann gilt für $C \in \mathcal{S}^n$ und $(t_1, \dots, t_n) \in \mathcal{H}(T)$:

$$P_{t_1, \dots, t_n}(f_\pi^{-1}C) = P_{t_{\pi_1}, \dots, t_{\pi_n}}(C).$$

- 2) Für alle $A \in \mathcal{S}^n$ und $(t_1, \dots, t_n, t_{n+1}, \dots, t_{n+m}) \in \mathcal{H}(T)$, $n, m \geq 1$, ist

$$P_{t_1, \dots, t_n}(A) = P_{t_1, \dots, t_n, t_{n+1}, \dots, t_{n+m}}(A \times E^m).$$

Ist außerdem

- 3) $(E, \mathcal{S}) = (E, \sigma(\mathcal{O}))$, wobei (E, \mathcal{O}) ein polnischer topologischer Raum ist,

so ist die Existenz eines stochastischen Prozesses mit endlichdimensionalen Randverteilungen \mathcal{P} gesichert.

- e) Die folgenden Aussagen sind äquivalent:

- i) (E, \mathcal{O}) ist polnischer topologischer Raum.
 ii) Es gibt eine die Topologie \mathcal{O} definierende Metrik, die vollständig ist und \mathcal{O} besitzt eine abzählbare Basis.
 iii) (E, \mathcal{O}) ist vollständiger separabler metrischer Raum.

Beispiele: \mathbb{R}^n ; \mathbb{C}^n ; kompakte Räume mit abzählbarer Basis, also insbesondere diskrete (abzählbare) Räume.

- f) Auf die topologischen Voraussetzungen an den Zustandsraum kann verzichtet werden, falls die Parametermenge abzählbar ist und eine Folge von Übergangswahrscheinlichkeiten (bedingte Verteilungen) gegeben ist (zur Definition siehe [BN84], S. 193, 209). Diese Verallgemeinerung des Konstruktionsprinzips für die Modellierung gekoppelter Experimente liefert der Satz von Ionescu Tulcea (siehe [Hin72], S. 134):

Satz (Ionescu Tulcea)

Seien (E_n, \mathcal{S}_n) , $n \in \mathbb{N}_+$, Meßräume und $(E, \mathcal{S}) = \left(\prod_{n=1}^{\infty} E_n, \bigotimes_{n=1}^{\infty} \mathcal{S}_n \right)$; sei Q_1^0 ein Wahrscheinlichkeits-

maß auf (E_1, \mathcal{S}_1) und Q_{n+1}^n eine Übergangswahrscheinlichkeit (Markov-Kern) von $\left(\prod_{i=1}^n E_i, \bigotimes_{i=1}^n \mathcal{S}_i \right)$ nach $(E_{n+1}, \mathcal{S}_{n+1})$, $n = 1, 2, \dots$

Dann gilt: Es gibt genau ein Wahrscheinlichkeitsmaß Q auf (E, \mathcal{S}) mit

$$Q \left(\prod_{i=1}^{\infty} A_i \right) = \int_{A_1} Q_1^0(dx_1) \int_{A_2} Q_2^1(x_1; dx_2) \dots \int_{A_n} Q_n^{n-1}(x_1, \dots, x_{n-1}; dx_n)$$

für alle Mengen $\prod_{i=1}^{\infty} A_i \in \mathcal{S}$ mit $A_i \in \mathcal{S}_i$, $i \in \mathbb{N}_+$, $A_i = E_i$ für alle $i > n$.

$Q := \bigotimes_{i=1}^{\infty} Q_i^{i-1}$ heißt das durch die Folge $(Q_i^{i-1} : i = 1, 2, \dots)$ bestimmte Wahrscheinlichkeitsmaß. \square

3 Markov-Ketten

Die in Anwendungen wohl am häufigsten verwendeten Modelle für zeitlich sich entwickelnde zufallsbeeinflusste Systeme sind Markovsche Prozesse. Wir werden uns mit einer — auch schon von A. A. Markov (1856–1922) in einer Serie von Arbeiten behandelten — Klasse spezieller Markov-Prozesse befassen:

3.1 Definition

Ein stochastischer Prozeß $X = (X_n : n \in \mathbb{N})$ auf (Ω, \mathcal{F}, P) mit höchstens abzählbarem Zustandsraum E heißt Markov-Kette (mit diskretem Parameter), falls gilt (Markov-Eigenschaft):

Für $0 \leq t_1 < t_2 < \dots < t_n$, $t_i \in \mathbb{N}$, $n \geq 2$ und $i_1, i_2, \dots, i_n \in E$ ist

$$P(X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}, X_{t_{n-2}} = i_{n-2}, \dots, X_{t_1} = i_1) = P(X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}), \quad (\text{M})$$

sofern die bedingte Wahrscheinlichkeit auf der linken Seite erklärt ist, d. h. sofern gilt, daß

$$P(X_{t_{n-1}} = i_{n-1}, \dots, X_{t_1} = i_1) > 0.$$

3.2 Anmerkung

- Wenn im folgenden nicht anderes gesagt wird, ist unter $X = (X_n : n \in \mathbb{N})$ stets ein stochastischer Prozeß auf (Ω, \mathcal{F}, P) mit höchstens abzählbarem Zustandsraum E zu verstehen. Die auf E verwendete σ -Algebra ist dann die Potenzmengen- σ -Algebra $\mathcal{P}(E)$.
- Häufig ist $E = \mathbb{N}$, $E = \mathbb{N} \cup \{\infty\}$, $E = \mathbb{Z} \cup \{-\infty, \infty\}$, ...
- Analog zu Definition 3.1 ist eine Markov-Kette mit diskretem Zeitparameter \mathbb{Z} zu definieren. \square

3.3 Lemma

Ein stochastischer Prozeß $X = (X_n : n \in \mathbb{N})$ ist genau dann eine Markov-Kette, wenn eine der drei folgenden Bedingungen erfüllt ist:

- (1) Für $n \geq 2$ und $i_1, \dots, i_n \in E$ ist

$$P(X_n = i_n \mid X_{n-1} = i_{n-1}, \dots, X_1 = i_1) = P(X_n = i_n \mid X_{n-1} = i_{n-1}),$$

sofern die linke Seite erklärt ist.

- (2) Für $0 \leq t_1 < t_2 < \dots < t_n < t_{n+1} < \dots < t_{n+m}$, $n \geq 2$, $m \geq 0$, $t_i \in \mathbb{N}$, und $i_k \in E$, $k = 1, \dots, n+m$ ist

$$\begin{aligned} P(X_{t_{n+m}} = i_{n+m}, \dots, X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}, \dots, X_{t_1} = i_1) \\ = P(X_{t_{n+m}} = i_{n+m}, \dots, X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1}), \end{aligned}$$

sofern die linke Seite erklärt ist.

- (3) Sei $\sigma(X_t : t \geq t_n)$ die von den Zufallsvariablen $\{X_t : t \geq t_n\}$ erzeugte Unter- σ -Algebra von \mathcal{F} . Für $0 \leq t_1 < t_2 < \dots < t_n$, $t_i \in \mathbb{N}$, $n \geq 2$, $i_1, \dots, i_n \in E$, und $M \in \sigma(X_t : t \geq t_n)$ ist

$$P(M \mid X_{t_n} = i_n, \dots, X_{t_1} = i_1) = P(M \mid X_{t_n} = i_n),$$

sofern die linke Seite erklärt ist.

Beweis:

- (M) \Rightarrow (2): Durch Induktion über m .
Im Falle $m = 0$ ist (2) die Markov-Eigenschaft selbst.

Für $m \geq 0$ sei die Behauptung richtig. Mit geeigneten t_j, i_j sei der folgende Ausdruck erklärt und es gelte $P(X_{t_\nu} = i_\nu, 1 \leq \nu \leq m+n) > 0$:

$$\begin{aligned}
& P(X_{t_\nu} = i_\nu, n \leq \nu \leq n+m+1 \mid X_{t_\nu} = i_\nu, 1 \leq \nu \leq n-1) \\
&= P(X_{t_{n+m+1}} = i_{n+m+1} \mid X_{t_\nu} = i_\nu, 1 \leq \nu \leq n+m) \\
&\quad \cdot P(X_{t_\nu} = i_\nu, n \leq \nu \leq n+m \mid X_{t_\nu} = i_\nu, 1 \leq \nu \leq n-1) \\
&= P(X_{t_{n+m+1}} = i_{n+m+1} \mid X_{t_{n+m}} = i_{n+m}) \\
&\quad \cdot P(X_{t_\nu} = i_\nu, n \leq \nu \leq n+m \mid X_{t_{n-1}} = i_{n-1}) \\
&= P(X_{t_{n+m+1}} = i_{n+m+1} \mid X_{t_\nu} = i_\nu, n-1 \leq \nu \leq n+m) \\
&\quad \cdot P(X_{t_\nu} = i_\nu, n \leq \nu \leq n+m \mid X_{t_{n-1}} = i_{n-1}) \\
&= P(X_{t_{n+m+1}} = i_{n+m+1}, \dots, X_{t_n} = i_n \mid X_{t_{n-1}} = i_{n-1})
\end{aligned}$$

Ist $P(X_{t_\nu} = i_\nu, 1 \leq \nu \leq n+m) = 0$ und $P(X_{t_\nu} = i_\nu, \nu = 1, \dots, n-1) > 0$, so ist der Ausgangsterm auch 0 (nach Definition). Es folgt aber auch:

$$\begin{aligned}
0 &= P(X_{t_\nu} = i_\nu, n \leq \nu \leq n+m \mid X_{t_\nu} = i_\nu, 1 \leq \nu \leq n-1) \\
&= P(X_{t_\nu} = i_\nu, n \leq \nu \leq n+m \mid X_{t_{n-1}} = i_{n-1}) \\
&\geq P(X_{t_\nu} = i_\nu, n \leq \nu \leq n+m+1 \mid X_{t_{n-1}} = i_{n-1}) \\
&\geq 0.
\end{aligned}$$

b) (2) \implies (3): Das Mengensystem

$$\{(X_{t_n+t_\nu} = i_\nu : \nu = 1, \dots, m) : 0 \leq t_1 < \dots < t_m; t_i \in \mathbb{N}; i_1, \dots, i_m \in E\}$$

ist ein \cap -stabiles Erzeugendensystem von $\mathcal{F}(X_t : t \geq t_n)$. Auf diesem stimmen nach (2) die Verteilungen

$$P(\bullet \mid X_{t_n} = i_n) \quad \text{und} \quad P(\bullet \mid X_{t_n} = i_n, \dots, x_{t_1} = i_1)$$

überein. Da beide Maße auf dem Erzeuger endlich sind, stimmen sie insgesamt überein.

c) (3) \implies (1) ist klar.

d) (1) \implies (M): Wir betrachten zunächst ein Beispiel. Es sei erklärt

$$\begin{aligned}
& P(X_4 = i_4 \mid X_3 = i_3, X_1 = i_1) \\
&= \sum_{i_2} \frac{P(X_4 = i_4, X_3 = i_3, X_2 = i_2, X_1 = i_1)}{P(X_3 = i_3, X_2 = i_2, X_1 = i_1)} \cdot \frac{P(X_3 = i_3, X_2 = i_2, X_1 = i_1)}{P(X_3 = i_3, X_1 = i_1)} \\
&\stackrel{(1)}{=} P(X_4 = i_4 \mid X_3 = i_3) \cdot \sum_{i_2} \frac{P(X_3 = i_3, X_2 = i_2, X_1 = i_1)}{P(X_3 = i_3, X_1 = i_1)} \\
&= P(X_4 = i_4 \mid X_3 = i_3)
\end{aligned}$$

(wobei nur über solche Zustände i_2 summiert wird, für die der Nenner des ersten Faktors positiv ist; mindestens ein solches i_2 existiert, denn sonst wäre der Ausgangsterm nicht erklärt). Mit hinreichend großem Notations- und Indexaufwand ist der allgemeine Fall genauso zu zeigen. \blacksquare

Die in (M) und den Charakterisierungen (1), (2), (3) aus Lemma 3.3 auftretenden Größen sind „elementare bedingte Wahrscheinlichkeiten“ im Sinne von [BN84], S. 95, da wir stets gefordert haben, daß „die linke Seite erklärt ist“. Falls der Zustandsraum einer „Markov-Kette“ nicht abzählbar ist, kann diese Forderung im allgemeinen nicht erfüllt werden (Beispiel: $E = \mathbb{R}$, mit stetigen Verteilungen). Die Verallgemeinerung, die sich aus dem Markovschen Konzept (M) bis heute etabliert hat, wird — auch mit Berücksichtigung stetiger Zeit — formuliert als

3.4 Definition

Ein stochastischer Prozeß $X = (X_t : t \in T)$ mit $T \subseteq \mathbb{R}$, $X_t : (\Omega, \mathcal{F}, P) \rightarrow (E, \mathcal{S})$, $t \in T$, heißt Markov-Prozeß, falls gilt:

Für alle $t_1 < t_2 < \dots < t_n$ mit $t_i \in T$, $i = 1, \dots, n$, $n \in \mathbb{N}$ ist

$$P^{X_{t_n} | (X_{t_{n-1}}, \dots, X_{t_1})} = P^{X_{t_n} | X_{t_{n-1}}} \quad P\text{-fast sicher.}$$

Die hier auftauchenden Größen sind „bedingte Verteilungen“, welche Spezialfälle sogenannter „bedingter Erwartungen“ sind ([Bau74], Kap. X). Für den in Anwendungen häufig zu findenden Fall, daß (E, \mathcal{S}) ein polnischer Raum mit Borelscher σ -Algebra ist, sind die bedingten Verteilungen als Übergangswahrscheinlichkeiten oder Markov-Kerne ([BN84], S. 209, Satz 5.4.11) interpretierbar. Die dabei und im allgemeinen Fall auftauchenden Schwierigkeiten lassen sich technisch am besten mit dem Kalkül der bedingten Erwartungswerte behandeln. Eine kurze Einführung gibt das Einführungskapitel von [Chu82].

Sowohl Definition 3.4 als auch (M) in Definition 3.1 formalisieren das Konzept eines stochastischen Prozesses als Modell für ein zeitlich sich entwickelndes, zufallsbeeinflusstes System, dessen zukünftige stochastische Entwicklung nur vom gegenwärtigen Zustand abhängt. Genauer: Die Übergangswahrscheinlichkeiten sind von der weiteren Vergangenheit unabhängig, wenn die Gegenwart bekannt ist; oder noch eingängiger: Die Zukunft hängt von der Vergangenheit nur über die Gegenwart ab.

Ein analoges Prinzip liegt vielen Theorien in anderen Wissenschaften und technischen oder wirtschaftlichen Anwendungen zugrunde. Prototyp und Denkschema für andere Wissenschaften war die klassische Mechanik, deren hier interessierende Grundannahme wie folgt zu skizzieren ist:

Für jedes mechanische System gibt es ein Regelwerk aus physikalischen Gesetzen, dargestellt in den „Bewegungsgleichungen“ des Systems (einem Differentialgleichungssystem). Aus den (hinreichend) vollständigen Daten über den Augenblickszustand des Systems kann mittels der Bewegungsgleichungen jeder zukünftige Zustand des Systems berechnet werden — und ganz symmetrisch dazu auch jeder vergangene Zustand des Systems rekonstruiert werden. (Die Bewegungsgleichungen der klassischen Mechanik sind invariant unter Zeitumkehr.)

Seine ganz explizite Ausformung findet dieses Prinzip im „Zustandskonzept“ der Systemtheorie der Ingenieurwissenschaften: Der „interne Zustand“ eines Systems enthält sämtliche Information, die notwendig ist, um aus einer erfolgten Eingabe und internem Zustand den Folgezustand zu errechnen. Derartige Systeme sind auch die (deterministischen) Automaten der Informatik, die Bewegungsgleichungen werden dort als Zustandsüberföhrungsfunktion bezeichnet ([Unb90], [EKKK74]).

Die Bewegungsgleichung der Mechanik und die Zustandsüberföhrungsfunktionen der Automatentheorie beschreiben deterministische Entwicklungsregeln; deren numerische Komplexität galt im wesentlichen als einziger Grund dafür, daß Vorhersagen und Rekonstruktion nur unzuverlässig möglich waren. Eine Intelligenz mit unendlichen Rechen- und Speicherkapazitäten (der sogenannte „Laplacesche Dämon“) wäre nach Auffassung der meisten Physiker des 19. Jahrhunderts in der Lage, die Entwicklung jedes physikalischen Systems exakt vorherzusagen. Da die Physik, in ihrer damals fast vollkommenen Ausprägung als Erklärung der Welt durch zugrundeliegende mechanische Prinzipien und Gesetze, Leitwissenschaft war, findet sich auch in anderen Wissenschaften, selbst den Sozial- und Wirtschaftswissenschaften, der Versuch, die beobachteten Phänomene durch „mechanische Gesetze“ zu erklären. Die heutige Systemtheorie (deterministischer Systeme) zeugt noch davon, denn diese wird noch regelhaft dort als Denkprinzip verwendet.

Allerdings: Nach den heute zumeist akzeptierten grundlegenden Gesetzen der Physik sind zum einen Meßungenauigkeiten bei der vollständigen Beschreibung des aktuellen Zustands eines Systems prinzipiell nicht zu vermeiden und zum anderen sind für grundlegende Systeme nur stochastische Bewegungsgleichungen formulierbar.

Ein Grund für den unbestreitbaren Erfolg des skizzierten Zustandskonzeptes der Mechanik ist die Möglichkeit, in den Modellen dieser Theorie effektive Rechenverfahren herleiten zu können; Theorie und Numerik von Differentialgleichungssystemen sind die Ergebnisse diesbezüglicher Forschung und Praxis.

Denselben Grund können wir anführen für den Erfolg Markovscher Modelle und deren Beliebtheit bei den Anwendern. Insbesondere für Markovsche Ketten werden wir die konkreten Auswertungsprobleme zur

Vorhersage stochastischer Entwicklungen auf bekannte Kalküle der Numerik zurückführen können. Zentrales Hilfsmittel wird der Matrizenkalkül sein. Dessen relative Simplität macht die Markov-Ketten zu einem Standardhilfsmittel bei der Modellierung naturwissenschaftlicher, technischer und sozialer Phänomene.

Vor dem Hintergrund dieser aktuellen Zustandsbeschreibung erscheint es geradezu merkwürdig, daß die Untersuchung derartiger Prozesse aus ganz anderen Anlässen heraus begonnen wurde.

A. A. Markovs Arbeiten der Jahre 1906–1912 etwa beschäftigen sich mit der Verallgemeinerung des Gesetzes der großen Zahlen und des zentralen Grenzwertsatzes auf Folgen abhängiger Variabler. Zur gleichen Zeit etwa wurden „Markovsche Ketten“ von E. H. Bruns (1848–1919) untersucht. Näheres findet sich bei Maistrov ([Mai74], S. 216 ff.) und Schneider ([Sch88]). Markov-Ketten wurden also eingeführt und behandelt als naheliegende Verallgemeinerung unabhängiger Folgen von Zufallsvariablen. (Letztere sind nach (M) auch die einfachsten Markov-Ketten.)

Neben diesen rein theoretischen Arbeiten hat Markov selbst angeblich nur ein einziges praktisches Beispiel für das Auftreten von Phänomenen gegeben, die durch Markov-Ketten beschrieben werden können: In A. A. Markovs „*Versuche einer statistischen Untersuchung über den Text des Romans ‚Eugen Onegin‘ zur Beleuchtung des Zusammenhangs der Kettenversuche*“ (russisch, [Mar13]), wurde die Folge der Buchstaben in Puschkins Roman als Folge von Zufallsvariablen angesehen, welche die Werte {Vokal, Konsonant} annehmen konnten. Es sei darauf hingewiesen, daß seit Ende der siebziger Jahre bei der Entwicklung von automatischen Spracherkennungssystemen Markovsche Modelle erneut, wenn auch mit komplexerer Struktur, eingesetzt werden. Einen Überblick geben Juang/Rabiner ([JR91]).

Unabhängig von dieser Entwicklung waren Markovsche Prozesse in der heutigen Formulierung gemäß Definition 3.4 schon kurze Zeit vorher verwendet worden zur Untersuchung und theoretischen Analyse realer Phänomene; A. Einstein ([Ein05]) führte die Markov-Eigenschaft der untersuchten stochastischen Prozesse aus der Annahme unterliegender mechanistischer Prinzipien heraus ein. Nicht so eindeutig sind die Modellierungsgründe bei L. Bachelier ([Bac00]), wo Markovsche Prozesse bei der mathematischen Beschreibung von Kursschwankungen an der Börse verwendet werden. Beide Arbeiten entwickelten heuristische Prozesse, die heute als Brownsche Bewegung oder Wiener Prozeß bezeichnet werden und Standardmodelle sind für stochastisch beeinflusste Diffusionsvorgänge. Diese technisch-naturwissenschaftlichen Modelle sind in den achtziger Jahren erneut von den Wirtschaftswissenschaftlern aufgegriffen worden, um, in Anlehnung an die Ideen Bacheliers, in der Finanzmathematik rationale Entscheidungsfindung, z. B. in der Anlagepolitik, zu unterstützen.

Weitere Anwendungsgebiete Markovscher Ketten und Prozesse sind Chemie und Biologie, in letzterem z. B. Geburts- und Todesprozesse, Vererbungsprozesse¹, Populationsprozesse; ebenso Soziologie, Psychologie (Lernmodelle), etc.

Ein Beispiel dafür, wie weitgehend Markovsche Beschreibungen realer Phänomene heute unterschiedliche Wissenschaften durchdringen, findet sich in [Hak78]. Für WirtschaftswissenschaftlerInnen ist in diesem Zusammenhang interessant, daß die Synergetik inzwischen auch in ihrer Wissenschaft als Methode und Denkschema Eingang gefunden hat.

Eine letzte Paralle zwischen klassischer Mechanik und Markov-Prozessen erhalten wir aus der Beobachtung, daß die Anwendung mechanischer Modelle auch dort quantitative Ergebnisse, die in der Praxis verwertbar sind, liefert, wo eigentlich Methoden der neueren Physik angewendet werden müßten. Dementsprechend verwendet man Markovsche Modelle häufig auch dort als Modellapproximation, wo eigentlich feinere Methoden näherliegend wären. Hintergedanke des Vorgehens ist: Besser irgendwelche, wenn auch kritisch zu wertende quantitative Ergebnisse zu finden, als ganz auf Ergebnisse zu verzichten! — Selbstverständlich ist bei der Interpretation derartiger gewonnener Größen Vorsicht angebracht und eine enge Zusammenarbeit mit Fachleuten des entsprechenden Anwendungsgebietes notwendig. Dennoch zeigt die Erfahrung: Die Markovsche Theorie liefert für viele Anwendungen hinreichend genaue, approximative Modelle auch für nicht-Markovsche Prozesse.

In der Praxis wird häufig noch weiter gegangen: Es gibt Techniken, mit denen nicht-Markovsche Prozesse „Markovisiert“ werden können, um dann den gut ausgebauten Rechenkalkül der Markov-Theorie

¹Eine sehr frühe diesbezügliche Arbeit stammt von S. N. Bernstein, siehe [Ber23].

einsetzen zu können. Die Grundidee des „Markovisieren“ ist dabei vergleichbar dem Prinzip des „internen Zustands“ der Systemtheorie: Nehme alle notwendige Kenntnis der Vergangenheit in den augenblicklichen Zustand auf — und zwar so weit, daß Du die stochastische Zukunft quantifizieren kannst.

3.5 Definition

Sei $X = (X_n : n \in \mathbb{N})$ eine Markov-Kette.

a) Ist dann die bedingte Wahrscheinlichkeit

$$p^{(m)}(n; i, j) := P(X_{m+n} = j \mid X_n = i) \quad \text{für } i, j \in E, m \geq 1, n \geq 0,$$

definiert, so wird sie als „ m -Schritt-Übergangswahrscheinlichkeit von i nach j zur Zeit n “ bezeichnet.

b) Ist für alle m, i, j

$$p^{(m)}(n; i, j) =: p^{(m)}(i, j)$$

von n unabhängig, so heißt X homogene Markov-Kette oder Markov-Kette mit stationären Übergangswahrscheinlichkeiten.

3.6 Anmerkung

a) Wenn im folgenden nicht anderes gesagt wird, ist mit „Markov-Kette X “ stets eine homogene Markov-Kette auf (Ω, \mathcal{F}, P) mit höchstens abzählbarem Zustandsraum E und stationären Übergangswahrscheinlichkeiten $p^{(m)}(i, j)$, $m \geq 1$, $i, j \in E$, bezeichnet.

b) Mit

$$p^{(m)} = (p^{(m)}(i, j) : i, j \in E) = (P(X_m = j \mid X_0 = i) : i, j \in E), \quad \text{für } m \geq 1,$$

bezeichnen wir die Matrix der m -Schritt-Übergangswahrscheinlichkeiten und schreiben kurz

$$p(i, j) := p^{(1)}(i, j) = P(X_1 = j \mid X_0 = i), \quad p := (p(i, j) : i, j \in E).$$

p heißt Übergangsmatrix von X . Den Vektor

$$p_0 = (P(X_0 = i) : i \in E)$$

bezeichnen wir als Anfangs- oder Startverteilung von X und

$$p_n = (P(X_n = i) : i \in E) \quad \text{für } n \in \mathbb{N}$$

heißt absolute Wahrscheinlichkeit oder Zustandsverteilung von X zur Zeit n . Mit der angegebenen Bedeutung schreiben wir auch — sofern X vorgegeben ist — kurz:

$$p_n = (p_n(i) : i \in E) \quad \text{für } n \in \mathbb{N}.$$

c) Wir setzen $p^{(0)}$ fest durch

$$p^{(0)}(i, j) = \delta_{i,j} \quad \text{für } i, j \in E,$$

so daß $p^{(0)}$ die Einheitsmatrix (mit jeweils passender Dimension) ist.

d) Die p_n sind stochastische Vektoren, da

$$\sum_{i \in E} p_n(i) = 1 \quad \text{und} \quad p_n(i) \geq 0 \quad \text{für } i \in E$$

gilt. Die Übergangsmatrizen sind stochastische Matrizen, da die Zeilen stochastische Vektoren sind, d. h. die Matrixelemente sind nichtnegativ und die Zeilensummen sind 1.

e) Wir identifizieren regelhaft Zähldichten und Wahrscheinlichkeitsmaße und verwenden für beide die gleichen Symbole. \square

3.7 Satz

a) Sei X eine homogene Markov-Kette mit Startverteilung

$$p_0 = (P(X_0 = i) : i \in E)$$

und Ein-Schritt-Übergangsmatrix

$$p = (P(X_1 = i | X_0 = j) = p(j, i) : i, j \in E).$$

Dann sind sämtliche endlich-dimensionalen Randverteilungen von X schon durch p_0 und p bestimmt. Die endlich-dimensionalen Randverteilungen bestimmen die Verteilungen von X vollständig.

b) Sei $q_0 = (q(i) : i \in E)$ ein stochastischer Vektor und $q = (q(i, j) : i, j \in E)$ eine stochastische Matrix. Dann existiert eine homogene Markov-Kette Y mit Startverteilung q_0 und Übergangsmatrix q auf einem geeigneten Wahrscheinlichkeitsraum $(\widehat{\Omega}, \widehat{\mathcal{F}}, \widehat{P})$.

Beweis:

a) Als abzählbarer Raum mit der diskreten Topologie ist E vollständiger, separabler metrischer Raum. Der Existenzsatz von Kolmogorov (Anmerkung 2.3 d)) sagt damit, daß zur Bestimmung der Verteilungen von X die endlich-dimensionalen Randverteilungen genügen. Da die $P^{(X_{t_n}, \dots, X_{t_1})}$ diskret sind, genügt die Kenntnis ihrer Zähldichten. Weiter kann man sich auf die Untersuchung der folgenden Ausdrücke beschränken ($n \geq 1, i_1, \dots, i_n \in E$):

$$\begin{aligned} P(X_n = i_n, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) \\ &= P(X_0 = i_0) \prod_{t=1}^n P(X_t = i_t | X_s = i_s : s = 0, \dots, t-1) \\ &= P(X_0 = i_0) \prod_{t=1}^n P(X_t = i_t | X_{t-1} = i_{t-1}) \\ &= p_0(i_0) \prod_{t=1}^n p(i_{t-1}, i_t), \end{aligned}$$

denn derartige Wahrscheinlichkeiten liefern über geeignete Summation sämtliche endlich-dimensionalen Randverteilungen.

b) Konstruiere aus q_0 und q eine konsistente Familie von endlich-dimensionalen Wahrscheinlichkeiten wie in a). Der Satz von Kolmogorov liefert die Behauptung.

Dabei ergibt sich $\widehat{\Omega} = E^{\mathbb{N}}$ als Raum aller E -wertigen Folgen, und $Y = (Y_n : n \in \mathbb{N})$ ist die Folge der Projektionen

$$\bar{\omega} = (\omega_0, \omega_1, \dots) \in \widehat{\Omega} \quad \implies \quad Y_n(\bar{\omega}) = \omega_n \quad (n \in \mathbb{N})$$

(siehe [Chu67], S. 7, oder [Fre83], S. 3). ■

Der Satz 3.7 zeigt, daß die (homogenen) Markov-Ketten nach den zugehörigen Übergangsmatrizen klassifiziert werden können. Aus jeder dieser Klassen wird dann durch Angabe eines Startvektors eine konkrete Markov-Kette herausgegriffen. Es wird sich im folgenden zeigen, daß wesentliche Eigenschaften von Markov-Ketten nur von der Struktur der Übergangsmatrizen abhängen.

Der Matrizenkalkül wird damit eine für uns wesentliche Rolle spielen. Wie dies möglich ist, zeigen die beiden folgenden Aussagen, bei denen wir die Bezeichnungen gemäß 3.5 und 3.6 voraussetzen.

3.8 Satz

Die Familie $(p^{(m)} : m \in \mathbb{N})$ der m -Schritt-Übergangsmatrizen genügt den (Chapman-Kolmogorov-) Gleichungen

$$p^{(m+n)} = p^{(m)} \cdot p^{(n)}, \quad m, n \in \mathbb{N}.$$

Beweis: Da X stationäre Übergangswahrscheinlichkeit hat, gilt für $i, j \in E$:

$$\begin{aligned} p^{(m+n)}(i, j) &= P(X_{m+n} = j \mid X_0 = i) \\ &= \sum_{k \in E} P(X_{m+n} = j, X_m = k \mid X_0 = i) \\ &= \sum_{k \in E} P(X_{m+n} = j \mid X_m = k, X_0 = i) P(X_m = k \mid X_0 = i) \\ &= \sum_{k \in E} P(X_{m+n} = j \mid X_m = k) P(X_m = k \mid X_0 = i) \\ &= \sum_{k \in E} P(X_n = j \mid X_0 = k) P(X_m = k \mid X_0 = i) \\ &= \sum_{k \in E} p^{(m)}(i, k) p^{(n)}(k, j). \quad \blacksquare \end{aligned}$$

3.9 Korollar

a) Für die m -Schritt-Übergangswahrscheinlichkeiten gilt:

$$p^{(m)} = p^m, \quad m \in \mathbb{N}.$$

b) Die absoluten Wahrscheinlichkeiten werden berechnet über

$$p_m = p_0 \cdot p^m, \quad m \geq 0.$$

Beweis: Induktion mittels Satz 3.8. ■

3.10 Beispiel (Restlebensdauerprozeß bei Erneuerungs- und Ersatzproblemen)

Die Lebensdauer (Arbeitszeit) einer Systemkomponente sei verteilt nach einer auf \mathbb{N}_+ konzentrierten Verteilung mit Zähldichte $f = (f_k : k \in \mathbb{N}_+)$. Fällt die Komponente aus, so wird sie durch eine identische Komponente ersetzt.

Wegen der Annahme zufallsbeeinflusster Lebensdauern wird in diesem Zusammenhang „identisch“ interpretiert als: physikalisch nicht unterscheidbar in bezug auf das vorliegende Problem und mit identisch verteilten Lebensdauern.

Zusätzlich wird in der Regel angenommen, daß die Lebensdauern der sukzessiv eingesetzten Komponenten unabhängig voneinander sind — im Modell als stochastische Unabhängigkeit wiedergegeben.

Da häufig die bei Beobachtungsbeginn zur Zeit 0 eingebaute Komponente schon eine Weile arbeitet, hat sie zur Zeit 0 nicht mehr die Lebensdauer-Zähldichte f ; wir nehmen an, daß die Restlebensdauer der zur Zeit 0 arbeitenden Komponente eine Zähldichte $g = (g_k : k \in \mathbb{N}_+)$ hat und daß diese unabhängig von den anderen Komponenten arbeitet.

Wir beschränken uns auf den in der Praxis häufig auftretenden Fall, daß die Zeit zum Ersetzen der Komponenten vernachlässigbar klein ist gegenüber den Lebensdauern der Komponenten.

Es sei Y_n die Restlebensdauer der zur Zeit $n \geq 0$ arbeitenden Komponente. Dann ist $Y = (Y_n : n \in \mathbb{N})$ eine homogene Markov-Kette mit Zustandsraum \mathbb{N}_+ , definiert auf einem unterliegenden Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) , auf dem auch die Lebensdauern als Zufallsvariablen (und Ausgangsdaten) definiert sind. Y ist dann in seiner stochastischen Entwicklung gegeben durch die Startverteilung

$$p_0 = (p_0(k) : k \in \mathbb{N}_+) \quad \text{mit} \quad p_0(k) = P(Y_0 = k) = g_k, \quad k \geq 1,$$

und die Übergangsmatrix

$$p = (p(i, j) : i, j \in \mathbb{N}_+) \quad \text{mit} \quad \begin{cases} p(i, j) = \delta_{i, i-1}, & \text{für } i \geq 2, j \geq 1 \\ p(1, j) = f_j, & \text{für } j \geq 1 \\ p(i, j) = 0 & \text{sonst.} \end{cases}$$

Falls also eine Erneuerung stattfindet (d. h. die arbeitende Komponente die Restlebensdauer 0 erreicht), wird dies hier nicht durch $Y_n = 0$ dargestellt. In stetiger Zeit betrachtet ist der Zustand also die Restlebensdauer am Anfang des jeweiligen Intervalls $[n, n+1)$. Ein typischer Pfad des Prozesses läßt sich wie in Abbildung 1 dargestellt skizzieren.

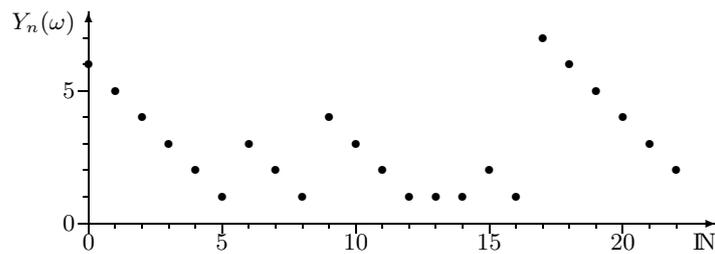


Abbildung 1: Restlebensdauerprozeß in diskreter Zeit

Der deutlicheren Schreibweise wegen wird dies meist über rechtsstetige Treppenfunktionen wie in Abbildung 2 dargestellt.

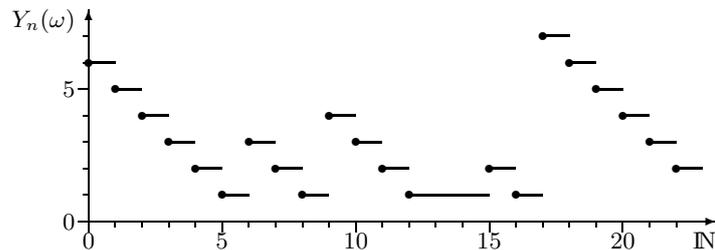


Abbildung 2: Restlebensdauerprozeß in diskreter Zeit als rechtsstetige Treppenfunktion

Der physikalische Ablauf des Systems findet nichtsdestoweniger in stetiger Zeit statt; die arithmetischen Lebensdauerverteilungen erlauben eine (im allgemeinen einfachere) Darstellung und rechnerische Auswertung mit diskreten Methoden. (Die Restlebensdauer zur Zeit 6.5 ist für das gegebene $\omega \in \Omega$ gerade 2.5 Zeiteinheiten.) Der angegebene Pfad ist also nur eine vereinfachende, wenn auch völlig hinreichende Beschreibung der Systementwicklung, wie Abbildung 3 sie widerspiegelt.

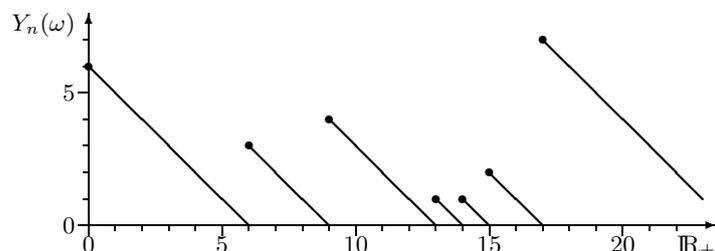


Abbildung 3: Restlebensdauerprozeß in stetiger Zeit

Dabei ist $\tilde{Y} = (\tilde{Y}_t : t \in \mathbb{R}_+)$ ein (geeignet definierter Markov-)Prozeß in stetiger Zeit mit stetigem Zustandsraum. Derartige Prozesse werden auch als „stückweise linear“ bezeichnet ([GK74], S. 167). \square

Die im Beispiel 3.10 statuierte Behauptung, daß Y eine Markov-Kette ist, haben wir *nicht* bewiesen. Mit der korrekten Festlegung der Übergangswahrscheinlichkeiten gibt man sich, wenn wie im obigen Beispiel die Markov-Eigenschaft *offensichtlich* ist, in der Praxis meist zufrieden.

Das „offensichtlich“ kann aber durchaus ein delikates Problem sein. Falls dies so ist, hilft manchmal eine *pfadweise* Konstruktion von Y aus den Ausgangsdaten, hier einer Folge $X = (X_n : n = 1, 2, \dots)$ von unabhängigen Zufallsvariablen mit $X_1 \sim g$, $X_n \sim f$ für $n \geq 2$, die Einsicht zu befördern. Als Beispiel dafür mögen Erneuerungsprozesse in diskreter Zeit dienen:

3.11 Beispiel (Erneuerungs- und Zählprozesse)

$X = (X_n : n \geq 1)$ sei eine Folge von unabhängigen Zufallsvariablen mit Werten in $\overline{\mathbb{N}}_+$:

$$X_n : (\Omega, \mathcal{F}, P) \rightarrow (\overline{\mathbb{N}}_+, \mathcal{P}(\overline{\mathbb{N}}_+)), \quad n \geq 1.$$

Es gelte

$$X_1 \sim g = (g_k : k \in \overline{\mathbb{N}}_+) \quad \text{und} \quad X_i \sim f = (f_k : k \in \overline{\mathbb{N}}_+) \quad \text{für } i \geq 2,$$

wobei f und g Zähldichten auf $\overline{\mathbb{N}}_+$ seien.

a) Die Folge $S = (S_k : (\Omega, \mathcal{F}, P) \rightarrow (\overline{\mathbb{N}}_+, \mathcal{P}(\overline{\mathbb{N}}_+)), k = 0, 1, 2, \dots)$, gegeben durch

$$S_0(\omega) = 0 \quad \text{und} \quad S_k(\omega) = \sum_{i=1}^k X_i(\omega), \quad k = 1, 2, \dots$$

für $\omega \in \Omega$, heißt Erneuerungsprozeß mit Erneuerungsichte f . Die $(X_i : i = 1, 2, \dots)$ heißen Lebensdauern oder Erneuerungsintervalle, die $(S_k : k = 1, 2, \dots)$ heißen Erneuerungszeiten. (0 wird also nicht als Erneuerungszeit gezählt.)

b) Der stochastische Prozeß

$$N = (N_t : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{N}_0, t \in \mathbb{N}_0),$$

gegeben durch

$$N_t(\omega) := \sum_{k=1}^{\infty} 1_{(S_k(\omega) \leq t)} = \sup\{k \in \mathbb{N}_0 : S_k(\omega) \leq t\}$$

heißt der zu $(S_k : k \in \mathbb{N})$ bzw. $(X_i : i \in \mathbb{N}_+)$ gehörige Zählprozeß. \square

Für die Freundinnen und Freunde der Maßtheorie sei angemerkt, daß die Meßbarkeit der (S_k) und (N_t) zu zeigen ist, bevor in Beispiel 3.11 von „stochastischen Prozessen“ gesprochen werden darf. Hier folgt dies aus elementaren Regeln über die Verknüpfung meßbarer Funktionen ([BN84], Satz 5.1.11). Dasselbe gilt auch für das folgende Korollar.

3.12 Korollar

a) Mit den Bezeichnungen aus Beispiel 3.11 gilt für den in Beispiel 3.10 definierten Restlebensdauerprozeß $Y = (Y_n : n = 0, 1, \dots)$:

Die Restlebensdauer der zur Zeit 0 arbeitenden Komponente sei

$$X_1 : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{N}_+, \mathcal{P}(\mathbb{N}_+)),$$

die Lebensdauer der i -ten neu eingesetzten Komponente sei

$$X_i : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{N}_+, \mathcal{P}(\mathbb{N}_+)), \quad i = 2, 3, \dots$$

Dann ist

$$Y_n = S_{N_n+1} - n, \quad n \geq 0.$$

b) Sei $Z = (Z_n : n \in \mathbb{N})$ das Alter der zur Zeit $n \in \mathbb{N}$ arbeitenden Komponente. (Falls noch die zur Zeit 0 arbeitende Komponente lebt: das Alter nach 0.) Dann ist

$$Z_n = n - S_{N_n}, \quad n \geq 0.$$

Im Korollar 3.12 sind die Übergangswahrscheinlichkeiten leicht hinzuschreiben, falls $g = f$ gilt. Ist dies nicht der Fall, benötigen wir etwas mehr Überlegung!

3.13 Beispiel (Lagerhaltung)

In einem Lager wird eine Ware vorrätig gehalten, um einen fortlaufenden Bedarf zu befriedigen. Wir nehmen an, daß die Ware stückweise gezählt, geliefert und verkauft wird. Die Lagerhaltungspolitik wird durch ein Paar kritischer Werte $(s, S) \in \mathbb{N}^2$ mit $s < S$ beschrieben. Der Lagerbestand wird in den Zeitpunkten $0, 1, 2, \dots$ überprüft. Ist der Bestand zur Zeit $n \in \mathbb{N}$ kleiner oder gleich s , so wird durch eine sofortige Bestellung und Lieferung der Bestand auf S zur Zeit $(n+)$ aufgestockt. Ist der Bestand zur Zeit $n \in \mathbb{N}$ größer als s , so wird keine Auffüllung des Lagers vorgenommen.

Es ist bekannt, daß in vielen Fällen eine derartige Bestellpolitik optimal ist bei Annahme vernünftiger Kostenfunktionen. (Gezeigt werden kann dies mit den Methoden der stochastischen dynamischen Optimierung; eine Diskussion auch der Anwendbarkeit dieser Politik findet sich z. B. bei [Sch81].)

Sei $X = (X_n : n \in \mathbb{N})$ der beschreibende Prozeß für den Lagerbestand, Z_n der in einem Intervall $[n, n+1)$ aufgetretene Bedarf. Dann gilt für $n = 0, 1, \dots$ und vorgegebenen Anfangsbestand $X_0 \leq S$:

$$X_{n+1} = \begin{cases} X_n - Z_n, & \text{falls } s < X_n, Z_n \leq X_n \\ S - Z_n, & \text{falls } X_n \leq s, Z_n \leq S \\ 0 & \text{sonst} \end{cases}$$

Sind die $\{Z_n : n \in \mathbb{N}\}$ Zufallsvariablen mit Werten in \mathbb{N} und X_0 eine Zufallsvariable mit Werten in $\{0, 1, \dots, S\}$, so ist auch X ein stochastischer Prozeß in diskreter Zeit mit endlichem Zustandsraum $\{0, 1, \dots, S\}$. X ist im allgemeinen keine Markov-Kette. Es gilt:

$$P(X_{n+1} = j \mid X_\nu = j_\nu, 1 \leq \nu \leq n) = P(X_{n+1} = j \mid X_n = j_n)$$

dann und nur dann, wenn gilt

$$P(Z_n = k \mid X_\nu = j_\nu, 1 \leq \nu \leq n) = P(Z_n = k \mid X_n = j_n), \quad k, j, j_\nu \in \mathbb{N}, \quad n \geq 0, \quad \nu = 1, \dots, n. \quad \square$$

3.14 Beispiel (Qualitätskontrolle)

Am Ende eines Produktionsprozesses werden die hergestellten Stücke einer Routine-Kontrolle unterworfen. Falls der Produktionsablauf in Ordnung ist, d. h. falls keine systematischen Fehler gemacht werden, wird das Auftreten von Fehlern rein zufällig sein mit zeitlich nicht veränderten Fehlerwahrscheinlichkeiten.

Ein geeignetes Modell wäre dann eine Folge $X = (X_n : n \in \mathbb{N}_+)$ von unabhängigen, identisch auf $\{0, 1\}$ verteilten Zufallsvariablen, wobei „ $1 = X_n$ “ bedeutet, daß das n -te Stück fehlerhaft ist. X ist eine Markov-Kette. \square

3.15 Beispiel (Bernoulli-Prozeß)

$X = (X_n : n \in \mathbb{N}_+)$ sei ein stochastischer Prozeß auf (Ω, \mathcal{F}, P) mit Werten in $\{0, 1\}$. Sind die $\{X_n : n \in \mathbb{N}_+\}$ stochastisch unabhängig, identisch verteilt mit $P(X_1 = 1) = p \in [0, 1]$, so heißt X Bernoulli-Prozeß mit Parameter p . X ist dann eine Markov-Kette mit der Übergangsmatrix $p = (p(i, j) : i, j \in \{0, 1\})$ gegeben als

$$p = \left(\begin{array}{c|cc} p(i, j) & 0 & 1 \\ \hline 0 & 1-p & p \\ 1 & 1-p & p \end{array} \right).$$

Die Startverteilung von X ist (wegen der Zählung der Zeit als $1, 2, \dots$)

$$p_1 = (1-p, p) = (P(X_1 = 0), P(X_1 = 1)).$$

Die endlich-dimensionalen Randverteilungen von X sind dann gegeben als

$$P(X_{t_\nu} = i_\nu, \nu = 1, \dots, n) = \prod_{i=1}^n p^{i_\nu} (1-p)^{1-i_\nu}$$

für $n \in \mathbb{N}_+$, Zeitpunkte $1 \leq t_1 < t_2 < \dots < t_n$ und $i_\nu \in \{0, 1\}$ ($\nu = 1, \dots, n$).

Von größerem Interesse z. B. in der Qualitätskontrolle ist der aus X abgeleitete Prozeß, der die Anzahl der fehlerhaften Produkte unter den ersten n Produkten zählt. Für $n \in \mathbb{N}$ und $\omega \in \Omega$ sei

$$S_n(\omega) = \begin{cases} 0 & \text{für } n = 0, \\ \sum_{i=1}^n X_i(\omega) & \text{für } n \geq 1. \end{cases}$$

Dann ist $S = (S_n : n \in \mathbb{N})$ eine Markov-Kette mit Übergangswahrscheinlichkeiten

$$\begin{aligned} P(S_{n+1} = k + 1 \mid S_n = k) &= p, \\ P(S_{n+1} = k \mid S_n = k) &= 1 - p, \end{aligned} \quad n \in \mathbb{N}, \quad k \in \mathbb{N},$$

und alle anderen Wahrscheinlichkeiten sind 0.

Für die absoluten Wahrscheinlichkeiten gilt:

$$P^{S_0} = \delta_0, \quad P^{S_n} = b(n, p), \quad n \geq 1.$$

(Weitere Untersuchungen des Bernoulli-Prozesses und daraus abgeleiteter weiterer Prozesse findet man in [Çin75], S. 43.) \square

Die Berechnung von n -Schritt-Übergangswahrscheinlichkeiten war nach Korollar 3.9 durch Potenzierung von stochastischen Matrizen auszuführen. Dies ist im allgemeinen ein numerisch schwieriges Problem, da häufig sehr kleine Werte für die auftretenden bedingten Wahrscheinlichkeiten vorliegen. Ein einfaches Resultat erhält man aber im Fall $|E| = 2$ als Verallgemeinerung von Beispiel 3.15.

3.16 Beispiel

$X = (X_n : n \in \mathbb{N})$ sei eine Markov-Kette mit Zustandsraum $E = \{0, 1\}$ und Übergangsmatrix

$$p = \left(\begin{array}{c|cc} p(i, j) & 0 & 1 \\ \hline 0 & 1 - a & a \\ 1 & b & 1 - b \end{array} \right), \quad a, b \in [0, 1].$$

a) Falls $0 < a + b < 2$ gilt, ist

$$p^n = \frac{1}{a+b} \left(\left(\begin{array}{cc} b & a \\ b & a \end{array} \right) + (1-a-b)^n \left(\begin{array}{cc} a & -a \\ -b & b \end{array} \right) \right).$$

b) Falls $a = b = 0$ gilt, ist

$$p^n = p = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right).$$

c) Falls $a = b = 1$ gilt, sind

$$p^{2n} = \left(\begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right) \quad \text{und} \quad p^{2n+1} = p \quad \text{für } n \geq 0.$$

(Die Beweise sind durch Induktion zu führen.) \square

3.17 Beispiel (M/G/1-FCFS Bedienungssystem in diskreter Zeit)

Als Fortsetzung von Beispiel 3.14 nehmen wir an, daß in jedem Zeitintervall $[n, n+1)$ ein produziertes Stück kontrolliert wird, daß der Produktionsablauf in Ordnung ist und daß defekte Stücke zu einer Reparaturstation gebracht werden, wo sie zu Beginn des nächsten Zeitintervalles $[n+1, n+2)$ eintreffen.

Die Reparaturstation besteht aus einem Bedienungsschalter und einem (linearen) Warteraum unbegrenzter Kapazität. Findet ein ankommendes Gerät den Schalter leer vor, beginnt seine Reparatur sofort, ohne

Zeitverlust. Ist der Schalter belegt, so wird das Gerät in die Warteschlange, und zwar an deren Ende, eingeordnet (First-Come-First-Served). Wird eine Reparatur beendet, so wird das Gerät am Ende des gerade laufenden Zeitintervalls $[r, r + 1)$ aus dem Schalter genommen und es verläßt das System; sofern ein weiteres Gerät wartet, beginnt dessen Reparatur im Intervall $[r + 1, r + 2)$, $r \in \mathbb{N}$. Dabei wird stets das Gerät von der Schlängenspitze (welches schon am längsten im System ist) als nächstes bedient.

Die Zufälligkeit der Fehler bei korrekt arbeitender Produktion rechtfertigt die Annahme, daß die Reparaturzeiten unabhängig identisch verteilt und unabhängig vom Ankunftsstrom (der ein Bernoulli-Prozeß ist) sind; die Quantisierung des Reparaturvorgangs impliziert, daß die Reparaturzeiten diskret auf $\{1, 2, \dots\}$ verteilt sind. Es sei $\pi = (\pi(k) : k = 1, 2, \dots)$ die Zähldichte der Reparaturzeitverteilung.

Aus den am System (Kontrolle und Reparatur) gemachten Beobachtungen und den daraus gezogenen Folgerungen muß bei der Modellierung durch eine Markov-Kette begründet werden, warum dies Modell geeignet ist. (Daß ein Modell korrekt ist, kann in letzter Konsequenz nicht bewiesen werden!) Bei der Modellierung durch Markov-Ketten ergibt sich wie oben beschrieben das Problem: Welche Daten des Systems muß man zu einem Zeitpunkt $n \in \mathbb{N}$ kennen, um die weitere stochastische Entwicklung (über die Zeit $n + 1, n + 2, \dots$) ohne Rückgriff auf die Vergangenheit vor n zu bestimmen? Dabei wird nur die Kenntnis der zukünftigen Verteilungen gefordert — nicht etwa ein Wissen über tatsächliche Realisierungen.

Im vorliegenden Beispiel kann argumentiert werden:

- i) Notwendig ist sicher die Kenntnis der Schlängellänge zur Zeit n (dabei zählen wir stets das eventuell in Bedienung befindliche Stück mit).
- ii) Hinreichend ist die Kenntnis der Schlängellänge im allgemeinen nicht: Sei beispielsweise die Bedienungszeit-Zähldichte gegeben durch $\pi(3) = 1$. Sind zur Zeit $n \in \mathbb{N}$ genau 2 Geräte in der Reparaturstation, so wird eines von ihnen, nennen wir es kurz G , repariert, während das andere wartet.

Dann hängt die Wahrscheinlichkeit, daß G zur Zeit $(n + 1)(-)$ das System verläßt, davon ab, wie lange G schon bedient wurde bzw. (was äquivalent ist) wie viele Zeiteinheiten Bedienungszeit G noch zu bekommen hat. Als Lösung bietet sich also an:

Bei Bedienungsbeginn von G wird eine „Zusatzvariable“ auf 3 (mit Wahrscheinlichkeit 1) gesetzt und mit Ablauf jedes Zeitquantums um 1 heruntersetzt; hat die Zusatzvariable bei Beginn des Zeitintervalls $[n, n + 1)$ den Wert 1, so wissen wir, daß G am Ende des Quantums repariert ist, und daß dann das Gerät an der Schlängenspitze repariert wird, wobei die Zusatzvariable erneut mit Wahrscheinlichkeit 1 auf 3 gesetzt wird (bzw. bei allgemeiner Reparaturzeitverteilung π mit Wahrscheinlichkeit $\pi(k)$ auf k gesetzt wird).

Für die angestrebte „Markovisierung“ der Systembeschreibung werden wir also sowohl die Schlängellänge als auch die Restbedienungszeit bei Quantenbeginn notieren.

- iii) Notizen über den Eingangsstrom brauchen wir nicht mitzuführen, da er vom Bedienungsprozeß unabhängig ist und als Bernoulli-Prozeß in jedem Zeittakt Ankünfte unabhängig vom Rest des Systems und seiner Vergangenheit produziert.

Eine geeignete Markovsche Beschreibung der zeitlichen Entwicklung des Systems geschieht also mit dem Zustandsraum \mathbb{N}^2 , wobei $X_n(\omega) = (k, r)$ bedeutet, daß

- a) für $k \geq 1, r \geq 1$ noch k Geräte im System sind und der in Bedienung befindliche noch r restliche Bedienungszeitquanten fordert,
- b) für $k = 0, r = 0$ das System leer ist.

(Beachte: Nicht alle Elemente aus \mathbb{N}^2 sind sinnvolle Zustände — es ist aber einfach, einen minimalen Zustandsraum anzugeben.)

Nach Satz 3.7 genügt es (wenn wir uns entschieden haben, daß mit \mathbb{N}^2 als Zustandsraum eine Markovsche Beschreibung machbar ist), Startverteilung und 1-Schritt-Übergangsmatrix anzugeben.

Die Feststellung der Übergangswahrscheinlichkeit $p(\bullet, \bullet)$ geschieht wie folgt:

$$\begin{aligned} p((0,0), (0,0)) &= 1 - p \\ p((0,0), (1,r)) &= p \cdot \pi(r), & r \geq 1 \\ p((1,1), (0,0)) &= 1 - p \\ p((1,1), (1,r)) &= p \cdot \pi(r), & r \geq 1 \\ p((k,1), (k-1,r)) &= (1-p) \cdot \pi(r), & k \geq 2, r \geq 1 \\ p((k,1), (k,r)) &= p \cdot \pi(r), & k \geq 2, r \geq 1 \\ p((k,r), (k,r-1)) &= 1 - p, & k \geq 1, r \geq 2 \\ p((k,r), (k+1,r-1)) &= p, & k \geq 1, r \geq 2 \end{aligned}$$

und für alle anderen Zustandspaare

$$p(x, y) = 0, \quad x, y \in \mathbb{N}^2.$$

Die Festlegung der Startverteilung kann z. B. über $P^{X_0}\{(0,0)\} = 1$ geschehen, für viele Untersuchungen läßt man dies jedoch noch offen. \square

3.18 Anmerkung

In Beispiel 3.17 wäre eine Modellierung über die „erhaltene Bedienungszeit“ ebenfalls möglich. Eine Entscheidung, für welche Beschreibung mehr spricht, ist nicht allgemein zu fällen.

Im Modell für einen Erneuerungsprozeß als Alters- bzw. Restlebensdauerprozeß hatten wir die gleiche Wählbarkeit der Markovschen Beschreibung vorgefunden. Die „Markovisierung durch Zusatzvariable für Alter bzw. Restlebensdauer“ ist eine häufig verwendete (Standard-)Methode. \square

3.19 Beispiel (Irrfahrten)

- a) Irrfahrt auf \mathbb{Z} : Sei $(Y_n : n = 1, 2, \dots)$ eine unabhängig identisch verteilte Folge von Zufallsvariablen mit Werten in $\{-1, 1\}$ auf (Ω, \mathcal{F}, P) . Es gelte $P(Y_1 = 1) = \lambda \in [0, 1]$. Der Summenprozeß $S = (S_n : n \in \mathbb{N})$ auf (Ω, \mathcal{F}, P) mit Werten in \mathbb{Z} , definiert durch

$$S_0(\omega) = 0, \quad S_n(\omega) = \sum_{i=1}^n Y_i(\omega), \quad n \geq 1,$$

für $\omega \in \Omega$ heißt (klassische) Irrfahrt. S ist eine Markov-Kette mit Startverteilung p_0 als Einpunktverteilung in 0 und Übergangsmatrix $p = (p(i, j) : i, j \in \mathbb{Z})$ gegeben durch:

$$p(i, j) = \begin{cases} \lambda, & \text{für } j = i + 1 \\ 1 - \lambda, & \text{für } j = i - 1 \\ 0, & \text{sonst.} \end{cases}$$

S beschreibt die Bewegung eines Teilchens auf dem Gitter \mathbb{Z} , welches stets genau einen Schritt macht, das kein Gedächtnis hat und dessen Verhalten sich mit der Zeit (lokal) nicht ändert.

Für den Fall $\lambda = 0.5$ spricht man von symmetrischer Irrfahrt. Ist $\lambda > 0.5$, wird man eine „Drift“ des Teilchens nach $+\infty$ erwarten.

- b) Irrfahrt mit Rändern: Als Zustandsraum hat man eine Teilmenge

$$\{a, a + 1, a + 2, \dots, b - 1, b\} = E \subseteq \mathbb{Z}, \quad |a|, |b| < \infty.$$

Festzulegen bleibt das Verhalten der Irrfahrt an den Rändern. Zwei bedeutsame Fälle:

- i) absorbierende Ränder: $p(a, a) = p(b, b) = 1$
 ii) reflektierende Ränder: $p(a, a + 1) = \lambda, \quad p(a, a) = 1 - \lambda$
 $p(b, b - 1) = 1 - \lambda, \quad p(b, b) = \lambda$

Es gibt beliebige andere Formen und Mischformen.

c) Anwendungen der Verallgemeinerung mit

$$\left. \begin{aligned} p(i, i+1) &= \lambda, \\ p(i, i-1) &= \mu, \\ p(i, i) &= 1 - \lambda - \mu, \\ p(i, j) &= 0, \end{aligned} \right\} \begin{array}{l} i \in \mathbb{Z}, 0 < \lambda + \mu < 1, \\ \text{falls } |i - j| > 1 \end{array}$$

findet man in [BN84], S. 75 (Selektionsspiel).

d) Als Irrfahrten werden auch die Summenprozesse zu Folgen

$$Y_n : (\Omega, \mathcal{F}, P) \rightarrow (G, \mathcal{S}), \quad n \geq 1,$$

von unabhängig identisch verteilten Zufallsvariablen mit Werten in meßbaren (meist sogar topologischen) Halbgruppen bezeichnet. In diesem Sinne (mit $G = \mathbb{Z}$) ist die Theorie der Irrfahrten eine Verallgemeinerung der Erneuerungstheorie (wo $G = \mathbb{N}$ war). (Entsprechendes gilt für stetige Zufallsvariablen.) \square

In Lemma 3.3 war unter (3) gezeigt worden:

Sei $n \in \mathbb{N}$ ein vorgegebener Zeitpunkt, $M \in \sigma(X_t : t \geq n)$; dann gilt:

$$P(M \mid X_n = i_n, \dots, X_1 = i_1) = P(M \mid X_n = i_n) = E(1_M \mid X_n = i_n),$$

sofern die linke Seite erklärt ist. Dies kann formuliert werden als:

Falls $X_n = i_n$ gegeben ist, so sind die Zukunft von X nach n und die Vergangenheit von X vor n (bedingt) unabhängig. Daraus abgeleitet werden kann die folgende Aussage, die eine intuitiv einleuchtende Eigenschaft homogener Markov-Ketten ist:

Ist $f : E^{\mathbb{N}} \rightarrow \mathbb{R}$ eine meßbare, beschränkte Funktion und $Y := f(X_n, X_{n+1}, \dots)$, so gilt

$$E(Y \mid X_0, \dots, X_n) = E(Y \mid X_n) = E(f(X_n, X_{n+1}, \dots) \mid X_n) = E(f(X_0, X_1, \dots) \mid X_0).$$

(Dabei folgt die zweite Gleichheit aus der Homogenität von X .)

Dies kann als eine Regenerationseigenschaft angesehen werden:

Wenn zum (festen deterministischen) Zeitpunkt n gilt: $X_n = i$, so ist die Entwicklung von X nach n unter dieser Bedingung dieselbe wie die Entwicklung von X ist, falls zur Zeit 0 gilt: $X_0 = i$. Außerdem ist dann — wie oben schon bemerkt — unter $X_n = i$ die Entwicklung von X nach n unabhängig von der Entwicklung von X vor n .

Die Regenerationseigenschaft „Neustart mit identischem stochastischem Verhalten bei Erreichen des vorgegebenen Zustands, unabhängig vom bisherigen Ablauf“ hat sich in den letzten 15 Jahren als wirksames Hilfsmittel bei praktischen Untersuchungen herausgestellt. Insbesondere beruht darauf die sogenannte Technik der *regenerativen Simulation*. Diese kann leicht am Beispiel eines Wartesystems gemäß Beispiel 3.17 dargestellt werden.

Gesucht sei irgendein Leistungsparameter (z. B. Durchsatz, mittlere Schlängellänge, ...) des Bedienungssystems. Dazu starten wir das System leer, d. h. $X_0 = (0, 0)$ und setzen

$$\tau_0(0, 0) \equiv 0, \quad \tau_n(0, 0) = \inf\{k \in \mathbb{N} : k > \tau_{n-1}(0, 0), X_k = (0, 0)\} \quad \text{für } n \geq 1$$

als Folge der Eintrittszeiten von X in den Leerzustand.

Aus der Regenerationseigenschaft folgt, daß die Abschnitte $(X_{\tau_n}, X_{\tau_{n+1}}, \dots, X_{\tau_{n+1}-1})$ für $n \in \mathbb{N}$ (mit $\tau_n := \tau_n(0, 0)$) voneinander unabhängig und identisch verteilt sind.

In jedem „Regenerationsintervall“ wird jetzt das entsprechende Leistungsmaß geschätzt aufgrund der vorliegenden Realisation von X . Hat man dabei N derartige Schätzergebnisse, so bilden diese eine Realisierung von N unabhängig identisch verteilten Zufallsvariablen — und für derartige Stichproben hat man die üblichen statistischen Auswertungsverfahren zur Verfügung.

Diese Art von Auswertung findet ihre Hauptanwendung derzeit wohl bei der stochastischen Simulation komplexer Systeme. Ihre Anwendung wird mit den eben skizzierten Argumenten begründet. Die Korrektheit dieser Begründung ist allerdings noch nicht von uns gezeigt! Dies läßt sich wieder am obigen Beispiel klarmachen:

Sei $\tau := \tau_1 := \tau_1(0, 0)$ die erste Rückkehrzeit in den Leerzustand, $f : E^{\mathbb{N}} \rightarrow \mathbb{R}$ die meßbare, beschränkte Funktion, die aus den Werten $(X_\tau, X_{\tau+1}, X_{\tau+2}, \dots)$ den Schätzer für das gesuchte Leistungsmaß im zweiten Regenerationszyklus liefert. Dann haben wir (wie oben für festes n):

$$\mathbb{E}(f(X_\tau, X_{\tau+1}, \dots) \mid X_0, X_1, \dots, X_\tau) = \mathbb{E}(f(X_\tau, X_{\tau+1}, \dots) \mid X_\tau) = \mathbb{E}(f(X_0, X_1, \dots) \mid X_0)$$

und analog für die späteren Regenerationenblicke. Formuliert analog zu (M) über Übergangswahrscheinlichkeiten hätten wir daraus:

$$P(X_{\tau+1} = j \mid X_\tau = i_0, X_{\tau-1} = i_1, \dots) = P(X_{\tau+1} = j \mid X_\tau = i_0) = P(X_1 = j \mid X_0 = i_0)$$

für $j, i_0, i_1, \dots \in E$ und zufällige Zeiten $\tau : (\Omega, \mathcal{F}, P) \rightarrow (\overline{\mathbb{N}}, \mathcal{P}(\overline{\mathbb{N}}))$.

Dies ist natürlich eine viel stärkere Eigenschaft von X als (M)! — Wir werden als nächstes die korrekte, für zeitdiskrete Markov-Ketten beweisbare, „starke Markov-Eigenschaft“ formulieren und wegen der Bedeutung für Anwendungen auch beweisen.

3.20 Beispiel

Mit den Bezeichnungen des Beispiels 3.17 sei $\tau = \tau_1(0, 0)$ die erste Rückkehrzeit von X in den Leerzustand. σ sei die Zufallszeit, die definiert ist als derjenige Zeitpunkt, nach dem die erste Rückkehr nach $(0, 0)$ stattfindet: $\sigma = \tau - 1$. Dann gilt

$$P(X_{\sigma+1} = j \mid X_\sigma = i_0, X_{\sigma-1} = i_1, \dots) = P(X_{\sigma+1} = j \mid X_\sigma = i_0) = \begin{cases} 1, & \text{falls } j = (0, 0) \\ 0 & \text{sonst,} \end{cases}$$

und das ist natürlich in der Regel nicht gleich $P(X_1 = j \mid X_0 = i_0)$. □

Eine „starke Markov-Eigenschaft“ ist nach diesem Beispiel also nicht für alle Zufallszeiten zu beweisen. Analysiert man weitere Fälle, so ergibt sich, daß für σ in Beispiel 3.20 entscheidend ist:

Die Frage, ob σ zur Zeit $n \in \mathbb{N}$ eingetreten ist, läßt sich nur beantworten, wenn der Prozeß X , auf den sich σ bezieht, noch über den Zeitpunkt n hinaus beobachtet wird.

Derartige Zufallsvariablen werden wir als zulässige Zufallszeiten deshalb ausschließen. Für Anwendungen ist das erfahrungsgemäß keine wesentliche Einschränkung.

3.21 Definition

Sei (Ω, \mathcal{F}, P) ein Wahrscheinlichkeitsraum, $\Delta_0 \in \mathcal{F}$ und $Y : (\Delta_0, \Delta_0\mathcal{F}) \rightarrow \overline{\mathbb{N}}$ eine meßbare Abbildung. Sei $\Delta \in \Delta_0\mathcal{F}$ gegeben als $\Delta = \{\omega \in \Omega : Y(\omega) \in \mathbb{N}\}$. Dann heißt Δ_0 der Definitionsbereich der Zufallsvariablen Y mit Werten in $\overline{\mathbb{N}}$, Δ ihr Endlichkeitsbereich. (Keine näheren Angaben bedeutet: $\Delta = \Delta_0 = \Omega$. Auch für $\Omega \neq \Delta_0$ spricht man manchmal von einer Zufallsvariablen $Y : (\Omega, \mathcal{F}, P) \rightarrow \overline{\mathbb{N}}$ mit Definitionsbereich Δ_0 .)

3.22 Definition

$X = (X_n : n \in \mathbb{N})$ sei eine Markov-Kette auf (Ω, \mathcal{F}, P) mit Zustandsraum E . $\tau : (\Omega, \mathcal{F}, P) \rightarrow \overline{\mathbb{N}}$ sei eine Zufallsvariable mit Endlichkeitsbereich $\Delta \in \mathcal{F}$ und $P(\Delta) > 0$.

a) τ heißt Markov-Zeit für X (Stoppzeit für X , optional random variable relative to X), falls gilt:

$$\{\tau = n\} \in \sigma(X_k : k \leq n) \quad \text{für alle } n \in \mathbb{N}.$$

b) Die σ -Algebra aller Mengen $\Lambda \in \Delta\mathcal{F}$, für die gilt

$$\Lambda \cap \{\tau \leq n\} \in \sigma(X_k : k \leq n)$$

heißt Prä- τ -Algebra oder σ -Algebra der τ -Vergangenheit und wird mit \mathcal{F}_τ oder auch mit $\sigma(X_k : k \leq \tau)$ bezeichnet.

c) Der stochastische Prozeß $Y = (Y_n : n \in \mathbb{N})$, definiert durch

$$Y_n(\omega) = Y(n, \omega) = X(\tau(\omega) + n, \omega) \quad \text{für } \omega \in \Delta \text{ und } n \in \mathbb{N}$$

mit Definitionsbereich Δ heißt Post- τ -Prozeß bezüglich X . Die σ -Algebra $\sigma(Y_n : n \in \mathbb{N})$ heißt Post- τ -Algebra oder σ -Algebra der τ -Zukunft und wird mit $\sigma(X_k : k \geq \tau)$ bezeichnet.

3.23 Beispiel

$X = (X_n : n \in \mathbb{N})$ sei eine Markov-Kette mit Zustandsraum E . $G \subseteq E$ sei nichtleer und

$$\tau_e(G) : (\Omega, \mathcal{F}, P) \rightarrow \overline{\mathbb{N}}, \quad \omega \mapsto \begin{cases} \inf\{n \in \mathbb{N} : X_n(\omega) \in G\} & \text{falls das Infimum existiert,} \\ \infty & \text{sonst} \end{cases}$$

die Ersteintrittszeit von X in G . Dann ist $\tau_e(G)$ eine Markov-Zeit für X . Es gilt nämlich

$$\{\omega \in \Omega : \tau_e(G)(\omega) = n\} = \bigcap_{k=1}^{n-1} \{\omega \in \Omega : X_k(\omega) \notin G\} \cap \{\omega \in \Omega : X_n(\omega) \in G\} \in \sigma(X_k : k \leq n).$$

Anschaulich: Wir brauchen nur die von X bis zum Zeitpunkt n erzeugte Information berücksichtigen, um zu entscheiden, ob X zur Zeit n zum ersten Mal in G ist.

Falls G eine sogenannte „transiente Menge“ ist (mit derartigen Mengen werden wir uns später beschäftigen), so gilt:

$$P(X_n \notin G \quad \forall n \in \mathbb{N}) > 0;$$

das bedeutet aber

$$P(\tau_e(G) = \infty) > 0,$$

und damit können wir auf $\{\omega \in \Omega : \tau_e(G) = \infty\} =: \Delta_0$ die Zufallsvariable $\tau'_e(G)$, die zweite Eintrittszeit von X in G nicht definieren. Derartige Fälle sind mit der Festlegung in Definition 3.21 zwanglos zu behandeln. \square

3.24 Anmerkung

Mit den Bezeichnungen von Definition 3.22 gilt:

a) τ ist Markov-Zeit für X genau dann, wenn für alle $n \in \mathbb{N}$ gilt:

$$\{\tau \leq n\} \in \sigma(X_k : k \leq n).$$

b) Bezeichnet man mit $(\Delta, \Delta\mathcal{F}, P(\bullet | \Delta))$ die Reduktion von (Ω, \mathcal{F}, P) auf Δ , so ist der Post- τ -Prozeß Y ein stochastischer Prozeß auf $(\Delta, \Delta\mathcal{F}, P(\bullet | \Delta))$.

c) Eine explizite Beschreibung von \mathcal{F}_τ ist in der Regel nicht einfach, da (Ω, \mathcal{F}) unspezifiziert ist. Falls (Ω, \mathcal{F}) der sogenannte kanonische Raum ist (siehe [Fre83], S. 3, oder jedes Buch, in dem der Kolmogorowsche Existenzsatz bewiesen wird, der die Konstruktion des kanonischen Raumes benutzt), läßt sich \mathcal{F}_τ anschaulich machen. \square

3.25 Beispiel

$S = (S_n : n \in \mathbb{N})$ sei die in Beispiel 3.19 a) definierte Irrfahrt auf \mathbb{Z} mit $S_0 \equiv 0$. Dann sind mit $G = \{0\}$ gemäß Beispiel 3.23 $\tau_e(\{0\}) \equiv 0$ und $\tau_1 := \tau'_e(\{0\})$ die erste Rückkehrzeit nach 0 eine Markov-Zeit für S .

Sei $\Lambda := \{\omega \in \Omega : S_i(\omega) \leq H \quad \forall i < \tau_1(\omega)\}$ für vorgegebenes $H > 0$ das Ereignis

$$\Lambda \hat{=} \{\text{vor der ersten Rückkehr zur 0 überschreitet die Irrfahrt nicht die Höhe } H\}.$$

Wir zeigen, daß $\Lambda \in \mathcal{F}_{\tau_1}$ gilt. Für $n \in \mathbb{N}$ ist nämlich

$$\begin{aligned} \Lambda \cap \{\tau_1 \leq n\} &= \{\omega \in \Omega : S_i(\omega) \leq H \quad \forall i < \tau_1(\omega) \quad \wedge \quad \tau_1(\omega) \leq n\} \\ &= \bigcup_{k=1}^n \{\omega \in \Omega : S_i(\omega) \leq H \quad \forall i < \tau_1(\omega) \quad \wedge \quad \tau_1(\omega) = k\} \\ &= \bigcup_{k=1}^n \underbrace{\{\omega \in \Omega : S_i(\omega) \leq H \quad \forall i < k\}}_{\in \sigma(S_i : i \leq k)} \cap \underbrace{\{\omega \in \Omega : \tau_1(\omega) = k\}}_{\in \sigma(S_i : i \leq k)} \in \sigma(S_i : i \leq n). \quad \square \end{aligned}$$

3.26 Satz

Sei $\tau : (\Omega, \mathcal{F}, P) \rightarrow \overline{\mathbb{N}}$ eine Markov-Zeit für X mit Endlichkeitsbereich Δ und $P(\Delta) > 0$. Dann ist der Post- τ -Prozeß $Y = (Y_n : n \in \mathbb{N})$ eine Markov-Kette auf $(\Delta, \Delta\mathcal{F}, P(\bullet | \Delta))$ mit Zustandsraum E und Übergangsmatrix p . Die Startverteilung von Y ist gegeben durch

$$P(Y_0 = i | \Delta) = \sum_{n=0}^{\infty} P(\tau = n, X_n = i | \Delta) \quad \text{für } i \in E.$$

Es gilt außerdem für $\Lambda \in \sigma(X_k : k \leq \tau)$ und $0 = t_0 < t_1 < \dots < t_l$ sowie $i_0, i_1, \dots, i_l \in E$ mit $l \geq 0$:

$$P(\Lambda, Y_{t_\nu} = i_\nu, 0 \leq \nu \leq l) = P(\Lambda, Y_0 = i_0) \cdot \prod_{\nu=0}^{l-1} p^{(t_{\nu+1}-t_\nu)}(i_\nu, i_{\nu+1}). \quad (1)$$

Beweis:

I. Zum Beweis von (1):

i) Sei $\Lambda \in \sigma(X_k : k \leq \tau)$; dies ist genau dann der Fall, wenn

$$\Lambda_n := \Lambda \cap \{\tau \leq n\} \in \sigma(X_k : k \leq n), \quad n \in \mathbb{N},$$

gilt. Also gibt es für jedes $n \in \mathbb{N}$ ein $\Lambda_n \in \sigma(X_k : k \leq n)$ mit

$$\Lambda \cap \{\tau = n\} = \Lambda_n \cap \{\tau = n\}.$$

Sei $M \in \sigma(X_k : k \geq \tau)$. Dann gibt es für jedes $n \in \mathbb{N}$ ein $M_n \in \sigma(X_k : k \geq n)$ mit

$$M \cap \{\tau = n\} = M_n \cap \{\tau = n\}. \quad (2)$$

Wir zeigen dies mit dem Beweisprinzip für σ -Algebren:

Sei \mathcal{M} die Menge aller Elemente aus $\Delta\mathcal{F}$, welche die geforderte Eigenschaft haben.

$\alpha)$ \mathcal{M} enthält den Erzeuger

$$\left\{ \{\omega \in \Delta : X(\tau(\omega) + k) = i\} : i \in E, k \in \mathbb{N} \right\}$$

von $\sigma(X_k : k \geq \tau)$. Denn es gilt

$$\{\omega \in \Delta : X(\tau(\omega) + k) = i\} \cap \{\tau = n\} = \underbrace{\{\omega \in \Delta : X(n+k) = i\}}_{\in \sigma(X_k : k \geq n)} \cap \{\tau = n\}.$$

$\beta)$ \mathcal{M} ist eine σ -Algebra:

a. $\emptyset \in \mathcal{M}$, da $\emptyset \in \sigma(X_k : k \geq n)$ ist.

b. Sei $M \in \mathcal{M}$, $M_n \in \sigma(X_k : k \geq n)$ und sei (2) erfüllt. Dann ist $M_n^c \in \sigma(X_k : k \geq n)$ und es sind

$$(M \cap \{\tau = n\}) + (M^c \cap \{\tau = n\}) = (M_n \cap \{\tau = n\}) + (M_n^c \cap \{\tau = n\})$$

zwei disjunkte Zerlegungen von $\{\tau = n\}$. Da nach Voraussetzung die beiden ersten Summanden gleich sind, gilt Gleichheit auch der beiden zweiten Summanden.

c. Die Abgeschlossenheit von \mathcal{M} unter abzählbarer Vereinigung sieht man direkt.

ii) Für $\Lambda \in \sigma(X_k : k \leq \tau)$ und $M \in \sigma(X_k : k \geq \tau)$ gilt damit

$$P(\Lambda \cap M) = \sum_{n=0}^{\infty} \sum_{i \in E} P(\Lambda; \tau = n; X_\tau = i; M) = \sum_{n=0}^{\infty} \sum_{i \in E} P(\Lambda_n; \tau = n; X_n = i; M_n),$$

wobei die $\Lambda_n \in \sigma(X_k : k \leq n)$ und $M_n \in \sigma(X_k : k \geq n)$ entsprechend i) gewählt sind. Dies führt auf

$$\begin{aligned} P(\Lambda \cap M) &= \sum_{n=0}^{\infty} \sum_{i \in E} P(\Lambda_n; \tau = n; X_n = i) \cdot P(M_n | X_n = i; \tau = n; \Lambda_n) \\ &= \sum_{n=0}^{\infty} \sum_{i \in E} P(\Lambda; \tau = n; X_n = i) \cdot P(M_n | X_n = i), \end{aligned} \quad (3)$$

wobei die letzte Gleichung aus der Markov-Eigenschaft in der folgenden Form folgt: Für festes $n \in \mathbb{N}$ seien $A \in \sigma(X_k : k \geq n)$ und $B \in \sigma(X_k : k \leq n)$ sowie $i \in E$; dann gilt:

$$P(A | X_n = i, B) = P(A | X_n = i).$$

(Setze $B := \Lambda_n \cap \{\tau = n\}$.)

iii) Setzt man jetzt speziell

$$M = \bigcap_{\nu=1}^l \{Y_{t_\nu} = i_\nu\} \in \sigma(X_k : k \geq \tau),$$

so hat man entsprechend i) $\alpha)$

$$M_n = \bigcap_{\nu=1}^l \{X_{n+t_\nu} = i_\nu\} \in \sigma(X_k : k \geq n)$$

und es gilt (unabhängig von n wegen der Homogenität)

$$P(M_n | X_n = i_0) = \prod_{\nu=0}^{l-1} p^{(t_{\nu+1}-t_\nu)}(i_\nu, i_{\nu+1}).$$

Mit den speziellen M, M_n gehen wir in (3) und erhalten

$$\begin{aligned} P(\Lambda, Y_{t_\nu} = i_\nu, 1 \leq \nu \leq l) &= \sum_{i_0 \in E} \prod_{\nu=0}^{l-1} p^{(t_{\nu+1}-t_\nu)}(i_\nu, i_{\nu+1}) \sum_{n=0}^{\infty} P(\Lambda; \tau = n; X_n = i_0) \\ &= \sum_{i_0 \in E} \prod_{\nu=0}^{l-1} p^{(t_{\nu+1}-t_\nu)}(i_\nu, i_{\nu+1}) P(\Lambda; Y_0 = i_0). \end{aligned} \quad (4)$$

Da $\{Y_0 = i_0\} \in \sigma(X_k : k \leq \tau)$ ist, folgt auch $\Lambda \cap \{Y_0 = i_0\} \in \sigma(X_k : k \leq \tau)$. Ersetzen wir in (4) Λ durch $\Lambda \cap \{Y_0 = i_0\}$, so haben wir gerade (1).

II. Setzen wir in (1) $\Lambda = \Delta$ und dividieren durch $P(\Delta)$, so ergibt sich nach Bildung der bedingten Wahrscheinlichkeit, daß Y Markov-Kette auf $(\Delta, \Delta\mathcal{F}, P(\bullet | \Delta))$ mit Zustandsraum E ist. ■

3.27 Korollar

Mit den Bezeichnungen und Annahmen aus Satz 3.26 sei zusätzlich für ein $i_0 \in E$: $P(Y_0 = i_0 | \Delta) = 1$. Dann gilt:

$$P(\Lambda, Y_{t_\nu} = i_\nu, 1 \leq \nu \leq l) = P(\Lambda) \prod_{\nu=0}^{l-1} p^{(t_{\nu+1}-t_\nu)}(i_\nu, i_{\nu+1}) \quad (1)$$

und in allgemeinerer Form

$$P(\Lambda \cap M) = P(\Lambda)P(M | \Delta). \quad (2)$$

Beweis: $P(Y_0 = i_0 | \Delta) = 1$ bedeutet $P(Y_0 = i_0, \Delta) = P(\Delta)$. Aus $\Lambda \in \sigma(X_k : k \leq \tau) \subseteq \Delta\mathcal{F}$ folgt daher

$$P(\Lambda, Y_0 = i_0) = P(\Lambda, Y_0 = i_0, \Delta) = P(\Lambda\Delta) = P(\Lambda).$$

Für alle $i \neq i_0$ gilt $P(\Lambda, Y_0 = i, \Delta) = 0$, so daß die Formel (4) in Satz 3.26 schon die hier gesuchte Gestalt (1) hat. Für den Fall $M = \{Y_{t_\nu} = i_\nu, 1 \leq \nu \leq l\}$ ist dies auch schon (2), denn da Y eine auf $(\Delta, \Delta\mathcal{F}, P(\bullet | \Delta))$ definierte Markov-Kette ist, gilt

$$\begin{aligned} P(Y_{t_\nu} = i_\nu, 1 \leq \nu \leq l | \Delta) &= P(Y_{t_\nu} = i_\nu, 1 \leq \nu \leq l, Y_0 = i_0 | \Delta) \\ &= P(Y_0 = i_0 | \Delta) \prod_{\nu=0}^{l-1} p^{(t_{\nu+1}-t_\nu)}(i_\nu, i_{\nu+1}). \end{aligned}$$

Um für festes (aber beliebiges) Λ die Formel (2) für alle M zu zeigen, wenden wir den Eindeutigkeitsatz für (endliche) Maße auf die Funktionen $P(\Lambda \cap (\bullet))$ und $P(\Lambda)P(\bullet | \Delta)$ auf $\sigma(X_k : k \geq \tau)$ an, die auf einem \cap -stabilen Erzeuger dieser Post- τ -Algebra nach (1) übereinstimmen. ■

3.28 Korollar

Mit den Bezeichnungen und Annahmen aus Satz 3.26 sei zusätzlich für ein $i_0 \in E$: $P(Y_0 = i_0 | \Delta) = 1$. Ist dann außerdem $P(\Delta) = 1$, d. h. $\tau < \infty$ P -fast sicher, so sind die Ereignisse aus der τ -Vergangenheit und aus der τ -Zukunft stochastisch unabhängig, gegeben die τ -Gegenwart $\{Y_0 = X_\tau = i_0\}$.

Die Korollare 3.27, 3.28 sind die entscheidenden Schritte auf dem Weg zur „starken Markov-Eigenschaft“, die wir auf S. 19 ff. über eine Regenerationseigenschaft zur Erneuerung von Markov-Ketten nach gewissen Markov-Zeiten diskutiert hatten: Sie besagen gerade, daß jede Markov-Kette (in unserem Sinne gemäß Definition 3.1 und homogen) „stark Markovsch“ ist.

Setzt man in Gleichung (1) von Korollar 3.27 $\Lambda = \Omega$, so erhalten wir, daß nach jedem Eintritt in einen vorgegebenen Zustand $i_0 \in E$ die Kette sich so entwickelt, wie sie es täte, wenn sie zur Zeit 0 in i_0 gestartet wird.

Wird dann X in $i_0 \in E$ gestartet, so ist der Block $(X_0, X_1, \dots, X_{\tau_1(\{i_0\})-1})$ stochastisch unabhängig vom zweiten Block $(X_{\tau_1(\{i_0\})+1}, \dots, X_{\tau_2(\{i_0\})-1})$, sofern die erste Rückkehrzeit nach i_0 mit Wahrscheinlichkeit 1 endlich ist.

Daß diese Blockbildung iteriert werden kann, erhalten wir jetzt direkt. Wir formulieren wieder mittels der erzeugten σ -Algebren und in etwas allgemeinerem Kontext.

3.29 Satz

Die Zufallsvariablen τ_n , $n = 1, 2, \dots, l$, seien Markov-Zeiten für X mit Endlichkeitsbereichen Δ_n derart, daß gilt: $\Delta_n \supseteq \Delta_{n+1}$ und $\tau_n < \tau_{n+1}$ auf Δ_{n+1} . Für jedes $n \in \{1, \dots, l\}$ gebe es ein $i_n \in E$, so daß gilt

$$P\left(\{\omega \in \Omega : X_{\tau_n(\omega)}(\omega) = i_n\} \mid \Delta_n\right) = 1.$$

Seien Λ_n , $n = 1, \dots, l+1$, gegeben mit

$$\Lambda_n \in \sigma(X_k : k \geq \tau_{n-1}) \cap \sigma(X_k : k \leq \tau_n)$$

(wobei formal $\tau_0 = 0$ und $\tau_{l+1} = \infty$ sei). Dann gilt

$$P\left(\bigcap_{n=1}^{l+1} \Lambda_n\right) = \frac{\prod_{n=1}^{l+1} P(\Lambda_n)}{\prod_{n=1}^l P(\Delta_n)}.$$

Sind außerdem die Markov-Zeiten $\tau_1, \tau_2, \dots, \tau_l$ mit Wahrscheinlichkeit 1 endlich, so sind die $\Lambda_1, \dots, \Lambda_{l+1}$ stochastisch unabhängig.

Beweis: Sukzessives Anwenden von Korollar 3.27 ergibt (mit $\Delta_0 = \Omega$)

$$P\left(\bigcap_{n=1}^{l+1} \Lambda_n\right) = \prod_{n=1}^{l+1} P(\Lambda_n \mid \Delta_0 \Delta_1 \dots \Delta_{n-1}).$$

Die Behauptung folgt mit $\bigcap_{j=0}^{n-1} \Delta_j = \Delta_{n-1}$. ■

3.30 Lemma

X sei eine Markov-Kette, $A \subseteq E$, $* \notin E$; es sei

$$\alpha(\omega) = \begin{cases} \inf\{n \geq 0 : X_n(\omega) \in A\}, & \text{falls ein derartiges } n \in \mathbb{N} \text{ existiert,} \\ \infty & \text{sonst.} \end{cases}$$

Dann ist α eine Markov-Zeit für X , und wir können eine Markov-Kette $Z = (Z_n : n \in \mathbb{N})$ definieren durch

$$Z_n(\omega) = \begin{cases} X_n(\omega) & \text{für } 0 \leq n < \alpha(\omega), \\ * & \text{für } \alpha(\omega) \leq n. \end{cases}$$

Z ist auf (Ω, \mathcal{F}, P) definiert mit Zustandsraum $(E - A) \cup \{*\}$ und Übergangsmatrix $Q = (q(i, j) : i, j \in (E - A) \cup \{*\})$, gegeben durch:

$$q(k, j) = \begin{cases} p(k, j), & \text{falls } k \neq *, j \neq *; \\ \sum_{l \in A} p(k, l), & \text{falls } k \neq *, j = *; \\ \delta_{*j}, & \text{falls } k = *. \end{cases}$$

Z beschreibt das Verhalten von X bis zum Eintritt in A , danach wird Z in $*$ absorbiert. Es gilt: $P(\alpha > m) = P(Z_m \neq *)$, $m \in \mathbb{N}$, d. h. wir können die Verteilung der Ersteintrittszeit in A berechnen über die Übergangswahrscheinlichkeiten von Z .

3.21–3.29 entsprechen dem Kapitel I, § 13 in [Chu67] und sind dort als „Systemtheorems“ abgehandelt.

Die im folgenden eingeführten Begriffe und bewiesenen Sätze werden sich auf Eigenschaften Markovscher Ketten beziehen, die durch ihre Übergangsmatrix (im wesentlichen) bestimmt sind. Um die Darstellung anschaulich zu machen, bietet es sich aber an — und dies wird hier und in anderen Darstellungen stets getan — die Sprache stochastischer Prozesse zu verwenden. Entsprechend Satz 3.7 und der daran anschließenden Bemerkungen heißt dies, daß wir — ohne weitere Erwähnung — Ausführungen über verschiedene Markov-Ketten gleichzeitig machen, sofern diese die gleiche Übergangsmatrix besitzen.

Ein Beispiel:

$$P(X_{m+n} = j \mid X_m = i) = \pi$$

soll eine Aussage über die Übergangsmatrix der entsprechenden Markov-Kette sein. Für eine konkrete Markov-Kette, die durch eine Startverteilung festgelegt wird, kann jedoch $P(X_m = i) = 0$ sein, der obige Ausdruck streng genommen also nicht definiert sein, während unter anderer Startverteilung die absolute Wahrscheinlichkeit, zur Zeit m in i zu sein, größer Null sein mag. Weiter ist nun aufgrund der Homogenität

$$\pi = P(X_n = j \mid X_0 = i) = p^{(n)}(i, j),$$

und wir interpretieren dies in einer Markov-Kette mit Startverteilung $P(X_0 = i) = 1$; analog die obige Darstellung von π in einer Markov-Kette mit $P(X_m = i) > 0$. — In der folgenden Definition geben wir der Vollständigkeit halber die entsprechende Voraussetzung noch an.

Ziel ist dabei die genauere Untersuchung von Rückkehrzeiten und Eintrittszeiten.

3.31 Definition

Sei $X = (X_n : n \in \mathbb{N})$ eine Markov-Kette, die mit Wahrscheinlichkeit 1 in $i \in E$ gestartet werde. Für jedes $\omega \in \Omega$ sei die (endliche oder unendliche) Folge $\tau_0(i, \omega), \tau_1(i, \omega), \tau_2(i, \omega), \dots$ definiert durch:

$$\tau_0(i, \omega) = 0 \quad \forall i \in E, \omega \in \Omega;$$

Sei $\tau_{n-1}(i, \omega)$ definiert und Δ_{n-1} der Endlichkeitsbereich von $\tau_{n-1}(i, \omega)$, $n \geq 1$; dann ist für $\omega \in \Delta_{n-1}$

$$\tau_n(i, \omega) = \begin{cases} \inf\{m \in \mathbb{N} : m > \tau_{n-1}(i, \omega), X_m(\omega) = i\}, & \text{falls das Infimum existiert;} \\ \infty & \text{sonst.} \end{cases}$$

Dann heißt

$$\tau_n(i) : (\Delta_{n-1}, \Delta_{n-1}\mathcal{F}, P(\bullet | \Delta_{n-1})) \rightarrow \overline{\mathbb{N}}, \quad n \geq 1,$$

n -te Eintrittszeit in den Zustand $i \in E$. Die Zufallsvariable

$$\varrho_n(i) : (\Delta_{n-1}, \Delta_{n-1}\mathcal{F}, P(\bullet | \Delta_{n-1})) \rightarrow \overline{\mathbb{N}}, \quad \omega \mapsto \tau_n(i, \omega) - \tau_{n-1}(i, \omega), \quad n \geq 1,$$

heißt n -te Rückkehrzeit nach $i \in E$, und das zufällige Intervall $[\tau_n(i), \tau_{n+1}(i))$ heißt $(n+1)$ -tes Rückkehrintervall nach $i \in E$, $n \geq 0$. Siehe auch Abbildung 4.

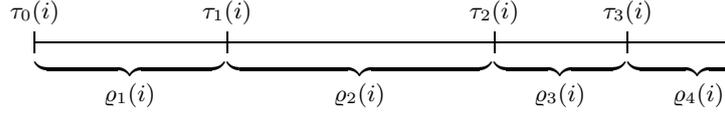


Abbildung 4: Eintrittszeiten τ_n und Rückkehrzeiten ϱ_n einer Markov-Kette

3.32 Lemma

Die Eintrittszeiten $\tau_n(i)$, $n \in \mathbb{N}$, sind Markov-Zeiten für X , $i \in E$, und es gilt für die Rückkehrzeitverteilungen

$$P(\varrho_n(i) = k | \Delta_{n-1}) = P(\varrho_1(i) = k | X_0 = i), \quad k \in \mathbb{N}, \quad n \in \mathbb{N}_+.$$

Ist außerdem

$$P(X_0 = i) = 1 \quad \text{und} \quad P(\tau_1(i) < \infty) = 1,$$

so ist die Folge $(\varrho_n(i) : n \in \mathbb{N}_+)$ unabhängig identisch verteilt. Weiter ist dann die Folge der i -Blöcke

$$\left\{ (X_{\tau_n(i)}, X_{\tau_n(i)+1}, \dots, X_{\tau_{n+1}(i)-1}) : n \in \mathbb{N} \right\}$$

unabhängig identisch verteilt.

Beweis:

a) Für $r \geq 1$ und $n \geq 1$ gilt

$$\{\tau_n(i) = r\} = \{X_r = i\} \cap \left\{ \sum_{h=1}^r 1_{\{i\}}(X_h) = n \right\} \in \sigma(X_k : k \leq r),$$

also ist $\tau_n(i)$ Markov-Zeit für X .

b) Für $r \geq 1$ ist

$$\{\varrho_n(i) = r\} = \{X_{\tau_{n-1}(i)+1} \neq i, \dots, X_{\tau_{n-1}(i)+r-1} \neq i, X_{\tau_{n-1}(i)+r} = i\}$$

ein Ereignis aus $\sigma(X_k : k \geq \tau_{n-1}(i)) \cap \sigma(X_k : k \leq \tau_n(i))$; außerdem gilt auf dem Endlichkeitsbereich Δ_{n-1} von $\tau_{n-1}(i)$: Der Post- $\tau_{n-1}(i)$ -Prozeß startet mit Wahrscheinlichkeit 1 in i . Damit liefert Gleichung (1) aus Korollar 3.27 die zweite Behauptung.

c) Ist $\varrho_1(i) = \tau_1(i)$ mit Wahrscheinlichkeit 1 endlich, so folgt aus dem eben gezeigten sukzessiv, daß alle $\varrho_1(i)$, $\varrho_2(i)$, \dots mit Wahrscheinlichkeit 1 endlich sind, damit auch alle $\tau_n(i)$, $n \geq 0$; damit kann Satz 3.29 angewendet werden, da wie oben gezeigt

$$\{\varrho_n(i) = r\} \in \sigma(X_k : k \geq \tau_{n-1}(i)) \cap \sigma(X_k : k \leq \tau_n(i))$$

gilt. Die (allgemeinere) Aussage über die i -Blöcke folgt genauso. ■

3.33 Korollar

Sei X eine Markov-Kette und $p_0 = P^{X_0}$ so gewählt, daß gilt: $P(X_n = i \text{ für ein } n \in \mathbb{N}) > 0$. Bezeichnen wir mit $\tau_0(i)$ die Ersteintrittszeit von X in i und mit $\Delta_0 = \{\omega \in \Omega : \tau_0(i) < \infty\}$ den Endlichkeitsbereich von $\tau_0(i)$, so kann X , restringiert auf $(\Delta_0, \Delta_0\mathcal{F}, P(\bullet | \Delta_0))$, definiert werden, und der Post- $\tau_0(i)$ -Prozeß $Y = (Y_n = X_{\tau_0(i)+n} : n \in \mathbb{N})$ kann gemäß Definition 3.31 strukturiert werden. Lemma 3.32 gilt dann entsprechend für Y .

3.34 Definition

- a) Die Wahrscheinlichkeit, in genau $n \geq 1$ Schritten zum ersten Mal von $i \in E$ nach $j \in E$ zu gelangen, sei

$$f_{ij}^{(n)} = P(X_n = j, X_k \neq j, 0 \leq k < n | X_0 = i).$$

Die Wahrscheinlichkeit, überhaupt von i nach j in endlicher Zeit zu gelangen, sei

$$f_{ij}^* := \sum_{n=1}^{\infty} f_{ij}^{(n)} = P(X_n = j \text{ für ein } n > 0 | X_0 = i) = P\left(\bigcup_{n=1}^{\infty} \{X_n = j\} \mid X_0 = i\right).$$

$(f_{ij}^{(n)} : n = 1, 2, \dots)$ ist also die (Zähldichte der) Verteilung der Ersteintrittszeit in j bei Start in i , $(f_{ii}^{(n)} : n = 1, 2, \dots)$ die (Zähldichte der) Verteilung von $\varrho_1(i)$.

Es sei $m_{ij} = E(f_{ij}^{(n)} : n \in \mathbb{N}_+)$ die erwartete Ersteintrittszeit in j bei Start in i , insbesondere also $m_{ii} = E(\varrho_1(i))$.

- b) Die Wahrscheinlichkeit, bei Start in i unendlich oft nach j zu gelangen, sei

$$\begin{aligned} P(X_n = j \text{ i. o.} | X_0 = i) &:= P\left(\limsup_{n \rightarrow \infty} \{\omega \in \Omega : X_n(\omega) = j\} \mid X_0 = i\right) \\ &= P\left(\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} \{\omega \in \Omega : X_n(\omega) = j\} \mid X_0 = i\right). \end{aligned}$$

(„i. o.“ steht für „infinitely often“.)

- c) Ein Zustand $i \in E$ heißt
- $$\left\{ \begin{array}{l} \text{transient} \iff f_{ii}^* < 1 \\ \text{rekurrent} \iff f_{ii}^* = 1 \\ \text{positiv rekurrent} \iff m_{ii} < \infty \\ \text{nullrekurrent} \iff m_{ii} = \infty \end{array} \right.$$

Die Begriffe Transienz und Rekurrenz sind zentral für die Untersuchung des Stabilitätsverhaltens und des asymptotischen Verhaltens von Markov-Ketten. Ihrem Studium werden wir uns zunächst widmen. Dafür ist die sogenannte „Ersteintrittsmethode“ ein wirksames Hilfsmittel. Bei dieser handelt es sich um eine Version der Formel von der totalen Wahrscheinlichkeit, welche die Wahrscheinlichkeit gesuchter Ereignisse zerlegt gemäß dem ersten Eintreten eines aufspaltenden anderen Ereignisses.

Da es sich um eine Verallgemeinerung des sogenannten „Erneuerungsargumentes“ der Erneuerungstheorie handelt, werden die dabei aufgestellten Gleichungen auch „Markovsche Erneuerungsgleichungen“ genannt. Wir zeigen das Vorgehen beispielhaft:

3.35 Lemma

Für $i, j \in E$ gilt:

$$P(X_n = j \text{ i. o.} | X_0 = i) = f_{ij}^* \cdot P(X_n = j \text{ i. o.} | X_0 = j),$$

d. h. die Wahrscheinlichkeit, „bei Start in i unendlich oft j zu erreichen“, ist das Produkt der Wahrscheinlichkeit, „in endlich vielen Schritten von i nach j zu gelangen“, mit der Wahrscheinlichkeit, „unendlich oft von j nach j zu gelangen“.

Beweis:

$$\begin{aligned}
P(X_n = j \text{ i. o.} \mid X_0 = i) &= \sum_{n=1}^{\infty} P\left(\begin{array}{l} X_k \neq j, 1 \leq k \leq n-1, X_n = j, \\ X_l = j \text{ für unendlich viele } l > n \end{array} \mid X_0 = i \right) \\
&= \sum_{n=1}^{\infty} P\left(X_l = j \text{ für unendlich viele } l > n \mid \begin{array}{l} X_k \neq j, 1 \leq k \leq n-1, X_n = j, \\ X_0 = i \end{array} \right) \\
&\quad \cdot P(X_k \neq j, 1 \leq k \leq n-1, X_n = j \mid X_0 = i) \\
&= \sum_{n=1}^{\infty} P(X_l = j \text{ für unendlich viele } l > n \mid X_n = j) \cdot f_{ij}^{(n)} \\
&= \sum_{n=1}^{\infty} P(X_l = j \text{ i. o.} \mid X_0 = j) \cdot f_{ij}^{(n)} \\
&= P(X_n = j \text{ i. o.} \mid X_0 = j) \cdot f_{ij}^*.
\end{aligned}$$

(Wir haben also das interessierende Ereignis aufgespalten nach disjunkten Ereignissen, deren Wahrscheinlichkeit bekannt (oder einfacher) sind und unter deren Bedingung die bedingte Wahrscheinlichkeit des interessierenden Ereignisses bekannt (oder einfacher) ist.) ■

3.36 Korollar

Sei $i \in E$. Dann gilt:

$$\begin{aligned}
i \text{ rekurrent} &\iff P(X_n = i \text{ i. o.} \mid X_0 = i) = 1, \\
i \text{ transient} &\iff P(X_n = i \text{ i. o.} \mid X_0 = i) = 0,
\end{aligned}$$

Beweis:

a) Sei i transient, nach Definition also $f_{ii}^* < 1$. Aus Lemma 3.35 mit $i = j$ folgt

$$P(X_n = j \text{ i. o.} \mid X_0 = j) = 0$$

(zu unterscheiden ist dabei $f_{ii}^* = 0$ und $f_{ii}^* > 0$).

b) Sei i rekurrent, also $f_{ii}^* = 1$. Mit den Bezeichnungen aus Definition 3.31 ist $(f_{ii}^* : n \in \mathbb{N})$ die Zähldichte der Rückkehrzeit von i nach i , und für $n \geq 1$ ist

$$\tau_n(i) = \sum_{k=1}^n \varrho_k(i).$$

Da jedes $\varrho_k(i)$ mit Wahrscheinlichkeit 1 endlich ist, findet also auch *jede* Rückkehr nach i mit Wahrscheinlichkeit 1 in endlicher Zeit statt. Angenommen, es wäre

$$P(X_n = i \text{ i. o.} \mid X_0 = i) < 1.$$

Dann wäre also

$$P(X_n = i \text{ nurendlichoft} \mid X_0 = i) > 0,$$

und damit für mindestens ein endliches $k \in \mathbb{N}$

$$P(X_n = i \text{ fürgenaukZeitpunkten} \mid X_0 = i) = P(\rho_1(i) + \rho_2(i) + \dots + \rho_k(i) = \infty \mid X_0 = i) > 0.$$

Das ist aber ein Widerspruch zu

$$P(\tau_n(i) < \infty \mid X_0 = i) = 1, \quad \forall n \in \mathbb{N}.$$

■

Eine weitere beliebte Ersteintrittsgleichung ist

3.37 Lemma

Für $i, j \in E$ und $n \geq 1$ gilt:

$$p^{(n)}(i, j) = \sum_{\nu=1}^n f_{ij}^{(\nu)} p^{(n-\nu)}(j, j).$$

Häufig sind die Verteilungen der Rückkehrzeiten nicht direkt zugänglich. Ein manchmal verwertbares Kriterium zur Prüfung auf Rekurrenz oder Transienz liefert der folgende Satz, dessen direkter Beweis länglich ist (siehe [Chu67], S. 22–23).

3.38 Satz

Ein Zustand $i \in E$ ist rekurrent genau dann, wenn die Reihe $\sum_{n=0}^{\infty} p^{(n)}(i, i)$ divergiert. Ist $i \in E$ transient, so

gilt $\sum_{n=0}^{\infty} p^{(n)}(i, i) = \frac{1}{1 - f_{i,i}^*}$. Außerdem ist die Summe $\sum_{n=0}^{\infty} p^{(n)}(i, i)$ die erwartete Anzahl von Aufenthalten von X in i .

3.39 Beispiel

a) Eindimensionale Irrfahrt auf \mathbb{Z} (nach Beispiel 3.19):

Für $n \in \mathbb{N}$ gilt:

$$p^{(2n+1)}(0, 0) = 0, \quad p^{(2n)}(0, 0) = \binom{2n}{n} \lambda^n (1-\lambda)^n = \binom{-1/2}{n} (-4\lambda(1-\lambda))^n.$$

(Dabei folgt die Wahrscheinlichkeit der $p^{(2n)}(0, 0)$ aus den folgenden Überlegungen: Jeder Pfad von 0 nach 0 in $2n$ Schritten hat die Wahrscheinlichkeit $\lambda^n (1-\lambda)^n$ und $\binom{2n}{n}$ ist die Anzahl dieser Pfade; die zweite Darstellung zeigt man durch Induktion.)

Wir haben zu untersuchen:

$$\sum_{n=0}^{\infty} \binom{2n}{n} \lambda^n (1-\lambda)^n = \sum_{n=0}^{\infty} \binom{-1/2}{n} (-4\lambda(1-\lambda))^n.$$

Für den Fall $\lambda \neq \frac{1}{2}$, d. h. $4\lambda(1-\lambda) < 1$, konvergiert die Reihe und hat (mittels Newtons Binomialformel) den Wert

$$(1 - 4\lambda(1-\lambda))^{-\frac{1}{2}} = |1 - 2\lambda|^{-1}.$$

Der Zustand 0 ist dann also transient.

Den Fall $\lambda = \frac{1}{2}$ erhält man durch Grenzübergang in der folgenden Weise:

$$\sum_{n=0}^{\infty} \binom{2n}{n} \left(\frac{1}{4}\right)^n \geq \sum_{n=0}^{\infty} \binom{2n}{n} (\lambda(1-\lambda))^n = (1 - 4\lambda(1-\lambda))^{-\frac{1}{2}} \xrightarrow[\lambda \in (0,1), \lambda \neq \frac{1}{2}]{\lambda \rightarrow \frac{1}{2}} \infty.$$

Der Zustand 0 ist dann also rekurrent.

Um zu entscheiden, ob positive Rekurrenz vorliegt, muß die erwartete Rückkehrzeitverteilung untersucht werden. Durch mehrfache Anwendung der starken Markov Eigenschaft und der Symmetrieeigenschaft der Irrfahrt, die aus $\lambda = \frac{1}{2}$ folgt, erhalten wir mit Ersteintrittsargumenten

$$E[\tau_1(0) \mid X_0 = 0] = 1 + E[\tau_1(0) \mid X_0 = 1]$$

und

$$E[\tau_1(0) \mid X_0 = 1] = \frac{1}{2} \cdot 1 + \frac{1}{2}(1 + E[\tau_1(0) \mid X_0 = 2]),$$

$$E[\tau_1(0) \mid X_0 = 2] = E[\tau_1(1) \mid X_0 = 2] + E[\tau_1(0) \mid X_0 = 1] = 2 \cdot E[\tau_1(0) \mid X_0 = 1],$$

$$E[\tau_1(0) \mid X_0 = 0] = 2 + E[\tau_1(0) \mid X_0 = 1].$$

Die erste und die letzte Gleichung sind nur verträglich für

$$E[\tau_1(0) \mid X_0 = 1] = \infty,$$

was auch

$$E[\tau_1(0) \mid X_0 = 0] = \infty$$

nach sich zieht. (Genauer zur Rückkehrzeitverteilung siehe [Hin72], S. 59.) Der Zustand 0 ist in der symmetrischen Irrfahrt auf \mathbb{Z} also nullrekurrent. Aus Symmetriegründen muß dies für alle Zustände gelten.

Ebenfalls aus Symmetriegründen sind in der nicht-symmetrischen Irrfahrt auf \mathbb{Z} alle Zustände transient.

- b) Für symmetrische Irrfahrten in höheren Dimensionen, d. h. Markov-Ketten auf \mathbb{Z}^d , $d \geq 2$, die sich bei jedem Schritt in jede mögliche Koordinatenrichtung (um einen Schritt) mit gleicher Wahrscheinlichkeit bewegen, kann man zeigen:

Für $d = 2$ liegt wieder Nullrekurrenz vor, für $d \geq 3$ aber Transienz (vgl. [KT75], S. 67). \square

3.40 Beispiel

Ist der Zustandsraum einer Markov-Kette endlich, so existiert mindestens ein positiv rekurrenter Zustand.

Beweis: Irgendwo muß der Prozeß sich stets aufhalten, und es stehen für unbegrenzt viele Zeitpunkte nur endlich viele Aufenthaltsorte zur Verfügung. \blacksquare

Die jetzt folgenden Klassifikationen und Sätze erleichtern Rekurrenzuntersuchungen und damit schließlich die Bestimmung des asymptotischen Verhaltens Markovscher Ketten beträchtlich.

3.41 Definition

Sei X eine Markov-Kette mit Zustandsraum E und Übergangsmatrix p und $i, j, k \in E$. Wir definieren:

- j ist von i aus erreichbar (i führt nach j , in Zeichen: $i \hookrightarrow j$) genau dann, wenn es ein $m \geq 0$ gibt mit $p^{(m)}(i, j) > 0$.
- i und j kommunizieren (i und j sind gegenseitig erreichbar, in Zeichen: $i \longleftrightarrow j$) genau dann, wenn $i \hookrightarrow j$ und $j \hookrightarrow i$.

3.42 Korollar

Die Relation \longleftrightarrow ist eine Äquivalenzrelation auf E und zerlegt damit E in disjunkte Teilmengen

$$C(i) \neq \emptyset, \quad i \in E: \quad \bigcup_{i \in E} C(i) = E.$$

Beweis: \longleftrightarrow ist reflexiv wegen $p^{(0)}(i, i) = 1$, symmetrisch per Definition und transitiv wegen der Chapman-Kolmogorov-Gleichung. Die Zerlegung des Zustandsraums ist eine Eigenschaft von Äquivalenzrelationen. \blacksquare

3.43 Definition

- a) Eine nichtleere Menge $A \subseteq E$ heißt abgeschlossen, falls für alle $i \in A$ gilt

$$\sum_{j \in A} p(i, j) = 1.$$

- Eine abgeschlossene Menge heißt minimal, wenn sie keine nichtleere echte Untermenge enthält, die abgeschlossen ist.
- Eine abgeschlossene Teilmenge $A \subseteq E$ heißt unzerlegbar, falls es keine nichtleere, abgeschlossene, echte Teilmenge $B \subseteq A$ gibt derart, daß auch $A - B$ abgeschlossen ist.
- Eine abgeschlossene Menge heißt irreduzibel, wenn alle ihre Elemente kommunizieren. X heißt irreduzibel, wenn E irreduzibel ist.
- $i \in E$ heißt absorbierend, falls $p(i, i) = 1$ ist.

3.44 Lemma

- a) $i \in E$ ist absorbierend genau dann, wenn $\{i\}$ abgeschlossen ist.
- b) Eine abgeschlossene Menge $A \subseteq E$, $|A| > 1$, ist irreduzibel genau dann, wenn keine nichtleere, abgeschlossene, echte Teilmenge existiert.

Beweis:

- b) i) Seien $\emptyset \neq A \subseteq E$ irreduzibel und $\emptyset \neq B \subset A$. Es gibt also $a \in A$ mit $a \in A - B$ und $b \in A$ mit $b \in B$. Damit gibt es eine Folge $b = c_0, c_1, \dots, c_{n-1}, c_n = a$ mit

$$\prod_{l=0}^{n-1} p(c_l, c_{l+1}) > 0,$$

denn sonst wäre $p^{(m)}(b, a) = 0$ für alle $m \geq 0$. Also gibt es ein $l \in \{0, \dots, n-1\}$ mit $c_l \in B$, $c_{l+1} \in A - B$ und $p(c_l, c_{l+1}) > 0$. Damit ist

$$\sum_{j \in B} p(c_l, j) < 1,$$

B also nicht abgeschlossen.

- ii) Ist $A \subseteq E$ abgeschlossen, $|A| > 1$ und $\emptyset \neq B \subset A$ ebenfalls abgeschlossen, so können die Elemente aus $A - B$ nicht von B aus erreichbar sein, damit kommunizieren nicht alle Elemente aus A . ■

3.45 Anmerkung

Sei $A \subseteq E$ abgeschlossen. Dann kann (eventuell nach einer Permutation der Zustände) p in nebenstehender Weise strukturiert werden. Dabei ist Q eine stochastische Matrix, 0 die geeignet dimensionierte Nullmatrix.

| | | |
|---------------|-----------|---------------|
| $p(i, j)$ | $j \in A$ | $j \in E - A$ |
| $i \in A$ | Q | 0 |
| $i \in E - A$ | * | |

Ist $B \subseteq A$, $B \neq A$ ebenfalls abgeschlossen, $A - B$ aber nicht, so kann p wie nebenstehend angeordnet werden. Dabei sind R und $(V|W)$ stochastisch und $V \neq 0$. □

| | | | |
|---------------|-----------|---------------|---------------|
| $p(i, j)$ | $j \in B$ | $j \in A - B$ | $j \in E - A$ |
| $i \in B$ | R | 0 | 0 |
| $i \in A - B$ | V | W | |
| $i \in E - A$ | * | | |

3.46 Beispiel

Sei $E = \mathbb{N}$ und $p = (p(i, j) : i, j \in \mathbb{N})$ Übergangsmatrix auf E mit $p(i, i + 1) = 1$ für $i \in \mathbb{N}$ sowie $p_0 = (\delta_{0k} : k \in \mathbb{N})$. Diese Angaben bestimmen eine Markov-Kette X , deren Zustandsraum die abgeschlossenen Mengen $\{k, k + 1, k + 2, \dots\}$, $k \geq 0$, enthält. Es existiert keine minimale nichtleere abgeschlossene Klasse. Anschaulich haben wir hier eine deterministische Wanderung von 0 aus unbegrenzt aufwärts vorliegen. Dies einfache System macht häufig auftretende Begriffe einfach klar oder zeigt merkwürdige Phänomene auf:

- a) Es wird keine der existierenden abgeschlossenen Klassen jemals verlassen.
- b) In einen Zustand, der einmal verlassen wurde, kehrt der Prozeß nie wieder zurück. Es existieren also keine rekurrenten Zustände und alle Zustände sind transient. □

3.47 Beispiel

- a) Sei $A \subseteq E$ abgeschlossen und

$$\alpha(\omega) = \begin{cases} \inf\{n \in \mathbb{N} : X_n(\omega) \in A\}, & \text{falls ein solches } n \text{ existiert,} \\ \infty & \text{sonst} \end{cases}$$

die Ersteintrittszeit in A . Der Endlichkeitsbereich Δ von α habe positive Wahrscheinlichkeit: $P(\Delta) > 0$. Dann ist der Post- α -Prozeß $Y = (Y_n : n \in \mathbb{N})$ eine Markov-Kette auf $(\Delta, \Delta\mathcal{F}, P(\bullet | \Delta))$ mit Zustandsraum E . Da A abgeschlossen ist, kann Y nur Zustände aus A erreichen und als Markov-Kette mit Zustandsraum A , Übergangsmatrix $p_A = (p(i, j) : i, j \in A)$ und Startverteilung $(P(Y_0 = i) = P(X_\alpha = i) : i \in A)$ angesehen werden.

- b) Wird im Lagerhaltungsmodell des Beispiels 3.13 als anfänglicher Lagerbestand X_0 eine \mathbb{N} -wertige Zufallsvariable zugelassen, so ist die Ersteintrittszeit α in $\{0, 1, \dots, S\}$ Markov-Zeit für X und der Post- α -Prozeß hat den endlichen Zustandsraum $\{0, 1, \dots, S\}$, während X auf \mathbb{N} lebt. Ein möglicher Übergang zu endlichen Prozessen erleichtert häufig Grenzwertuntersuchungen.

Wir werden also, wenn für unsere Untersuchungen das Einschwingverhalten in der Anfangszeit des Prozesses nicht wesentlich ist, wenn möglich das Vorgehen aus a) anwenden mit $A = \{0, 1, \dots, S\}$.

- c) Eine entsprechende Überlegung vereinfacht auch die Untersuchung des Langzeitverhaltens im Beispiel 3.10 für den Restlebensdauerprozeß. \square

3.48 Beispiel

Im M/G/1/ ∞ -FCFS-Modell des Beispiels 3.17 ist der (erweiterte) Schlangenlängenprozeß $X = (X_n : n \in \mathbb{N})$ mit Zustandsraum $E = \{(0, 0)\} \cup (\mathbb{N}_+ \times \mathbb{N}_+)$ genau dann irreduzibel, wenn die Bedienzeitverteilung $\pi(\bullet)$ keinen endlichen Träger hat. \square

3.49 Definition

Für die Übergangsmatrix p der Markov-Kette X sei

$$N_{(i)} := \{n \in \mathbb{N}_+ : p^{(n)}(i, i) > 0\}.$$

Ist $N_{(i)} \neq \emptyset$, so heißt der größte gemeinsame Teiler d_i von $N_{(i)}$ die Periode von i . Ist $d_i = 1$, so heißt i auch aperiodisch. Ist $N_{(i)} = \emptyset$, so ist keine Periode für i erklärt.

Wir werden im folgenden zeigen, daß die bisher für Zustände definierten Eigenschaften sogenannte „Klasseneigenschaften“ sind für die durch die Äquivalenzrelation „ \longleftrightarrow “ erzeugten Klassen. Genauer: Ist $C(i)$ die Äquivalenzklasse von i bezüglich „ \longleftrightarrow “, und ist für $j \in C(i)$ eine dieser Eigenschaften, beispielsweise Rekurrenz, nachgewiesen, so sind auch alle anderen $k \in C(i)$ rekurrent. Dies erweist sich in der Praxis als äußerst wichtig, da die meisten Modelle, die durch Markov-Ketten beschrieben werden, über einem irreduziblen Zustandsraum definiert sind. Rekurrenz zum Beispiel braucht dann nur für einen — beliebigen — Zustand gezeigt werden.

Bei Warteschlangenmodellen bietet sich hierfür fast immer der Leerzustand (in Beispiel 3.17 also $(0, 0)$) an.

3.50 Satz

- a) Sei $i \in E$ ein rekurrenter Zustand und gelte $i \leftrightarrow j \in E$. Dann ist auch j rekurrent und es gilt außerdem

$$P(X_n = j \text{ i. o.} \mid X_0 = i) = P(X_n = i \text{ i. o.} \mid X_0 = j) = 1.$$

- b) Rekurrenz und Transienz sind Klasseneigenschaften.

Beweis:

- a) Wir zeigen zuerst, daß unter den vorausgesetzten Bedingungen $P(X_n = j \text{ i. o.} \mid X_0 = i) = 1$ ist.

Seien $N > N' > 0$ beliebige natürliche Zahlen. Dann gilt:

$$\begin{aligned} & P(X_m = i \text{ für ein } m \geq N, X_l \neq j \ \forall l \geq N' \mid X_0 = i) \\ &= \sum_{m=N}^{\infty} P\left(X_m = i; X_n \neq i \text{ für } N \leq n < m; \right. \\ & \quad \left. X_l \neq j \ \forall l \geq N' \mid X_0 = i\right) \end{aligned}$$

$$\begin{aligned}
&\leq \sum_{m=N}^{\infty} P\left(\begin{array}{l} X_m = i; X_n \neq i \text{ für } N \leq n < m; \\ X_l \neq j \text{ für } N' \leq l \leq N; X_l \neq j \forall l > m \end{array} \middle| X_0 = i \right) \\
&= \sum_{m=N}^{\infty} P\left(X_l \neq j \forall l > m \middle| \begin{array}{l} X_m = i; X_n \neq i \text{ für } N \leq n < m; \\ X_l \neq j \text{ für } N' \leq l \leq N; X_0 = i \end{array} \right) \\
&\quad \cdot P\left(\begin{array}{l} X_m = i; X_n \neq i \text{ für } N \leq n < m; \\ X_l \neq j \text{ für } N' \leq l \leq N \end{array} \middle| X_0 = i \right) \\
&= \sum_{m=N}^{\infty} P(X_l \neq j \forall l > m \mid X_m = i) \cdot P\left(\begin{array}{l} X_m = i; X_n \neq i \text{ für } N \leq n < m; \\ X_l \neq j \text{ für } N' \leq l \leq N \end{array} \middle| X_0 = i \right) \\
&= (1 - f_{ij}^*) \cdot \sum_{m=N}^{\infty} P\left(\begin{array}{l} X_m = i; X_n \neq i \text{ für } N \leq n < m; \\ X_l \neq j \text{ für } N' \leq l \leq N \end{array} \middle| X_0 = i \right) \\
&= (1 - f_{ij}^*) \cdot P\left(\begin{array}{l} X_m = i \text{ für ein } m \geq N; \\ X_l \neq j \text{ für } N' \leq l \leq N \end{array} \middle| X_0 = i \right).
\end{aligned}$$

Direkt aus der Definition in 3.34 b) und der Stetigkeit von P (*) folgt

$$\begin{aligned}
P(X_m = i \text{ i. o.} \mid X_0 = i) &= P\left(\bigcap_{N=1}^{\infty} \bigcup_{m=N}^{\infty} \{X_m = i\} \middle| X_0 = i \right) \\
&= \lim_{N \rightarrow \infty} P\left(\bigcup_{m=N}^{\infty} \{X_m = i\} \middle| X_0 = i \right) \\
&\stackrel{*}{=} \lim_{N \rightarrow \infty} P(X_m = i \text{ für ein } m \geq N \mid X_0 = i).
\end{aligned}$$

Entsprechend erhält man beim Übergang $N \rightarrow \infty$ auf der äußerst linken und äußerst rechten Seite der obigen Ungleichung

$$\begin{aligned}
P(X_m = i \text{ i. o.}; X_l \neq j \forall l \geq N' \mid X_0 = i) \\
\leq P(X_m = i \text{ i. o.}; X_l \neq j \forall l \geq N' \mid X_0 = i) \cdot (1 - f_{ij}^*).
\end{aligned}$$

Da nach Voraussetzung $f_{ij}^* > 0$ ist, folgt für alle $N' \in \mathbb{N}$:

$$P(X_m = i \text{ i. o.}; X_l \neq j \forall l \geq N' \mid X_0 = i) = 0.$$

Also gilt:

$$\begin{aligned}
1 &= P(X_m = i \text{ i. o.} \mid X_0 = i) \\
&= P(X_m = i \text{ i. o.}; X_l = j \text{ i. o.} \mid X_0 = i) + P(X_m = i \text{ i. o.}; X_l = j \text{ nur endlich oft} \mid X_0 = i) \\
&= P(X_m = i \text{ i. o.}; X_l = j \text{ i. o.} \mid X_0 = i),
\end{aligned}$$

denn

$$\begin{aligned}
P(X_m = i \text{ i. o.}; X_l = j \text{ nur endlich oft} \mid X_0 = i) \\
\leq \sum_{N'=1}^{\infty} P(X_m = i \text{ i. o.}; X_l \neq j \forall l \geq N' \mid X_0 = i) = 0.
\end{aligned}$$

Insgesamt ist also $P(X_l = j \text{ i. o.} \mid X_0 = i) = 1$, da $P(X_m = i \text{ i. o.} \mid X_0 = i) = 1$ vorausgesetzt war. Nach Lemma 3.35 muß damit gelten

$$P(X_n = j \text{ i. o.} \mid X_0 = j) = 1,$$

so daß nach Korollar 3.36 auch j rekurrent ist.

b) ist damit auch schon gezeigt. ■

3.51 Beispiel

X sei Markov-Kette mit Zustandsraum \mathbb{N} und der Übergangsmatrix

$$\begin{pmatrix} p_0 & 1-p_0 & 0 & 0 & \dots & 0 & \dots \\ p_1 & 0 & 1-p_1 & 0 & \dots & 0 & \dots \\ p_2 & 0 & 0 & 1-p_2 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \\ p_r & 0 & 0 & 0 & \dots & 1-p_r & \dots \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \ddots \end{pmatrix},$$

wobei $0 < p_i < 1$ für $i \in \mathbb{N}$ gelte. X ist also irreduzibel mit \mathbb{N} als einziger Klasse. Für Rekurrenz- und Transienzuntersuchungen genügt es daher, den Zustand „0“ zu betrachten. Es gilt:

$$f_{00}^{(1)} = p_0, \quad f_{00}^{(n)} = \prod_{i=0}^{n-2} (1-p_i)p_{n-1}, \quad n \geq 2.$$

Für $r \geq -1$ setzen wir

$$U_r = \begin{cases} (1-p_0)(1-p_1)\dots(1-p_r) & \text{für } r \geq 0, \\ 1 & \text{für } r = -1; \end{cases}$$

dann gilt

$$f_{00}^{(n)} = \prod_{i=0}^{n-2} (1-p_i)(1-(1-p_{n-1})) = U_{n-2} - U_{n-1}, \quad n \geq 1.$$

Es folgt für $m \geq 0$:

$$\sum_{n=1}^{m+1} f_{00}^{(n)} = \sum_{r=1}^{m-1} U_r - \sum_{r=0}^m U_r = 1 - U_m.$$

Es gilt aber (siehe [KT75], S. 70)

$$U_m = \prod_{i=0}^m (1-p_i) \xrightarrow{m \rightarrow \infty} 0 \iff \sum_{i=0}^{\infty} p_i \text{ divergiert.}$$

Also haben wir insgesamt:

$$\sum_{i=0}^{\infty} p_i = \infty \iff f_{00}^* = 1 \iff 0 \text{ ist rekurrent.}$$

Aus Satz 3.50 b) folgt also:

$$\sum_{i=0}^{\infty} p_i = \infty \iff \text{Alle } i \in E \text{ sind rekurrent.} \quad \square$$

3.52 Satz

Nullrekurrenz und positive Rekurrenz sind Klasseneigenschaften.

Beweis: Es reicht zu zeigen, daß für $i, j \in E$, $i \longleftrightarrow j$, gilt: Ist i positiv rekurrent, so kann j nicht nullrekurrent sein (oder umgekehrt).

Zunächst bemerken wir: Aus Satz 3.50 folgt, daß i und j rekurrent sind, also nach Satz 3.50 und Korollar 3.36

$$P(X_n = j \text{ i. o.} \mid X_0 = i) = P(X_n = j \text{ i. o.} \mid X_0 = j) = 1$$

gilt. Aus Lemma 3.35 folgt $f_{ij}^* = 1$ und aus Symmetriegründen $f_{ji}^* = 1$. Die Ersteintrittszeit $\varrho_1(i, j)$ und $\varrho_1(j, i)$ von i nach j bzw. von j nach i haben also nicht defekte Verteilungen. Entsprechendes gilt für $\varrho_1(i)$, $\varrho_1(j)$ (siehe Definition 3.31).

Die Rückkehrzeit von i nach i wird aufgespalten nach den Ereignissen

A_0 := „kein Durchgang durch j vor der Rückkehr nach i “

A_m := „genau m Durchgänge durch j vor der Rückkehr nach i “, $m = 1, 2, \dots$

Die Anzahl der Durchgänge durch j vor der Rückkehr nach i ist geometrisch verteilt wegen der Gültigkeit der starken Markov-Eigenschaft. Wird eine entsprechende Aufspaltung der Rückkehrzeit $\varrho_1(j)$ nach j vorgenommen, treten nach entsprechender Bedingung die gleichen bedingten Rückkehrzeitmittelwerte bzw. Ersteintrittszeitmittelwerte auf. Damit können $E(\varrho_1(i))$ und $E(\varrho_1(j))$ nur gemeinsam endlich oder unendlich sein. ■

3.53 Satz

Die Periodenlänge ist eine Klasseeigenschaft, d. h. aus $j \in C(i)$ folgt $d_i = d_j$ (sofern definiert).

Beweis: Sei $i \leftrightarrow j$, d_i sei definiert und $i \neq j$. Also gibt es $m, n, s \in \mathbb{N}_+$ mit

$$p^{(m)}(i, j) > 0, \quad p^{(n)}(j, i) > 0, \quad p^{(s)}(i, i) > 0.$$

Es gilt:

$$p^{(n+ks+m)}(j, j) \geq p^{(n)}(j, i) \cdot p^{(ks)}(i, i) \cdot p^{(m)}(i, j) > 0,$$

so daß $N_{(j)} \neq \emptyset$ ist und d_j existiert. d_j teilt $(n+s+m)$ und $(n+2s+m)$, also auch $(n+2s+m) - (n+s+m) = s$. Damit muß d_j auch d_i (den größten gemeinsamen Teiler aller solchen s) teilen. Da d_j existiert, kann diese Argumentation mit vertauschten i und j durchgeführt werden, was $d_i = d_j$ bedeutet. ■

Die periodische Struktur einer Markov-Kette läßt sich noch genauer beschreiben.

3.54 Satz

Der Zustand i habe Periode d_i . Sei $j \in C(i)$, $j \neq i$.

a) Dann existiert eine eindeutig bestimmte Zahl $r_j \in \{0, 1, \dots, d_i - 1\}$, so daß gilt:

$$p^{(n)}(i, j) > 0 \quad \implies \quad n \equiv r_j \pmod{d_i}.$$

b) Außerdem gibt es ein $n(j)$ mit

$$n \geq n(j) \quad \implies \quad p^{(nd_i+r_j)}(i, j) > 0.$$

c) Die Klasse $C(i)$ zerfällt in disjunkte, nichtleere Teilklassen $C_0(i), C_1(i), \dots, C_{d_i-1}(i)$. Diese Zerlegung ist unabhängig von der Wahl von i (bis auf zyklische Permutation der Indizes).

Wird formal $C_p(i)$ auch für $p \in \mathbb{N}$, $p \geq d_i$, definiert durch

$$C_p(i) = C_q(i) \quad \iff \quad p \equiv q \pmod{d_i},$$

so gilt:

$$\sum_{j \in C_{r+n}(i)} p^{(n)}(l, j) = 1, \quad \text{falls } l \in C_r(i).$$

Beweis:

a) Seien $m \neq m' \in \mathbb{N}$ mit $p^{(m)}(i, j) > 0$, $p^{(m')}(i, j) > 0$ und $n \in \mathbb{N}$ mit $p^{(n)}(j, i) > 0$. Also gilt $p^{(m+n)}(i, i) > 0$, $p^{(m'+n)}(i, i) > 0$ und d_i teilt sowohl $m+n$ als auch $m'+n$. Damit teilt d_i auch $m-m'$. Das impliziert, daß m und m' den gleichen Divisionsrest bei Division durch d_i haben.

b) siehe [Chu67], S. 14.

c) Die Zerlegung erfolgt nach den zugehörigen $r_j \in \{0, 1, \dots, d_i - 1\}$, die in a) gefunden wurden. Für den Rest des Beweises siehe [Chu67], S. 15. ■

3.55 Beispiel

In einer Reparaturhalle müssen defekte Werkstücke eine Reihe von Arbeitsgängen A_0, A_1, \dots, A_5 durchlaufen, eventuell mehrmals hintereinander. Ist das Werkstück wiederhergestellt, wird es aus dem Arbeitsprozeß sofort herausgenommen und ein neues Werkstück beginnt ohne Zeitverlust den Reparaturzyklus. Wir nehmen an, daß stets Werkstücke auf Reparatur warten.

Wir legen fest: Reparaturanforderungen treten über A_0 in den Zyklus ein. Nach Bearbeitung in A_1 geht das Werkstück entweder mit Wahrscheinlichkeit $1 - p$ nach A_2 oder mit Wahrscheinlichkeit p nach A_5 . In A_5 wird eine Kontrolle durchgeführt und entschieden, ob das Werkstück wiederhergestellt ist (die Wahrscheinlichkeit hierfür ist $1 - q$) oder noch einen Zyklus, beginnend mit A_0 , durchlaufen muß (mit Wahrscheinlichkeit q). Die restliche Ablaufsteuerung ist deterministisch (siehe Abbildung 5).

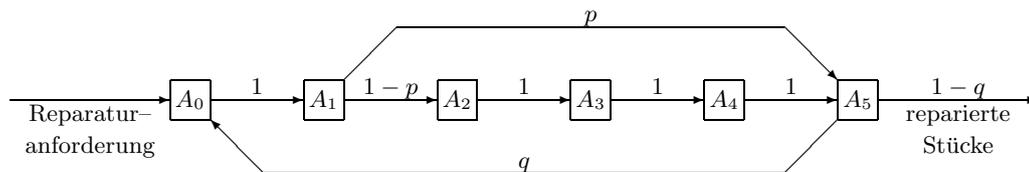


Abbildung 5: Reparaturablaufplan

Wir nehmen an, daß die Verzweigungsentscheidungen unabhängig voneinander und den sonstigen Daten sind.

Es sei $X_n = A_i$ genau dann, wenn das zur Zeit $n \in \mathbb{N}$ in der Halle bearbeitete Werkstück sich im Arbeitsgang A_i befindet. Dann ist unter den angegebenen Modellannahmen $X = (X_n : n \in \mathbb{N})$ eine Markov-Kette, falls P^{X_0} festgelegt ist. X ist auf dem Zustandsraum $E = \{A_0, A_1, \dots, A_5\}$ irreduzibel und alle Zustände haben die Periode 3. Mit Bezug auf A_0 sind die Restklassenzahlen der Zustände aus E

$$\begin{array}{cccccc} A_0 & A_1 & A_2 & A_3 & A_4 & A_5 \\ \hline 0 & 1 & 2 & 0 & 1 & 2 \end{array} \implies \begin{array}{ccc} C_0(A_0) & C_1(A_0) & C_2(A_0) \\ \hline A_0, A_3 & A_1, A_4 & A_2, A_5 \end{array}.$$

Nimmt man einen anderen Bezugspunkt, so werden die Restklassenzahlen zyklisch permutiert. \square

3.56 Anmerkung

- a) Ist X irreduzibel mit Periode $d > 1$ und ordnet man entsprechend einer Folge $C_0(i), C_1(i), \dots, C_{d-1}(i)$ den Zustandsraum E an, so hat die Übergangsmatrix die folgende Blockform auf dem Teilraum $C(i)$:

$$\left(\begin{array}{c|cccccc} p(\bullet, \bullet) & C_0(i) & C_1(i) & C_2(i) & C_3(i) & \dots & C_{d-1}(i) \\ \hline C_0(i) & 0 & p_0 & 0 & 0 & \dots & 0 \\ C_1(i) & 0 & 0 & p_1 & 0 & \dots & 0 \\ C_2(i) & 0 & 0 & 0 & p_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ C_{d-2}(i) & 0 & 0 & 0 & 0 & \dots & p_{d-2} \\ C_{d-1}(i) & p_{d-1} & 0 & 0 & 0 & \dots & 0 \end{array} \right)$$

Dabei sind die 0-Blockmatrizen der Diagonale quadratisch; die p_i sind stochastisch, aber im allgemeinen nicht quadratisch. Die Blöcke sind nicht notwendig endlich.

- b) Die Erreichbarkeitsfragen und Periodizitätseigenschaften der stochastischen Matrizen hängen nur von der Platzierung der Nullen ab. Man kann deshalb die Definitionen auf nicht-negative Matrizen sinnvoll übertragen (siehe [Sen80]). \square

4 Asymptotisches Verhalten, Stabilität und stationäre Verteilungen von Markov-Ketten

Ein Warteschlangensystem (siehe Beispiel 3.17) wird in der Praxis kurz als „stabil“ bezeichnet, wenn die Kapazität des Bedienungsgerätes groß genug ist, um sämtliche eintreffende Arbeitsanforderung auch abzuarbeiten — und dies auch in nicht allzu langer Zeit nach Ankunft der Aufträge. Mit anderen Worten: Das Bediengerät muß so angelegt sein, daß der Leerzustand des Systems immer wieder (in endlicher Zeit) erreicht wird, und zudem soll die Zeit zwischen zwei Leerzuständen im Mittel endlich sein. In der Terminologie des letzten Abschnitts heißt dies: Der Leerzustand des Schlangenlängenprozesses ist positiv rekurrent, falls das System stabil ist. Instabil wird das System also genannt, wenn der Leerzustand transient oder nullrekurrent ist. Die anschauliche Konsequenz ist dann, daß die Schlangenlänge (im Mittel) unbegrenzt anwächst bzw. daß es im Mittel unendlich lange dauert, bis ein leer gestartetes System wieder leer wird. Ist dann der Schlangenlängenprozeß irreduzibel, übertragen sich diese Eigenschaften auf die Gesamtheit der Zustände und wir benennen den gesamten Prozeß bzw. sogar „das Wartesystem“ entsprechend.

In einem etwas anderen Sinne wird das Wort „stabil“ verwendet, wenn untersucht wird, ob das System sich in einen „Gleichgewichtszustand“ einschwingt, d. h. sich „stabilisiert“ und dann „stationäres“ Verhalten zeigt. Dabei kann in der Regel nicht angenommen werden, daß der dann erreichte „stationäre Zustand“ eindeutig bestimmt ist wie in der asymptotischen Theorie für Lösungen von Differentialgleichungen: Diese bewegen sich unter geeigneten Bedingungen in einem deterministischen Ruhezustand. Bei unseren zufallsbeeinflussten Systemen werden wir auch asymptotisch die Zufallsfluktuationen behalten. Was wir zu erreichen hoffen können ist, daß die Wahrscheinlichkeiten sich einschwingen, und zwar im Sinne einer Konvergenz von Zähldichten, was wegen des diskreten Zustandsraumes gleichbedeutend ist zur schwachen Konvergenz der Zustandsverteilungen gegen eine Grenzverteilung.

Das letzte Thema dieses Kapitels entsteht aus der Frage, ob es möglich ist, eine Markov-Kette so zu starten, daß schon Stationarität vorliegt, also das Einschwingen schon stattgefunden hat.

Es stellt sich heraus, daß alle drei Probleme eng miteinander zusammenhängen, was nicht sehr überraschend ist, wenn man im Beispiel eines Warteschlangensystems dies etwas weiter inhaltlich durchdenkt.

Wir klären zunächst die Begriffe und legen wieder fest, daß — wenn nicht anderes gesagt ist — $X = (X_n : n \in \mathbb{N})$ mit $X_n : (\Omega, \mathcal{F}, P) \rightarrow (E, \mathcal{P}(E))$ für $n \in \mathbb{N}$ eine homogene Markov-Kette gemäß Kapitel 3 ist.

4.1 Definition

Eine Markov-Kette X heißt stationär, falls für alle $n \in \mathbb{N}_+$ und alle Zeitpunkte t_1, t_2, \dots, t_n mit $0 \leq t_1 < t_2 < \dots < t_n$ gilt:

$$P^{(X_{t_1}, \dots, X_{t_n})} = P^{(X_{t_1+h}, \dots, X_{t_n+h})} \quad \text{für alle } h \in \mathbb{N}.$$

4.2 Definition

Gegeben sei eine Markov-Kette $X = (X_n : n \in \mathbb{N})$.

- Ein nichtnegativer Vektor $q = (q(i) : i \in E) \in \mathbb{R}_+^E$ heißt invariantes Maß für X , falls gilt: $q = q \cdot p$ („Stationaritäts- oder Gleichgewichtsgleichung“). q ist also ein linker Eigenvektor zum Eigenwert 1 der Übergangsmatrix p von X .
- Ein invariantes Wahrscheinlichkeitsmaß für X wird als stationäre Verteilung von X bezeichnet. Gebräuchlich sind dafür auch die Bezeichnungen Gleichgewicht q , Gleichgewichtsverteilung q , Äquilibrium q , stationäre Anfangsverteilung q .

Die folgende Beobachtung begründet die letzte Bezeichnung:

4.3 Korollar

Ist q ein invariantes Wahrscheinlichkeitsmaß für X und gilt $p_0 = P^{X_0} = q$, so ist X stationär und die eindimensionalen Randverteilungen sind zeitlich invariant:

$$p_n = p_0 \cdot p^n = q, \quad n \geq 0.$$

Beweis: $p_n = q$ für $n \geq 1$ folgt durch Induktion aus der Definition. Aus der Homogenität von X folgt analog zu Satz 3.7 a), daß $p^{(X_{t_1+h}, \dots, X_{t_n+h})}$ eindeutig bestimmt ist durch p_{t_1+h} und p . Die Stationarität folgt damit aus $p_0 = p_{t_1+h}$ und direktem Ausrechnen der endlichdimensionalen Randverteilungen, sowie deren Umrechnen mittels Formel von der totalen Wahrscheinlichkeit. ■

4.4 Definition

Gegeben sei eine Markov-Kette $X = (X_n : n \in \mathbb{N})$ mit Startverteilung $p_0 = P^{X_0}$. X besitzt eine Grenzverteilung unter p_0 , falls

$$\pi(k) := \lim_{n \rightarrow \infty} p_n(k) \text{ existiert für alle } k \in E$$

und es gilt

$$\sum_{k \in E} \pi(k) = 1,$$

d. h. π ist ein Wahrscheinlichkeitsmaß auf E . π wird dann auch als asymptotische Verteilung von X unter p_0 bezeichnet. Falls die Grenzverteilung π von p_0 unabhängig ist, sagen wir kurz: X besitzt die Grenzverteilung π .

4.5 Korollar

q sei ein invariantes Wahrscheinlichkeitsmaß für X . Wird dann X unter q gestartet, d. h. $p_0 = q$, so besitzt X eine Grenzverteilung (unter q), nämlich q .

Jede stationäre Verteilung ist in diesem Sinne auch eine Grenzverteilung. Falls also X mehrere stationäre Verteilungen $q_i, i \in I$, besitzt, so existieren auch mehrere Grenzverteilungen, jeweils unter den verschiedenen q_i gestartet.

Einen Überblick über die auftretenden Möglichkeiten gewinnen wir anhand einfacher Beispiele.

4.6 Beispiel

Die in Beispiel 3.45 definierte deterministische, in Schritten der Höhe 1 aufsteigende Bewegung sollte anschaulich keine stationäre Verteilung besitzen.

Formal haben wir: Ist p_0 auf $[k, \infty)$ konzentriert mit $p_0(k) > 0$, so ist $p_1 = p_0 \cdot p$ auf $[k+1, \infty)$ konzentriert mit $p_1(k+1) > 0$. Mit dem gleichen Argument folgt:

$$\sum_{k=0}^{n-1} p_n(k) = 0 \quad \text{für alle } n \geq 1,$$

also

$$\lim_{n \rightarrow \infty} p_n(k) = 0 \quad \text{für alle } k \in \mathbb{N}.$$

Damit kann auch keine Grenzverteilung existieren. □

4.7 Beispiel

Im Beispiel 3.16 haben wir die Potenzen der Übergangsmatrizen für Markov-Ketten mit Zustandsraum $E = \{0, 1\}$ untersucht:

$$p(i, j) \begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-a & a \\ 1 & b & 1-b \end{array} \quad (a, b \in [0, 1]).$$

- $a = b = 0$. Dann ist $p = I$ und jede Verteilung auf $\{0, 1\}$ stationäre und Grenzverteilung (unter dieser Anfangsverteilung).
- $a = b = 1$. Wir lösen die Stationaritätsgleichung

$$X = (X(0), X(1)) = X \cdot p$$

und erhalten als Lösungsmenge $\{X = (\eta, \eta) : \eta \in \mathbb{R}\}$. Als einzige Wahrscheinlichkeitslösung haben wir also die Gleichverteilung $q = (\frac{1}{2}, \frac{1}{2})$ auf E , welche damit stationäre und Grenzverteilung ist (unter dieser Anfangsverteilung). Wegen

$$(X(1), X(2)) \cdot p = (X(2), X(1))$$

kann es keine weitere Grenzverteilung geben. Außerdem folgt, daß unter keiner anderen Startverteilung außer $q = (\frac{1}{2}, \frac{1}{2})$ Konvergenz der p_n gegen eine Grenzverteilung vorliegt.

- c) $0 < a+b < 2$. Wir lösen die Stationaritätsgleichung und erhalten als Lösungsraum $\{c \cdot (b, a) : c \in \mathbb{R}\}$. Als einzige Wahrscheinlichkeitslösung haben wir also

$$q = \frac{1}{a+b} \cdot (b, a),$$

welche eindeutig bestimmte stationäre Verteilung ist und damit auch wieder Grenzverteilung unter dieser Anfangsverteilung. Die Frage nach weiteren Grenzverteilungen können wir hier numerisch beantworten:

Nach Beispiel 3.16 gilt

$$\begin{aligned} p_n = p_0 \cdot p^n &= \frac{1}{a+b} \left((p_0(0), p_0(1)) \begin{pmatrix} b & a \\ b & a \end{pmatrix} + (p_0(0), p_0(1))(1-a-b)^n \begin{pmatrix} a & -a \\ -b & b \end{pmatrix} \right) \\ &= \frac{1}{a+b} \left((b, a) + (p_0(0), p_0(1))(1-a-b)^n \begin{pmatrix} a & -a \\ -b & b \end{pmatrix} \right) \\ &\rightarrow \frac{1}{a+b} (b, a). \end{aligned}$$

$q = \frac{1}{a+b} (b, a)$ ist also die eindeutig bestimmte asymptotische und Grenzverteilung von X , und von jeder Anfangsverteilung aus konvergiert die Folge der Zustandsverteilungen eindeutig gegen q (im Sinne der schwachen Konvergenz).

- d) Falls in c) $a = 0$ bzw. $b = 0$ ist, haben wir den absorbierenden Zustand 0 bzw. 1.
 e) Im allgemeinen gilt unter den Bedingungen von c), daß $p_n \neq q$ ist für alle $n \in \mathbb{N}$, falls nicht $p_0 = q$ ist. \square

4.8 Beispiel

Wir betrachten ein System in den diskreten Zeitpunkten 0, 1, 2, ... und unterscheiden als mögliche Zustände des Systems:

- 0 $\hat{=}$ System arbeitet,
- 1 $\hat{=}$ System ist ausgefallen und wird repariert (ist in „Reparatur“),
- 2 $\hat{=}$ System ist in Ordnung, arbeitet aber nicht (ist „in Ruhe“).

Aus Beobachtungen ergibt sich: Die Wahrscheinlichkeit, daß das System zur Zeit $n+1$ im Zustand $x \in E = \{0, 1, 2\}$ ist, hängt nur von x und dem Zustand ab, in dem sich das System zur Zeit n befindet, $n \in \{0, 1, 2, \dots\} = \mathbb{N}$.

Genauer ergibt sich außerdem: Arbeitet das System zur Zeit n , so arbeitet es zur Zeit $n+1$ mit Sicherheit nicht und die Zustände „in Reparatur“ und „in Ruhe“ sind gleichwahrscheinlich.

Falls das System zur Zeit n in Ruhe oder Reparatur ist, so arbeitet es zur Zeit $n+1$ mit Sicherheit.

Wird das Systemverhalten durch eine Markov-Kette beschrieben, erhalten wir $X = (X_n : n \in \mathbb{N})$, mit $X_n : (\Omega, \mathcal{A}, P) \rightarrow (E, \mathcal{P}(E))$ für $n \in \mathbb{N}$ und mit Übergangsmatrix

$$\left(\begin{array}{c|ccc} p(\bullet, \bullet) & 0 & 1 & 2 \\ \hline 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 \end{array} \right).$$

Dann gilt: X ist irreduzibel und alle Zustände haben die Periode 2. Die n -Schritt-Übergangswahrscheinlichkeiten sind:

$$p^{2n} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad p^{2n-1} = p, \quad n \in \mathbb{N}_+.$$

Die einzige stationäre Verteilung für X ist $p = (\frac{1}{2}, \frac{1}{4}, \frac{1}{4})$.

Bei einer Anfangsverteilung $p_0 = (p_0(0), p_0(1), p_0(2))$ gilt für die eindimensionalen Randverteilungen von X für $n \geq 1$:

$$p_n(0) = \begin{cases} p_0(0) & n \text{ gerade} \\ p_0(1) + p_0(2) & n \text{ ungerade} \end{cases}$$

$$p_n(1) = \begin{cases} \frac{1}{2}(p_0(1) + p_0(2)) & n \text{ gerade} \\ \frac{1}{2}p_0(0) & n \text{ ungerade} \end{cases}$$

$$p_n(2) = \begin{cases} \frac{1}{2}(p_0(1) + p_0(2)) & n \text{ gerade} \\ \frac{1}{2}p_0(0) & n \text{ ungerade} \end{cases}$$

Für $n \geq 1$ sind also für $i \in E$ die Folgen $(p_n(i) : n \in \mathbb{N})$ konstant (und zwar genau dann, wenn gilt $p_0(0) = p_0(1) + p_0(2)$) oder alternierend (und zwar genau dann, wenn gilt $p_0(0) \neq p_0(1) + p_0(2)$). Im ersten Fall liegt Konvergenz gegen eine Grenzverteilung vor, welche wegen $p_0(0) + p_0(1) + p_0(2) = 1$ und $p_n(1) = p_n(2) = \frac{1}{2}p_0(0)$ gerade $(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}) = q$ ist. Im zweiten Fall liegt keine Konvergenz gegen eine Grenzverteilung vor. \square

4.9 Satz

Sei X eine Markov-Kette mit Zustandsraum E und Übergangsmatrix p . Dann gilt:

Die absoluten Zustandswahrscheinlichkeiten $p_n = (p_n(i) : i \in E)$ konvergieren punktweise (d. h. für alle $i \in E$) gegen einen Vektor $a = (a(i) : i \in E)$ unabhängig von der Startverteilung p_0 genau dann, wenn die Grenzwerte

$$b(j) = \lim_{n \rightarrow \infty} p^{(n)}(i, j), \quad j \in E,$$

existieren und nicht von i abhängen, d. h. wenn die Matrixpotenzen p^n gegen eine Matrix mit identischen Zeilen punktweise konvergieren:

Ist $b = (b(i) : i \in E)$, so gilt

$$p^{(n)} \xrightarrow{n \rightarrow \infty} \begin{pmatrix} b \\ b \\ \vdots \end{pmatrix}.$$

Außerdem gilt dann $a = b$.

Ist außerdem $a = (a(i) : i \in E)$ ein stochastischer Vektor, so ist a die eindeutig bestimmte stationäre Verteilung von X .

Beweis:

a) Die Limiten $b(j) = \lim_{n \rightarrow \infty} p^{(n)}(i, j)$ mögen existieren und nicht von $i \in E$ abhängen. Dann gilt:

$$\begin{aligned} \lim_{n \rightarrow \infty} p_n(i) &= \lim_{n \rightarrow \infty} (p_0 \cdot p^n)(i) = \lim_{n \rightarrow \infty} \left(\sum_{j \in E} p_0(j) p^{(n)}(j, i) \right) = \sum_{j \in E} p_0(j) \left(\lim_{n \rightarrow \infty} p^{(n)}(j, i) \right) \\ &= \sum_{j \in E} p_0(j) b(i) = b(i). \end{aligned}$$

Die Summation kann mit dem Grenzwert vertauscht werden unter Anwendung des Satzes von der majorisierten Konvergenz: Die Funktionen $p^{(n)}(i, \bullet)$ sind durch die bezüglich $(p_0(i) : i \in E)$ integrierbaren (summierbaren) Funktionen 0 und 1 von unten und oben beschränkt.

- b) Falls die absoluten Zustandsverteilungen unabhängig von der Startverteilung gegen a konvergieren, wähle $p_0(i) = \delta_{i_0, i}$ für ein $i_0 \in E$. Dann gilt:

$$p_n(i) = \sum_{j \in E} p_0(j) p^{(n)}(j, i) = p^{(n)}(i_0, i) \quad \text{und} \quad p_n(i) \rightarrow a(i),$$

unabhängig von i_0 .

- c) Falls $\lim_{n \rightarrow \infty} p^{(n)}(i, j) = b(j)$, $j \in E$, unabhängig von i existieren und einen stochastischen Vektor definieren, gilt (Chapman-Kolmogorov-Gleichung)

$$p^{(n)}(i, j) = \sum_{k \in E} p^{(n-1)}(i, k) p(k, j), \quad i, j \in E, \quad n \geq 1.$$

Aus dem Lemma von Fatou (siehe [Hin72], S. 119) folgt:

$$\lim_{n \rightarrow \infty} p^{(n)}(i, j) = \lim_{n \rightarrow \infty} \sum_{k \in E} p^{(n-1)}(i, k) p(k, j) \geq \sum_{k \in E} \lim_{n \rightarrow \infty} p^{(n-1)}(i, k) p(k, j) = \sum_{k \in E} b(k) p(k, j),$$

für alle $j \in E$ also:

$$b(j) \geq \sum_{k \in E} b(k) p(k, j) \geq 0. \quad (*)$$

Summation über $j \in E$ und Anwendung des Satzes von Fubini liefert:

$$1 = \sum_{j \in E} b(j) \geq \sum_{j \in E} \sum_{k \in E} b(k) p(k, j) = \sum_{k \in E} b(k) \sum_{j \in E} p(k, j) = \sum_{k \in E} b(k) = 1.$$

Damit muß für alle Summanden aus den Ungleichungen (*) schon Gleichheit gelten, und dies sind die Gleichgewichtsbedingungen für X . ■

Während im letzten Satz stationäre Verteilungen als Grenzwerte erhalten wurden, geben wir jetzt eine andere Konstruktion an, die eine Grundtatsache zur Begründung der sogenannten „Regenerativen Simulation“ darstellt. Wir verwenden dabei — und im folgenden bei geeigneter Gelegenheit — die folgenden abkürzenden Schreibweisen:

4.10 Definition

Sei $X = (X_n : n \in \mathbb{N})$ eine homogene Markov-Kette auf (Ω, \mathcal{F}, P) mit Zustandsraum.

Falls die Startverteilung p_0 in dieser Darstellung fixiert werden soll, wird der unterliegende Wahrscheinlichkeitsraum bezeichnet als $(\Omega, \mathcal{F}, P_{p_0})$. Wir haben also formal $P_{p_0}^{X_0} = p_0$.

Für Erwartungswerte unter P_{p_0} schreiben wir $E_{p_0}(\bullet)$. Falls $p_0 = \varepsilon_{i_0}$ die Einpunktverteilung in $i_0 \in E$ ist, schreiben wir statt $P_{\varepsilon_{i_0}}$ und $E_{\varepsilon_{i_0}}(\bullet)$ einfach P_{i_0} und $E_{i_0}(\bullet)$.

Mit diesen Schreibweisen erhalten wir also z. B.

$$P(X_2 = j \mid X_0 = i_0) = P_{i_0}(X_2 = j), \\ E(X_2 \mid X_0 = i_0) = E_{i_0}(X_2),$$

oder trivial

$$E(X_2) = E_{p_0}(X_2), \quad P(X_2 = j) = P_{p_0}(X_2 = j).$$

Auf diese Weise läßt sich ein Wechsel der Startverteilung schreibtechnisch einfach darstellen; außerdem werden später technische Umformungen leichter darstellbar. Ein Beispiel liefert der folgende Satz.

4.11 Satz

Sei X eine Markov-Kette, $i_0 \in E$ ein beliebiger, aber fest gewählter, rekurrenter Referenzzustand und $P(X_0 = i_0) = 1$. Für $j \in E$ sei ν_j die erwartete Anzahl von Eintritten in j zwischen zwei Eintritten in i_0 von X , d. h.

$$\begin{aligned} \nu_j &:= \mathbb{E}_{i_0} \left(\sum_{n=0}^{\varrho_1(i_0)-1} 1_{(X_n=j)} \right) = \sum_{n=0}^{\infty} P_{i_0}(X_n = j, \varrho_1(i_0) > n) = \mathbb{E} \left(\sum_{n=0}^{\varrho_1(i_0)-1} 1_{(X_n=j)} \mid X_0 = i_0 \right) \\ &= \sum_{n=0}^{\infty} P(X_n = j, \varrho_1(i_0) > n \mid X_0 = i_0). \end{aligned}$$

(Insbesondere gilt: $\nu_{i_0} = 1$.)

Dann ist der Vektor $\nu = (\nu_j : j \in E)$ ein stationäres Maß für X (insbesondere also aus \mathbb{R}_+^E) mit $\nu \neq 0$. Falls zur Unterscheidung bezüglich des Referenzzustandes nötig, bezeichnen wir den Vektor mit

$$\nu := \nu(i_0) = (\nu_j(i_0) : j \in E).$$

Beweis: Wir haben nach Voraussetzung und Definition $X_0 = i_0 = X_{\varrho_1(i_0)}$ P_{i_0} -fast sicher. Also läßt sich der Summationsindex verschieben zu:

$$\begin{aligned} \nu_j &= \mathbb{E}_{i_0} \left(\sum_{n=0}^{\varrho_1(i_0)-1} 1_{(X_n=j)} \right) \\ &= \mathbb{E}_{i_0} \left(\sum_{n=1}^{\varrho_1(i_0)} 1_{(X_n=j)} \right) \\ &= \sum_{n=1}^{\infty} \mathbb{E}_{i_0} (1_{(X_n=j, \varrho_1(i_0) \geq n)}) \\ &= \sum_{n=1}^{\infty} \mathbb{E}_{i_0} \left(\mathbb{E}_{i_0} (1_{(X_n=j)} \cdot 1_{(\varrho_1(i_0) > n-1)} \mid \sigma(X_k : k \leq n-1)) \right) \end{aligned}$$

(Der Erwartungswert einer Zufallsvariablen ist darstellbar als Erwartungswert der bedingten Erwartung dieser Zufallsvariablen gegeben eine Unter- σ -Algebra; siehe [Bau74], S. 295 (54.18); [Hin72], S. 180.)

$$= \sum_{n=1}^{\infty} \mathbb{E}_{i_0} \left(1_{(\varrho_1(i_0) > n-1)} \cdot \mathbb{E}_{i_0} (1_{(X_n=j)} \mid \sigma(X_k : k \leq n-1)) \right)$$

($\varrho_1(i_0)$ ist eine Stoppzeit bezüglich X , also gilt $\{\varrho_1(i_0) \leq n-1\} \in \sigma(X_k : k \leq n-1)$; die $\sigma(X_k : k \leq n-1)$ -meßbare Funktion $1_{(\varrho_1(i_0) > n-1)}$ kann aus dem bedingten Erwartungswert bezüglich $\sigma(X_k : k \leq n-1)$ herausgezogen werden; siehe [Bau74], S. 294 (54.16).)

$$= \sum_{n=1}^{\infty} \mathbb{E}_{i_0} (1_{(\varrho_1(i_0) > n-1)} \cdot p(X_{n-1}, j))$$

(Markov-Eigenschaft und Homogenität von X , wobei die Markov-Eigenschaft in der Form

$$P(X_{n+1} = j \mid \sigma(X_k : k \leq n)) = P(X_{n+1} = j \mid \sigma(X_n)) = p(X_n, j)$$

für beliebige Startverteilungen eingesetzt wurde; siehe [Asm87], S. 3.)

$$\begin{aligned} &= \sum_{n=1}^{\infty} \sum_{k \in E} \mathbb{E}_{i_0} (1_{(\varrho_1(i_0) > n-1)} \cdot p(X_{n-1}, j) \cdot 1_{(X_{n-1}=k)}) \\ &= \sum_{n=1}^{\infty} \sum_{k \in E} \mathbb{E}_{i_0} (1_{(\varrho_1(i_0) > n-1)} \cdot 1_{(X_{n-1}=k)} \cdot p(k, j)) \\ &= \sum_{k \in E} p(k, j) \sum_{n=1}^{\infty} \mathbb{E}_{i_0} (1_{(\varrho_1(i_0) > n-1)} \cdot 1_{(X_{n-1}=k)}) \end{aligned}$$

$$= \sum_{k \in E} p(k, j) \nu_k.$$

Wir haben also mit $\nu = (\nu_j : j \in E)$ gezeigt: $\nu = \nu \cdot p$. Zu zeigen bleibt noch $\nu_j < \infty$ für alle $j \in E$:

- a) Falls $j \notin C(i_0)$ ist, gilt $\nu_j = 0$.
- b) Sei $j \in C(i_0)$, d. h. es gibt ein $m \geq 1$ mit $p^{(m)}(j, i_0) > 0$. Es folgt:

$$\nu_{i_0} = \sum_{k \in E} \nu_k p^{(m)}(k, i_0) \geq \nu_j p^{(m)}(j, i_0)$$

und damit $\nu_j < \infty$. ■

Die Form des Beweises stammt aus [Asm87] (Theorem 3.2) und ist typisch für viele neuere Darstellungen. Die jeweiligen Umformungen geben jeweils intuitiv einleuchtende Sachverhalte wider; die hier angegebenen Erläuterungen sollen klar machen, welche Rechenregeln bei derartigen Schritten im Hintergrund stehen. (Das Buch [Asm87] von Asmussen scheint sich zu einem auch bei Anwendern zitierten Referenzwerk für angewandte Markov-Prozesse zu entwickeln.)

4.12 Satz

X sei irreduzibel und rekurrent. Dann existiert ein strikt positives Maß ν , welches stationär für *X* ist. ν ist bis auf einen multiplikativen Faktor eindeutig bestimmt, d. h. ist ν' ein weiteres stationäres Maß für *X*, so gilt $\nu = \nu' \cdot c$ für ein $c \in (0, \infty)$.

X ist genau dann positiv rekurrent, wenn ν endlich, d. h. zu einem Wahrscheinlichkeitsmaß normierbar ist.

Beweis: Sei $i_0 \in E$ beliebig, aber fest. Aus der Irreduzibilität folgt, daß für alle $j \in E$ die in Satz 4.11 definierten Größen $\nu_j(i_0) > 0$ sind. Die Endlichkeit der $\nu_j(i_0)$ war ebenfalls in Satz 4.11 gezeigt.

$\nu(i_0)$ ist genau dann normierbar, wenn gilt

$$\begin{aligned} \infty &> \sum_{j \in E} \nu_j(i_0) = \sum_{j \in E} \mathbf{E}_{i_0} \left(\sum_{n=0}^{\varrho_1(i_0)-1} 1_{(X_n=j)} \right) = \mathbf{E}_{i_0} \left(\sum_{n=0}^{\varrho_1(i_0)-1} \sum_{j \in E} 1_{(X_n=j)} \right) \\ &= \mathbf{E}_{i_0} \left(\sum_{n=0}^{\varrho_1(i_0)-1} 1 \right) = \mathbf{E}_{i_0}(\varrho_1(i_0)) = m_{i_0 i_0} \end{aligned}$$

(siehe Definition 3.34). Nach Definition ist dies aber genau dann der Fall, wenn *X* positiv rekurrent ist.

Ist ν' ein weiteres stationäres Maß für *X* mit $\nu' \neq 0$, dann ist auch ν' strikt positiv, denn aus $\nu' > 0$, d.h., $\nu'(j) > 0$ für ein $j \in E$, folgt für $i \in E$:

$$\nu'_i = \sum_{k \in E} \nu'_k p^{(m)}(k, i) \geq \nu'_j p^{(m)}(j, i)$$

für geeignetes $m > 0$.

Weiter folgt daraus insbesondere: Ist ν^* ein stationäres Maß für *X* mit $\nu^*(j) = 0$ für ein $j \in E$, so gilt $\nu^* \equiv 0$.

Wir nutzen dies aus, um in zwei Schritten (bis auf einen Faktor) die Eindeutigkeit zu zeigen:

- a) Sei ν' stationär für *X* mit $\nu'_{i_0} \geq 1 = \nu_{i_0}(i_0)$. Dann ist $\nu'_j \geq \nu_j(i_0)$ für alle $j \in E$ (koordinatenweise geordnet).

Sei \tilde{p} die Matrix, die aus *p* entsteht, indem die i_0 -te Spalte durch 0 ersetzt wird. Dann gilt für alle $n \geq 1$

$$\tilde{p}^n(k, j) = P_k(X_n = j, \varrho_1(i_0) > n)$$

(Tabu-Wahrscheinlichkeit für den n -stufigen Übergang von k nach j unter dem Tabu i_0). Nach Definition von $\nu(i_0)$ haben wir

$$\nu(i_0) = \varepsilon_{i_0} \cdot \sum_{n=0}^{\infty} \tilde{p}^n,$$

wobei $\varepsilon_{i_0}(j) = \delta_{i_0,j}$ die Einpunktverteilung in i_0 ist. Weiter gilt:

$$\nu' \geq \varepsilon_{i_0} + \nu' \cdot \tilde{p}.$$

Es ist nämlich $\nu'_{i_0} \geq 1 + 0$ für i_0 und $(\nu' \tilde{p})_j = (\nu' p)_j = \nu'_j$ für $j \neq i_0$. Daraus folgt jetzt direkt:

$$\nu' \geq \varepsilon_{i_0} + \nu' \tilde{p} \geq \varepsilon_{i_0} + (\varepsilon_{i_0} + \nu' \tilde{p}) \tilde{p} = \varepsilon_{i_0} (I + \tilde{p}) + \nu' \tilde{p}^2 \geq \dots \geq \varepsilon_{i_0} \sum_{n=0}^{N-1} \tilde{p}^n + \nu' \tilde{p}^N, \quad N \geq 2,$$

und für $N \rightarrow \infty$ folgt

$$\nu' \geq \varepsilon_{i_0} \sum_{n=0}^{\infty} \tilde{p}^n = \nu(i_0).$$

- b) Sei ν' stationär für X mit $\nu'_j > 0$ für alle $j \in E$. Dann kann $\nu'_{i_0} = 1$ gesetzt werden. Nach a) folgt dann aber $\nu'_j \geq \nu_j(i_0)$ für alle $j \in E$. Damit ist auch $\nu' - \nu(i_0)$ ein nichtnegatives, stationäres Maß für X . Da $\nu'_{i_0} = \nu_{i_0}(i_0) = 1$ war, folgt $(\nu' - \nu(i_0))_{i_0} = 0$ und damit $\nu' - \nu(i_0) \equiv 0$. Damit waren aber ν' und $\nu(i_0)$ bis auf einen Faktor schon gleich gewesen. ■

4.13 Korollar

Sei X irreduzibel und positiv rekurrent. Dann besitzt X eine stationäre Verteilung π , die durch

$$\pi(j) = \frac{1}{E_{i_0}(\varrho_1(i_0))} E_{i_0} \left(\sum_{n=0}^{\varrho_1(i_0)-1} 1_{(X_n=j)} \right) = \frac{1}{E_j(\varrho_1(j))} = \frac{1}{m_{jj}}, \quad j \in E,$$

gegeben ist.

Beweis: Die erste Darstellung folgt für $\pi(j)$ aus Satz 4.12 durch Normierung von $\nu(i_0)$. Insbesondere folgt $\pi(i_0) = m_{i_0 i_0}^{-1}$. Ersetze i_0 durch j als Referenzzustand, um die zweite Darstellung über die Eindeutigkeitsaussage in Satz 4.12 zu gewinnen. ■

Im Satz 4.9 war eine Verbindung zwischen stationären und Grenzverteilungen angesprochen. Wir werden jetzt einfach zu verifizierende Kriterien für die Existenz asymptotischer Verteilungen angeben. Aus den Beispielen 4.7 b) und 4.8 ersehen wir, daß für periodische Markov-Ketten Konvergenz der p^n nicht zu erwarten ist. Wir untersuchen deshalb zunächst den aperiodischen Fall:

4.14 Satz

Sei $X = (X_n : n \in \mathbb{N})$, $X_n : (\Omega, \mathcal{F}, P) \rightarrow (E, \mathcal{P}(E))$, eine Markov-Kette, welche irreduzibel und aperiodisch ist. Dann gilt unabhängig von der Wahl der Startverteilung p_0 :

$$\lim_{n \rightarrow \infty} P_{p_0}(X_n = j) = \frac{1}{m_{jj}} \quad \text{existiert,}$$

wobei $m_{jj} := E_j(\varrho_1(j))$ die erwartete Rückkehrzeit nach j ist. Für den Fall $m_{jj} = \infty$ setzen wir dabei $m_{jj}^{-1} := 0$.

Beweis: Sei $X' = (X'_n : n \in \mathbb{N})$, $X'_n : (\Omega, \mathcal{F}, P) \rightarrow (E, \mathcal{P}(E))$, eine weitere Markov-Kette auf demselben unterliegenden Wahrscheinlichkeitsraum mit derselben Übergangsmatrix p wie X . Die Startverteilung von X' sei p'_0 und kann von p_0 verschieden sein. X und X' seien stochastisch unabhängig bezüglich P .

(Daß die Konstruktion von X und X' auf einem gemeinsamen zugrundeliegenden Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) möglich ist, sehen wir wie folgt:

Konstruiere gemäß Satz 3.7 b) sowohl X als auch X' mit dem dort angegebenen Verfahren. Dabei erhalten wir X als $X_n : (\hat{\Omega}, \hat{\mathcal{F}}, \hat{P}) \rightarrow (E, \mathcal{P}(E))$ und X' als $X'_n : (\hat{\Omega}', \hat{\mathcal{F}}', \hat{P}') \rightarrow (E, \mathcal{P}(E))$, $n \in \mathbb{N}$. Wähle als gemeinsamen unterliegenden Wahrscheinlichkeitsraum die unabhängige Koppelung $(\Omega, \mathcal{F}, P) := (\hat{\Omega} \times \hat{\Omega}', \hat{\mathcal{F}} \otimes \hat{\mathcal{F}}', \hat{P} \otimes \hat{P}')$ und gehe über zu:

$$X_1 := (X_n \circ \text{pr}_1 : n \in \mathbb{N}) \quad \text{und} \quad X_2 := (X'_n \circ \text{pr}_2 : n \in \mathbb{N}).$$

Nun haben X_1 und X_2 beziehungsweise X und X' die gleichen Verteilungen und sind voneinander unabhängig bezüglich $\hat{P} \otimes \hat{P}'$. Die pr_i ($i = 1, 2$) seien dabei die Projektionen auf $\hat{\Omega} \times \hat{\Omega}'$.

Wir fixieren einen festen (beliebigen) Referenzzustand $i_0 \in E$ und definieren $T := \inf\{n \geq 0 : X_n = X'_n = i_0\}$ als die Zeit des ersten Zusammentreffens von X und X' in i_0 . T ist als Ersteintrittszeit von (X, X') in $(i_0, i_0) \in E^2$ eine Stoppzeit für (X, X') .

Für das asymptotische Verhalten von (X_n) (und damit auch von (X'_n)) unterscheiden wir zwei Fälle.

- a) Fall I: $P(T = \infty) = 0$ für alle Startverteilungen p_0 von X und p'_0 von X' („Erfolgreiche Kopplung“). In diesem Fall ist i_0 — und wegen der Irreduzibilität damit auch alle anderen Zustände — rekurrent. Starten wir nämlich X mit $p_0 = \varepsilon_{i_0}$, so folgt für die erste Rückkehrzeit nach i_0

$$\varrho_1(i_0) \leq T \quad P\text{-fast sicher,}$$

also

$$P(\varrho_1(i_0) \in \mathbb{N}) \geq P(T \in \mathbb{N}) = 1 - P(T = \infty) = 1.$$

Zur Konvergenz der absoluten Wahrscheinlichkeiten $P_{p_0}(X_n = j)$ für $n \rightarrow \infty$:

Aus der starken Markov-Eigenschaft und der Tatsache, daß X und X' zur Zeit T im gleichen Zustand i_0 sind, folgt:

$$P(T \leq n, X_n = j) = P(T \leq n, X'_n = j) \quad \text{für alle } j \in E. \quad \textcircled{*}$$

Also gilt:

$$\begin{aligned} P(X_n = j) &= P(X_n = j, T \leq n) + P(X_n = j, T > n) \\ &= P(X'_n = j, T \leq n) + P(X_n = j, T > n) \\ &= P(X'_n = j) - P(X'_n = j, T > n) + P(X_n = j, T > n). \end{aligned}$$

- i) Ist jetzt $m_{j_0 j_0} < \infty$ für ein $j_0 \in E$, so ist j_0 positiv rekurrent, wegen der Irreduzibilität also auch alle anderen Zustände.

Für $j \in E$ setzen wir speziell $p'_0(j) = m_{jj}^{-1}$. Nach Korollar 4.13 ist diese spezielle Startverteilung stationäre Anfangsverteilung für X' , es gilt also

$$P_{p'_0}(X'_n = j) = p'_0(j) = \frac{1}{m_{jj}}, \quad j \in E.$$

Aus $P(T = \infty) = 0$ folgt $P(T > n) \rightarrow 0$ für $n \rightarrow \infty$ und damit

$$P_{p_0}(X_n = j) = \frac{1}{m_{jj}} - P_{p'_0}(X'_n = j, T > n) + P_{p_0}(X_n = j, T > n) \xrightarrow{n \rightarrow \infty} \frac{1}{m_{jj}}.$$

- ii) Ist jetzt $m_{j_0 j_0} = \infty$ für ein $j_0 \in E$, so ist j_0 und auch jeder andere Zustand nullrekurrent. Nach Satz 4.11 gilt dann noch für alle $j \in E$ und einen beliebigen, aber festen Referenzzustand j_0

$$0 < \nu_j(j_0) =: \nu_j < \infty$$

und ν ist stationäres Maß für X und X' . Sei $\emptyset \neq B \subseteq E$ mit $|B| < \infty$. Setze jetzt

$$p'_0(j) = \begin{cases} \frac{\nu_j}{\sum_{i \in B} \nu_i} & \text{für } j \in B, \\ 0 & \text{für } j \notin B \end{cases}$$

als Startverteilung für X' . Dann gilt (komponentenweise):

$$p'_0 \leq \left(\nu_j \cdot \left(\sum_{i \in B} \nu_i \right)^{-1} : j \in E \right) =: \nu_B,$$

und also auch

$$P_{p'_0}(X'_n = j) = (p'_0 \cdot p^n)(j) \leq (\nu_B \cdot p^n)(j) = \frac{(\nu \cdot p^n)(j)}{\sum_{i \in B} \nu_i} = \frac{\nu_j}{\sum_{i \in B} \nu_j} < \infty.$$

Damit erhalten wir wiederum

$$\begin{aligned} P_{p_0}(X_n = j) &= P_{\nu_B}(X'_n = j) + P_{p_0}(T > n, X_n = j) - P_{\nu_B}(T > n, X'_n = j) \\ &\leq \frac{\nu_j}{\sum_{i \in B} \nu_i} + P_{p_0}(T > n) \\ &\xrightarrow{n \rightarrow \infty} \frac{\nu_j}{\sum_{i \in B} \nu_i}. \end{aligned}$$

Wegen

$$m_{j_0 j_0} = \sum_{i \in E} \nu_i = \infty$$

wird der letzte Ausdruck für $B \uparrow E$ beliebig klein, also

$$P_{p_0}(X_n = i_0) \xrightarrow{n \rightarrow \infty} 0.$$

(Anmerkung: Falls die Markov-Ketten X und X' periodisch — mit dann gleicher Periode — sind, kann Fall I wegen $P(T = \infty) = 0$, $\forall p_0, p'_0$ nicht eintreten.)

- b) Fall II: $P(T = \infty) > 0$ für ein Paar von Anfangsverteilungen p_0, p'_0 („Erfolgreiche Koppelung“). Unter den angegebenen Bedingungen ist auch $Z = (Z_n : n \in \mathbb{N})$ mit $Z_n = (X_n, X'_n)$ für $n \in \mathbb{N}$ eine irreduzible, aperiodische Markov-Kette mit Zustandsraum E^2 , Startverteilung $p_0 \otimes p'_0$ und Übergangsmatrix

$$\tilde{p}((i, i'), (j, j')) = p(i, j) \cdot p'(i', j') \quad \text{für } (i, i'), (j, j') \in E^2.$$

Für Z gibt es nach Voraussetzung eine Startverteilung, nämlich $\bar{p} := p_0 \otimes p'_0$, für welche gilt:

Bei Start unter \bar{p} erreicht Z den Zustand (i_0, i_0) mit positiver Wahrscheinlichkeit nicht in endlicher Zeit, formal:

$$P_{\bar{p}}(Z_n = (i_0, i_0) \text{ für ein } n \in \mathbb{N}) = P_{\bar{p}}(T \in \mathbb{N}) = \sum_{(k, l) \in E^2} \bar{p}(k, l) P_{(k, l)}(T \in \mathbb{N}) < 1.$$

Es muß also mindestens einen Zustand $(k, l) \in E^2$ geben derart, daß gilt:

$$P(Z_n = (i_0, i_0) \text{ für ein } n \in \mathbb{N} \mid Z_0 = (k, l)) < 1 \quad (\text{und } \bar{p}(k, l) > 0).$$

Aus der Irreduzibilität von Z folgt, daß $(i_0, i_0) \in E^2$ ein transienter Zustand ist. Damit sind aber alle Zustände von Z transient.

Sei \tilde{p}_0 jetzt eine beliebige Startverteilung von X und es werde X' ebenfalls unter \tilde{p}_0 gestartet. Dann gilt

$$P_{p_0}^{\sim}(X_n = j) = \left(P_{p_0 \otimes \tilde{p}_0}^{\sim} \left(\underbrace{Z_n = (j, j)}_{\subseteq \{\tau_j \geq n\}} \right) \right)^{\frac{1}{2}} \leq \left(P_{p_0 \otimes \tilde{p}_0}^{\sim}(\tau_j \geq n) \right)^{\frac{1}{2}} \xrightarrow{n \rightarrow \infty} 0,$$

wobei τ_j die letzte Aufenthaltszeit von Z in (j, j) ist; für diese gilt nach Korollar 3.36 $P(\tau_j < \infty) = 1$ aufgrund der Transienz von (j, j) .

Wir haben noch $m_{jj} = \infty$ zu zeigen. Für transientes $j \in E$ ist dies klar. Sei also j rekurrent und $m_{jj} < \infty$. Dann gilt für die nach Korollar 4.13 bestimmte stationäre Anfangsverteilung $\pi(\bullet) = \frac{\nu(\bullet)(j)}{m_{jj}}$ von X

$$0 < \frac{1}{m_{jj}} = P_\pi(X_n = j) \xrightarrow{n \rightarrow \infty} 0. \quad \blacksquare$$

Der Beweis von Satz 4.14 folgt [Tho85] und ist ein Beispiel für die sogenannte „Kopplungsmethode“, die auch als „Döblins Methode“ bezeichnet wird. Sie wurde von Döblin 1938 eingeführt und seit etwa 1975 weiter angewendet und ausgebaut. (Zum Werk Döblins siehe [Lin91].) Ausführlichere Hinweise und Anwendungen finden sich in [Asm87].

Die Bezeichnung „Kopplungsmethode“ findet ihre Berechtigung aus der Vorstellung heraus, daß hier die Prozesse X und X' auf dem gleichen unterliegenden Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) definiert sind. Die Realisierung eines $\omega \in \Omega$ bestimmt also sowohl für X als auch für X' den realisierten Pfad $(X_t(\omega) : t \in \mathbb{N})$ bzw. $(X'_t(\omega) : t \in \mathbb{N})$.

Die hier verwendete Technik kann über \otimes als „Verteilungskopplung“ bezeichnet werden, da nach dem ersten Zusammentreffen von X und X' in i_0 die Zustandsverteilungen gleich sind.

Eine andere Variante der Kopplungsmethode koppelt die Prozesse „pfadweise“ in dem Sinne, daß nach dem ersten Zusammentreffen von X und X' in i_0 die Entwicklung der Prozesse gleich bleibt: Falls $T(\omega) = n_0$ für $\omega \in \Omega$ gilt, folgt $X_n(\omega) = X'_n(\omega)$ für alle $n \geq n_0$ (siehe [Asm87]).

Zur Vervollständigung der Übersicht behandeln wir noch nicht-rekurrente Zustände ohne die Irreduzibilitätsannahme:

4.15 Satz

Sei $j \in E$ ein transienter Zustand der Markov-Kette X . Dann gilt unabhängig von der Startverteilung p_0 von X :

$$\lim_{n \rightarrow \infty} P_{p_0}(X_n = j) = 0.$$

Beweis: Analog zu den auf Z angewendeten Argumenten kann auch hier argumentiert werden. \blacksquare

Damit haben wir die wesentlichen Aussagen für das asymptotische Verhalten der Übergangsmatrix ebenfalls schon berechnet.

4.16 Satz

Sei $X = (X_n : n \in \mathbb{N})$ eine homogene Markov-Kette mit Übergangsmatrix p auf E .

a) Ist $j \in E$ transient, so gilt:

$$\lim_{n \rightarrow \infty} p^{(n)}(i, j) = 0 \quad \text{für jedes } i \in E.$$

b) Sei $j \in E$ rekurrent mit Periode d_j und erwarteter Rückkehrzeit $m_{jj} \leq \infty$.

i) Ist j von i aus nicht erreichbar, so gilt:

$$p^{(n)}(i, j) = 0 \quad \text{für alle } n \in \mathbb{N}. \quad (1)$$

ii) Ist $i \in C_r(j) \subseteq C(j)$ (der Äquivalenzklasse von i), dann gilt:

$$\lim_{n \rightarrow \infty} p^{(nd_j+r)}(i, j) = \frac{d_j}{m_{jj}} \quad \text{und} \quad p^{(m)}(i, j) = 0, \quad \text{falls } m \not\equiv r \pmod{d_j}. \quad (2)$$

Insbesondere gilt also:

$$\lim_{n \rightarrow \infty} p^{(nd_j)}(j, j) = \frac{d_j}{m_{jj}}, \quad \text{und} \quad p^{(m)}(j, j) = 0, \quad m \notin d_j \cdot \mathbb{N}. \quad (3)$$

Beweis:

- a) Setze $p_0 = \varepsilon_i$ in Satz 4.15.
 b) i) folgt direkt aus der Definition.
 ii) Falls X nicht irreduzibel ist, betrachten wir die Markov-Kette mit Übergangsmatrix

$$p[j] = (p(k, l) : k, l \in C(j)),$$

welche eine stochastische Matrix auf $C(j)$ ist.

Ist j aperiodisch, so folgen (2) und (3) aus Satz 4.14 für alle $i \in C(j)$, da $d_j = 1$ ist.

Falls $d_j > 1$ ist, kann Satz 4.14 angewendet werden auf die homogene Markov-Kette $X[j, d_j]$ auf dem Zustandsraum $C_0(j)$ mit Übergangsmatrix

$$p^{d_j}[j]|_{C_0(j)} = (p^{(d_j)}(k, l) : k, l \in C_0(j)),$$

um

$$\lim_{n \rightarrow \infty} p^{(nd_j)}(j, j) = \frac{d_j}{m_{jj}}, \quad \text{und} \quad p^{(m)}(j, j) = 0, \quad m \notin d_j \cdot \mathbb{N},$$

für $j \in C_0(j)$ zu erhalten, also (3). Der Faktor d_j im Nenner entsteht dadurch, daß die mittlere Rückkehrzeit nach j in der Zeitskala von $X[j, d_j]$ gerade $\frac{m_{jj}}{d_j}$ ist.

Aus (3) folgt direkt (2), indem wir uns klar machen, daß es gerade r Schritte dauert, bis X im richtigen Zyklus für das asymptotische Verhalten bezüglich j ist. ■

4.17 Korollar

Für $i, j \in E$ existiert stets der Cesaro-Limes

$$\pi(i, j) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\nu=1}^n p^{(\nu)}(i, j) = \frac{f_{ij}^*}{m_{jj}},$$

wobei für transientes j gesetzt wird: $m_{jj} = \infty$. Insbesondere gilt:

$$\pi(i) := \pi(i, i) = \begin{cases} 0 & i \text{ ist transient oder nullrekurrent,} \\ \frac{1}{m_{ii}} & i \text{ ist positiv rekurrent.} \end{cases}$$

Beweis: Eigenschaften allgemeiner Cesaro-Summierbarkeit, Beweis siehe z. B. [Hun83], S. 86). ■

4.18 Satz

Positive Rekurrenz und Nullrekurrenz sind Klasseigenschaften.

Beweis: Sei $i \leftrightarrow j$, i und j rekurrent sowie $p^{(m)}(i, j) > 0$, $p^{(n)}(j, i) > 0$. Sei d die Periode von i und j . Dann gilt für $l \in \mathbb{N}_+$:

$$p^{(m+ld+n)}(i, i) \geq p^{(m)}(i, j) \cdot p^{(ld)}(j, j) \cdot p^{(n)}(j, i).$$

Ist j positiv rekurrent, so hat für $l \rightarrow \infty$ die rechte Seite einen positiven Grenzwert, also auch die linke Seite, so daß i nicht nullrekurrent sein kann. Ist j nullrekurrent, so vertausche man in obiger Gleichung i und j . Dann geht für $l \rightarrow \infty$ die linke Seite gegen 0, so daß auch i nicht positiv rekurrent sein kann. ■

4.19 Korollar

X sei irreduzible Markov-Kette mit endlichem Zustandsraum E . Dann gilt:

- a) X hat keine transienten Zustände.
 b) X hat keine nullrekurrenten Zustände.

Beweis:

- b) Da X irreduzibel ist, können nur alle Zustände gleichzeitig nullrekurrent sein. Dies sei angenommen, d. h.

$$\lim_{n \rightarrow \infty} p^{(n)}(i, j) = 0 \quad \text{für alle } i, j \in E.$$

Mit

$$\sum_{j \in E} p^{(n)}(i, j) = 1 \quad \text{für jedes } n \geq 0$$

folgt:

$$1 = \lim_{n \rightarrow \infty} \sum_{j \in E} p^{(n)}(i, j) = \sum_{j \in E} \lim_{n \rightarrow \infty} p^{(n)}(i, j) = \sum_{j \in E} 0 = 0.$$

- a) läuft analog. ■

Aus dem Korollar 4.19 können wir sofort folgern:

4.20 Korollar

Da die nullrekurrenten Zustände abgeschlossene Klassen bilden, kann eine endliche Markov-Kette keine nullrekurrenten Zustände haben. Außerdem können nicht alle Zustände einer endlichen Markov-Kette transient sein.

Im Korollar 4.5 und den Beispielen in 4.6 hatten wir einen engen Zusammenhang zwischen stationären und Grenzverteilungen festgestellt, nach Satz 4.9 hängen wiederum Grenzverteilungen und die Limiten der Matrixpotenzen $p^{(n)}$ eng zusammen. Dies haben wir quantitativ erneut im Satzpaar 4.14, 4.16 wiedergefunden.

Die effektive Bestimmung invarianter Maße geschieht in der Regel über die Lösung der Stationaritätsgleichung (Definition 4.2), da die im Beweis von Satz 4.11 angegebenen stationären Verteilungen in der Regel nicht direkt zugänglich sind. Die dortige Darstellung hat allerdings große praktische Bedeutung als Begründung der regenerativen Simulation.

Den für die Anwendungen wichtigsten Fall haben wir schon bewiesen. Wir formulieren ihn hier noch einmal in der üblichen Terminologie.

4.21 Definition

Eine irreduzible, positiv rekurrente und aperiodische Markov-Kette heißt ergodisch.

4.22 Korollar (Ergodensatz für Markov-Ketten)

- a) Sei $X = (X_n : N \in \mathbb{N})$ eine ergodische Markov-Kette. Dann besitzt X eine eindeutig bestimmte stationäre Grenzverteilung $\pi = (\pi(i) : i \in E)$, für die gilt:

$$\pi(i) = \frac{1}{m_{ii}} \quad \text{für } i \in E$$

und

$$\lim_{n \rightarrow \infty} p^{(n)}(j, i) = \pi(i) \quad \text{für } i \in E.$$

- b) Sei $X = (X_n : n \in \mathbb{N})$ eine irreduzible Markov-Kette, die aperiodisch ist. Dann ist X ergodisch genau dann, wenn die Stationaritätsgleichung $q = q \cdot p$ eine Wahrscheinlichkeitslösung besitzt.

Beweis:

- a) ist gerade Satz 4.14 und Satz 4.16 b), Gleichung (2).
 b) Aus Satz 4.11 und Satz 4.14 folgt: Ergodizität impliziert die eindeutige Lösbarkeit von $q = q \cdot p$ durch ein Wahrscheinlichkeitsmaß. Besitzt $q = q \cdot p$ eine Wahrscheinlichkeitslösung, so kann X keine transienten Zustände haben, da nach Satz 4.15 dann

$$\lim_{n \rightarrow \infty} P_q(X_n = j) = 0 \quad \text{für } j \in E$$

sein müßte. Unter der somit gesicherten Annahme von Rekurrenz für X gibt Satz 4.12 bis auf einen Faktor das stationäre Maß für X an. Dieses ist nach der hier gesetzten Voraussetzung normierbar; nach Satz 4.12 ist X damit positiv rekurrent. ■

Eine genauere Untersuchung der Stationaritätsgleichung folgt:

4.23 Satz

Sei $C \subseteq E$ eine irreduzible (abgeschlossene) Klasse. Die einzigen Lösungen des linearen Gleichungssystems

$$u(i) = \sum_{k \in C} u(k)p(k, i), \quad i \in C, \quad (1)$$

die $\sum_{i \in C} |u(i)| < \infty$ erfüllen, sind durch

$$u(i) = c \cdot \pi(i) = c \cdot \pi(i, i), \quad i \in C, \quad c \in \mathbb{R},$$

gegeben (siehe Korollare 4.13 und 4.17). Ist C transient oder nullrekurrent, so gilt

$$u(i) = \pi(i) = 0, \quad i \in C;$$

ist C positiv rekurrent mit Periode $d \geq 1$ und Unterklassen C_0, C_1, \dots, C_{d-1} , so gilt für $r = 0, 1, \dots, d-1$

$$\sum_{i \in C_r} \pi(i) = \frac{1}{d} \quad (2)$$

und

$$\sum_{i \in C} \pi(i) = 1. \quad (3)$$

(Dabei ist $\pi(i) = m_{ii}^{-1} > 0$ nach Korollar 4.22.)

Beweis:

- i) Sei C positiv rekurrent; wir zeigen, daß die $(\pi(i) : i \in E)$ Gleichung (1) lösen. (In den anderen Fällen ist dies trivial.)

Es gilt nach Satz 4.16 b) ii)

$$\begin{aligned} d \cdot \pi(i) &= \lim_{n \rightarrow \infty} p^{(nd)}(i, i) = \lim_{n \rightarrow \infty} \sum_{k \in C_{-1} \pmod{d}(i)} p^{(nd-1)}(i, k)p(k, i) \\ &\stackrel{\text{Fatou}}{\geq} \sum_{k \in C_{-1} \pmod{d}(i)} \lim_{n \rightarrow \infty} p^{(nd-1)}(i, k)p(k, i) = \sum_{k \in C_{-1} \pmod{d}(i)} d \cdot \pi(k) \cdot p(k, i). \end{aligned}$$

Also

$$\pi(i) \geq \sum_{k \in C} \pi(k)p(k, i), \quad i \in C. \quad (*_1)$$

Weiter gilt für $i \in C$ und $r = 0, 1, \dots, d-1$:

$$1 = \lim_{n \rightarrow \infty} \sum_{j \in C_r(i)} p^{(nd+r)}(i, j) \geq \sum_{j \in C_r(i)} \lim_{n \rightarrow \infty} p^{(nd+r)}(i, j) = \sum_{j \in C_r(i)} d\pi(j),$$

also

$$\sum_{j \in C_r(i)} \pi(j) \leq \frac{1}{d} \quad \text{und} \quad \sum_{j \in C} \pi(j) \leq 1.$$

Diese Abschätzung erlaubt die Vertauschung der Summationsreihenfolge in

$$\sum_{i \in C} \pi(i) \stackrel{(*_1)}{\geq} \sum_{i \in C} \sum_{k \in C} \pi(k)p(k, i) = \sum_{k \in C} \pi(k) \sum_{i \in C} p(k, i) = \sum_{k \in C} \pi(k).$$

Mit $\pi(i) \geq 0$ folgt also aus $(*_1)$:

$$\pi(i) = \sum_{k \in C} \pi(k)p(k, i), \quad i \in C.$$

$(\pi(i) : i \in C)$ löst also Gleichung (1).

- ii) Sei $(u(i) : i \in C)$ eine weitere Lösung, die $\sum_{i \in C} |u(i)| < \infty$ erfüllt. Aus Gleichung (1) erhält man durch Iteration für $i \in E$:

$$u(i) = \sum_{k \in C_{-r \pmod{d}}(i)} u(k) p^{(nd+r)}(k, i), \quad n \in \mathbb{N}, \quad r \in \{0, 1, \dots, d-1\}. \quad (*_2)$$

Limesbildung in $(*_2)$, Vertauschen von Summe und Limes nach dem Satz von der majorisierten Konvergenz liefert mit Satz 4.16

$$u(i) = \left(\sum_{k \in C_{-r \pmod{d}}(i)} u(k) \right) d\pi(i), \quad r \in \{0, 1, \dots, d-1\}, \quad i \in E. \quad (*_3)$$

Ist i nullrekurrent oder transient, so ist mit $\pi(i) = 0$ auch $u(i) = 0$.

Ist i positiv rekurrent, so gilt $\pi(j) > 0$ für $j \in E$.

Da $(*_3)$ für $r \in \{0, 1, \dots, d-1\}$ gilt, hängt $\left(\sum_{k \in C_{-r \pmod{d}}(i)} u(k) \right)$ nicht von r oder i ab, von i deshalb nicht, da die Klasseneinteilung $C_0(i), \dots, C_{d-1}(i)$ nicht vom Repräsentanten i abhängt. Damit ist

$$\pi(i) \cdot c = u(i), \quad i \in E.$$

- iii) Um die Aussagen (2) und (3) zu beweisen, ersetzen wir im positiv rekurrenten Fall in $(*_3)$ für $k \in C$ die $u(k)$ durch $\pi(k)$. Division durch $d \cdot \pi(i)$ liefert (2) und damit (3). ■

Mit diesem Hilfsmittel können wir Korollar 4.22 noch etwas verschärfen, indem wir auf Aperiodizität als Voraussetzung verzichten.

4.24 Satz

Sei X eine irreduzible Markov-Kette. X ist genau dann positiv rekurrent, wenn eine der folgenden Bedingungen erfüllt ist:

- i) X besitzt ein invariantes Wahrscheinlichkeitsmaß.
- ii) Jede nichtverschwindende Lösung $x = (x(i) : i \in E)$ des Gleichungssystems $x = xp$ mit der Eigenschaft $\sum_{i \in E} |x(i)| < \infty$ ist von der Form $x(i) = c \cdot m_{ii}^{-1}$, $c \neq 0$.

Zerfällt der Zustandsraum von X in mehrere Klassen, so wird durch die Startverteilung p_0 von X festgelegt, wieviel „Wahrscheinlichkeits-Masse“ auf den einzelnen Klassen liegt. Ist dann die Klasse A abgeschlossen und positiv rekurrent, so kann (bis auf Normierung) auf die Markov-Kette, eingeschränkt auf A , der Satz 4.24 angewendet werden.

4.25 Korollar

Sei X eine Markov-Kette mit Startverteilung p_0 und Übergangsmatrix p . Seien A_α , $\alpha \in D$, die positiv rekurrenten Klassen von X mit $A_\alpha \neq A_{\alpha'}$, falls $\alpha \neq \alpha'$, und $\bigcup_{\alpha \in D} A_\alpha = A$.

X ist stationär genau dann, wenn es ein Wahrscheinlichkeitsmaß $(\lambda(\alpha) : \alpha \in D)$ gibt, so daß gilt:

$$p_0(i) = \begin{cases} 0, & \text{falls } i \notin A; \\ \lambda(\alpha)\pi(i), & \text{falls } i \in A_\alpha \text{ für ein } \alpha \in A. \end{cases}$$

4.26 Beispiel

Im Beispiel 3.39 hatten wir gezeigt, daß die Irrfahrt auf \mathbb{Z} rekurrent ist genau dann, wenn die Wahrscheinlichkeit $\lambda = 0.5$ war. Durch Rechnung hatten wir sogar Nullrekurrenz nachgewiesen. Wir können jetzt einfacher und anschaulich argumentieren:

Die räumlich auf \mathbb{Z} homogene Übergangsmatrix verlangt, daß im Falle positiver Rekurrenz die Rückkehrzeitverteilungen in jeden Zustand gleich sind, insbesondere auch ihre Erwartungswerte. Wären diese

endlich ($m_{ii} < \infty$), so wäre $x(i) = m_{ii}^{-1}$ von i unabhängig und $\sum_{i \in \mathbb{Z}} |x(i)| = \infty$. Also ist X nullrekurrent. \square

4.27 Beispiel

- a) Restlebensdauerprozeß (Fortsetzung von Beispiel 3.10)

Der Zustandsraum für den Restlebensdauerprozeß kann als \mathbb{N} angenommen werden. Dann gibt es aber möglicherweise Zustände, die nicht von anderen aus erreichbar sind, so daß keine Irreduzibilität vorliegt. E wird deshalb eingeschränkt auf

$$\{1, 2, 3, \dots, \sup\{k \in \mathbb{N} : g_k > 0 \vee f_k > 0\}\}.$$

Falls $\sup\{k : g_k > 0\} > \sup\{k : f_k > 0\}$ ist, tritt das folgende Phänomen auf: Nach dem Start des Prozesses treten mit positiver Wahrscheinlichkeit Zustände auf, die der Prozeß nie wieder erreichen kann, nachdem er sie verlassen hat. Hat X aber erst den Raum

$$\{1, 2, \dots, \sup\{k \in \mathbb{N} : f_k > 0\}\}$$

erreicht, so verläßt er E nicht mehr; E ist also abgeschlossen und außerdem irreduzibel. (Zustände aus $\{\sup\{k : f_k > 0\} + 1, \dots, \sup\{k : g_k > 0\}\}$ werden auch als unwesentlich bezeichnet; siehe [Chu67], S. 13. Unwesentliche Zustände haben in der Regel für des Langzeitverhalten des Prozesses keine Bedeutung.)

Wir nehmen im folgenden an, daß gilt:

$$E(g_{(\bullet)}) < \infty \quad \text{und} \quad \sup\{k : f_k > 0\} \geq \sup\{k : g_k > 0\}.$$

Dann ist X irreduzibel.

X ist periodisch mit Periode $d > 1$ genau dann, wenn d die kleinste Zahl ist mit:

$$f_k > 0 \implies k \in d \cdot \mathbb{N}, \quad k \in \mathbb{N}.$$

X ist transient genau dann, wenn $\sum_{k \in \mathbb{N}} f_k < 1$ ist, rekurrent also genau dann, wenn $\sum_{k \in \mathbb{N}} f_k = 1$ gilt.

Der Zustand 1 ist positiv rekurrent genau dann, wenn gilt $\sum_{k \in \mathbb{N}} k \cdot f_k < \infty$.

Falls X positiv rekurrent ist, existiert eine eindeutig bestimmte stationäre Verteilung, erhalten als Lösung des Systems $x = x \cdot p$, die im aperiodischen Fall auch eindeutige Grenzverteilung ist. Explizit erhalten wir als Gleichgewichtsbedingung:

$$x(i) = x(i+1) \cdot 1 + x(1) \cdot f_i, \quad i \geq 1.$$

Durch Einsetzen überprüft man, daß das System gelöst wird durch

$$x(i) = \frac{1}{\mu} \left(1 - \sum_{k=1}^{i-1} f_k \right), \quad i \geq 1,$$

wobei $\mu = \sum_{l=0}^{\infty} l \cdot f_l$ die mittlere Arbeitszeit (Lebensdauer) ist.

- b) Führen wir analog zu Beispiel 3.10 gemäß Korollar 3.12 b) die Modellierung und Untersuchung für den Altersprozeß durch, erhalten wir, daß auch die eindeutig bestimmte stationäre Altersverteilung die in a) errechnete Zähldichte hat. \square

4.28 Beispiel (M/M/1/∞-FCFS in diskreter Zeit²)

- a) Wir untersuchen das Beispiel 3.17 mit geometrisch verteilten Bedienungszeiten. Da dies gerade die Verteilung der Zeit bis zum ersten Erfolg in einer unabhängigen Folge von Bernoulli-Experimenten ist, benötigen wir für eine Markovsche Beschreibung des Systemablaufs keine Alters- bzw. Restbedienungszeitvariable mitzuführen. Der Schlangenlängenprozeß ist Markovsch mit Übergangsmatrix

$$\begin{aligned} p(0,0) &= 1-p, \\ p(0,1) &= p, \\ p(i,i-1) &= (1-p)q, & i \geq 1 \\ p(i,i) &= pq + (1-p)(1-q), & i \geq 1 \\ p(i,i+1) &= p(1-q), & i \geq 1 \end{aligned}$$

Das Gleichungssystem $x = x \cdot p$ hat dann die Form

$$\begin{aligned} x(0) &= x(0)p(0,0) + x(1)p(1,0) \\ x(i) &= x(i-1)p(i-1,i) + x(i)p(i,i) + x(i+1)p(i+1,i), \quad i \geq 1, \end{aligned}$$

und ist damit sukzessiv lösbar, wenn $x(0) > 0$ als Parameter betrachtet wird. Es ergibt sich

$$x(i) = x(0) \frac{1}{1-q} \cdot \left(\frac{p(1-q)}{(1-p)q} \right)^i, \quad i \geq 1.$$

$x = (x(i) : i \in \mathbb{N})$ läßt sich also genau dann zu einem Wahrscheinlichkeitsvektor normieren, wenn $\frac{p(1-q)}{(1-p)q} < 1$, d. h. $\frac{p}{q} < 1$ gilt.

Dies ist einfach interpretierbar, denn für eine geometrisch auf $\{1, 2, 3, \dots\}$ verteilte Zufallsvariable mit Parameter $p \in (0, 1)$ ist der Mittelwert gerade p^{-1} . Die obige Rekurrenzbedingung heißt also $q^{-1} < p^{-1}$, d. h. die mittlere Bedienungszeit ist kleiner als die mittlere Zwischenankunftszeit. Korollar 4.22 b) liefert also das notwendige und hinreichende Kriterium $\frac{p}{q} < 1$ für positive Rekurrenz des Schlangenlängenprozesses X . $\frac{p}{q}$ ist ein Maß für die Belastung des Systems und wird als Verkehrsintensität bezeichnet. Wir erhalten

$$x(0) = 1 - \frac{p}{q} \quad \text{und} \quad x(i) = \left(1 - \frac{p}{q} \right) \frac{1}{1-q} \left(\frac{p(1-q)}{(1-p)q} \right)^i, \quad i \geq 1.$$

- b) Außerdem kann gezeigt werden:

$$\frac{p}{q} > 1 \iff X \text{ transient}; \quad \frac{p}{q} = 1 \iff X \text{ nullrekurrent.}$$

- c) Läßt man wieder beliebig verteilte Bedienungszeiten zu mit mittlerer Bedienungszeit q^{-1} , so sind die in a) und b) formulierten Charakterisierungen für den in Beispiel 3.17 definierten Zustandsprozeß zutreffend. \square

4.29 Beispiel

- a) In einem M/M/1/∞-FCFS-System werden durch wartende Kunden Kosten verursacht. Warten in der Schlange $n-1 \geq 0$ Kunden, die nicht bedient werden, d. h. ist $X_{(\bullet)} = n \geq 1$, so entstehen Kosten

$f(n) \in \mathbb{R}$. Der Kostenprozeß ist dann $f(X) = (f(X_n) : n \in \mathbb{N})$, und wir definieren $\sum_{k=0}^n f(X_k)$ als in

$\{0, 1, \dots, n\}$ kumulierte Kosten und $\frac{1}{n+1} \sum_{k=0}^n f(X_k)$ als mittlere Kosten bis n bzw. in $\{0, 1, \dots, n\}$.

- b) Entsprechend zu a) kann zu beliebigen Markov-Ketten die Definition beliebiger Funktionen vorgenommen werden. Diese sind dann zwar wieder stochastische Prozesse, da auf dem Zustandsraum E die Potenzmengen- σ -Algebra als Ereignis- σ -Algebra angenommen war, im allgemeinen jedoch nicht Markovsch.
- c) Ein Spezialfall: $X = (X_n : n \in \mathbb{N})$ sei eine unabhängig identisch verteilte Folge, also eine spezielle Markov-Kette, und $f : E \rightarrow D$ eine Abbildung. Dann ist auch $f(X) := (f(X_n) : n \in \mathbb{N})$ eine unabhängig identisch verteilte Folge, somit wieder eine Markov-Kette. Das schwache Gesetz der großen Zahlen liefert

$$\frac{1}{n+1} \sum_{k=0}^n f(X_k) \xrightarrow{\text{stoch.}} \mathbb{E}(f(X_0))$$

(falls z. B. $\text{Var}(f(X_0)) < \infty$).

(Nach dem starken Gesetz der großen Zahlen gilt sogar P -fast sichere Konvergenz.) \square

Der zentrale Satz zur Behandlung derartiger Probleme wird bewiesen z. B. in [Chu67], S. 91 und Kapitel 14. Die umfangreichen notwendigen Vorbereitungen des Beweises beruhen im wesentlichen auf Folgerungen aus der starken Markov-Eigenschaft.

4.30 Satz

Sei $X = (X_n : n \in \mathbb{N})$ eine irreduzible, rekurrente Markov-Kette mit Zustandsraum E , und seien $f, g : E \rightarrow \mathbb{R}$ zwei Funktionen derart, daß mit der Lösung $(u(i) : i \in E)$, $u(i) > 0$, $i \in E$, des Systems $u = u \cdot p$ (Stationaritätsgleichung, siehe Satz 4.23) gilt: $\sum_{j \in E} u(j)f(j)$ und $\sum_{j \in E} u(j)g(j)$ sind endlich und nicht gleichzeitig 0. Dann gilt (für beliebige Startverteilung)

$$\lim_{n \rightarrow \infty} \frac{\sum_{s=0}^n f(X_s)}{\sum_{s=0}^n g(X_s)} = \frac{\sum_{j \in E} u(j)f(j)}{\sum_{j \in E} u(j)g(j)} \quad P\text{-fast sicher.} \quad (1)$$

4.31 Korollar

X sei irreduzibel und rekurrent mit stationärem Maß ν , gegeben nach Satz 4.11 und Satz 4.12 (für einen beliebigen, aber festen Referenzzustand). Dann gilt für alle $i, j, k, l \in E$:

$$\frac{\sum_{m=0}^n p^{(m)}(i, j)}{\sum_{m=0}^n p^{(m)}(l, k)} \xrightarrow{n \rightarrow \infty} \frac{\nu_j}{\nu_k}.$$

Beweis: Es gibt (siehe [Cin75], S. 160) eine zu Satz 4.30 analoge Aussage für Erwartungswerte:

$$\lim_{n \rightarrow \infty} \frac{\sum_{s=0}^n \mathbb{E} f(X_s)}{\sum_{s=0}^n \mathbb{E} g(X_s)} = \frac{\sum_{j \in E} u(j)f(j)}{\sum_{j \in E} u(j)g(j)},$$

unabhängig von der Startverteilung. Setze in diesem Ausdruck $f = 1_{\{j\}}$ und $g = 1_{\{k\}}$. (Für einen kürzeren direkten Beweis ohne Rückgriff auf den allgemeinen Satz siehe [Asm87], Prop. 4.4, S. 16.) \blacksquare

²siehe Beispiel 3.17

(Für einen kürzeren, direkten Beweis von Korollar 4.31, ohne Satz 4.30 zu benutzen, siehe [Asm87], Prop. 4.4, S. 16.)

4.32 Satz (Ergodensatz)

Mit den Bezeichnungen aus Satz 4.30 sei X positiv rekurrent.

a) Dann gilt:

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{s=0}^n f(X_s) = \sum_{j \in E} \pi(j) f(j) \quad P\text{-fast sicher,}$$

falls die rechte Seite absolut konvergiert, wobei $\pi = (\pi(j) : j \in E)$ die stationäre Verteilung ist.

b) Sei $\tau_n(i)$, $n = 0, 1, 2, \dots$, die Folge der Eintrittszeiten von X in i (siehe Definition 3.31 und Korollar 3.32), $N(i) = (N_n(i) : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{N})$ der zum Erneuerungsprozeß $(\tau_n(i) : n = 0, 1, \dots)$ gehörige Zählprozeß, d. h.

$$N_n(i) = \sup\{k \in \mathbb{N} : \tau_k(i) \leq n\}, \quad n = 0, 1, \dots$$

(siehe Beispiel 3.11). Dann gilt für $i \in E$

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} N_n(i) = \pi(i) \quad P\text{-fast sicher.}$$

c) Für $i \in E$ gilt

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \tau_{N_n(i)}(i) = 1 \quad P\text{-fast sicher.}$$

Beweis: In Satz 4.30 wird $u = \pi$ gesetzt gemäß Satz 4.12 und Satz 4.24. Setzen wir dann in Gleichung (1) von Satz 4.30 $g \equiv 1$, so folgt a).

Mit $f(j) = 1_{\{i\}}(j)$, $j \in E$, und $g \equiv 1$ folgt mit $N_n(i) = \sum_{k=0}^n 1_{\{i\}}(X_k)$ die Behauptung b).

Um c) zu zeigen, zerlegen wir

$$\frac{1}{n+1} \tau_{N_n(i)}(i) = \underbrace{\frac{N_n(i)}{n+1}}_{\rightarrow \pi(i) \text{ nach b)}} \cdot \underbrace{\frac{1}{N_n(i)} \sum_{k=1}^{N_n(i)} \varrho_k(i)}_{\rightarrow E(\varrho_1(i)) \text{ nach dem starken Gesetz der großen Zahlen, } P\text{-fast sicher}} \rightarrow \pi(i) \cdot m_{ii} = \pi(i) \frac{1}{\pi(i)} = 1. \quad \blacksquare$$

Sätze vom Typ 4.30, 4.31, 4.32 werden als „Ergodensätze“ bezeichnet, vergleiche auch Korollar 4.22, wo ein Spezialfall von Satz 4.32 a) behandelt wurde.

Insbesondere ist die Aussage

$$\frac{1}{n+1} \sum_{k=0}^n f(X_k) \rightarrow \sum_{j \in E} \pi(j) f(j) \quad P\text{-fast sicher}$$

in Satz 4.32 a) eine Behauptung, die derjenigen im Birkhoffschen Ergodensatz für stationäre Prozesse bzw. für dynamische Systeme entspricht.

Die anschauliche Interpretation dieses Satzes ist auch hier zutreffend: Asymptotisch ist das *zeitliche Mittel*

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{k=0}^n f(X_k(\omega))$$

für P -fast alle $\omega \in \Omega$ gleich dem *räumlichen Mittel*

$$\sum_{j \in E} f(j) \pi(j) = E_\pi(f(X_0)).$$

Die Herleitung eines solchen Verhaltens gelingt für physikalische Systeme, wenn sie einer sogenannten „(Quasi-)Ergodenhypothese“ genügen.

Für die Theorie und Praxis ergodischer Markov-Prozesse ist die vielleicht wichtigste Folgerung Satz 4.32 b): Im zeitlichen Mittel ist für (P -fast sicher) jeden Pfad $(X_n(\omega) : n \in \mathbb{N})$ die relative Häufigkeit der Aufenthalte in $i \in E$ gleich der Wahrscheinlichkeit, mit der sich X asymptotisch oder stationär in i aufhält.

Dies hat natürlich Anwendung beim Schätzen stationärer Verteilungen für Markov-Ketten:

$$\frac{1}{n+1} \sum_{k=0}^n 1_{\{i\}}(X_k)$$

ist ein geeigneter Schätzer für $\pi(i)$ aus einer Beobachtung (X_0, X_1, \dots, X_n) , die im Sinne der klassischen Stichprobentheorie keine Unabhängigkeit sichert für die Beobachtungen.

5 Erneuerungstheorie

Erneuerungs- und Ersatzprobleme haben wir in zeitdiskreten Markov-Ketten schon beschrieben (siehe Beispiel 3.10), außerdem in diskretem Kontext auch schon die allgemeinen Begriffe bereitgestellt (Beispiel 3.11). In diesem Kapitel werden wir, soweit möglich, diskrete und nichtdiskrete Modelle gleichzeitig behandeln. Sofern das nicht möglich ist, werden wir uns auf den nichtdiskreten Fall beschränken, da die entsprechenden Aussagen für den diskreten Fall meist Spezialfälle von Aussagen über Markov-Ketten sind.

„Diskrete“ Verteilungen in diesem Zusammenhang werden üblicherweise als „arithmetisch“ bezeichnet.

5.1 Definition

Ein Wahrscheinlichkeitsmaß P auf $(\overline{\mathbb{R}}_+, \overline{\mathbb{B}}_+)$ heißt arithmetisch mit Spanne (oder Span) λ , falls $\lambda > 0$ ist und $P(\lambda \cdot \mathbb{N}) = 1$ gilt. (Englisch: „lattice distribution“; nichtarithmetische Verteilungen heißen dann „non lattice“ distribution.)

5.2 Definition

$(X_i : (\Omega, \mathcal{F}, P) \rightarrow (\overline{\mathbb{R}}_+, \overline{\mathbb{B}}_+); i = 1, 2, \dots)$ sei eine Folge von unabhängig identisch verteilten Zufallsvariablen mit Verteilungsfunktion $F : [0, \infty) \rightarrow [0, 1]$, so daß $F(0) = 0$ gilt.

- a) Die Folge $(S_k : (\Omega, \mathcal{F}, P) \rightarrow (\overline{\mathbb{R}}_+, \overline{\mathbb{B}}_+); k = 0, 1, 2, \dots)$, gegeben durch

$$S_0 \equiv 0, \quad S_k = \sum_{i=1}^k X_i, \quad k \geq 1,$$

heißt Erneuerungsprozeß mit Erneuerungsverteilungsfunktion F . Die $(X_i : i \in \mathbb{N}_+)$ heißen Erneuerungsintervalle, die $(S_k : k \in \mathbb{N}_+)$ Erneuerungszeiten.

- b) Der stochastische Prozeß $N = (N_t : (\Omega, \mathcal{F}, P) \rightarrow \mathbb{N}_0; t \in \mathbb{R}_+)$, gegeben durch

$$N_t := \sum_{k=1}^{\infty} 1_{(S_k \leq t)} = \sup\{k \in \mathbb{N}_0 : S_k \leq t\},$$

heißt der zu $(S_k : k \in \mathbb{N})$ bzw. $(X_i : i \in \mathbb{N}_+)$ gehörige Zählprozeß.

5.3 Beispiel

- a) Zur Zeit 0 werde eine Sicherung mit Lebensdauerverteilungsfunktion $F : [0, \infty) \rightarrow [0, 1]$ in Betrieb gesetzt und bei Ausfall sofort (ohne Zeitverlust) durch eine neue Sicherung des gleichen Typs aus einer Massenproduktion ersetzt. Diese Voraussetzungen erlauben die Annahme, daß die Lebenszeiten der Sicherungen durch eine Folge von unabhängig identisch verteilten Zufallsvariablen modelliert werden.
- b) Zusätzlich zu den Annahmen in a) sei für das Einsetzen einer neuen Sicherung eine (zufällige) Zeit erforderlich, und die Folge der zur Auswechslung benötigten Zeiten sei unabhängig identisch verteilt und unabhängig von der Folge der Lebensdauern. Dann betrachte man jeweils den Zeitraum zwischen den Augenblicken, an denen gerade eine Sicherung eingesetzt ist.
- c) Häufig hat man bei Aufnahme des Betriebs in einem System keine Kenntnis darüber, wie lange das interessierende Teil eventuell vorher schon gearbeitet hat. Dann ist die Folge der (X_1, X_2, \dots) zwar unabhängig, aber X_1 hat eine andere Verteilung als $X_i, i \geq 2$.
- d) Die Zwischenankunftszeiten von Kunden an einem Bedienungssystem seien unabhängig identisch verteilte Zufallsvariablen. Besteht die Möglichkeit, daß nach einer endlichen Folge von Ankünften der Ankunftsstrom abbricht (mit positiver Wahrscheinlichkeit), so gilt für die Zwischenankunftsverteilungsfunktion (Erneuerungsverteilungsfunktion) F :

$$P(X_i \notin \mathbb{R}_+) > 0 \quad \text{und} \quad \lim_{t \rightarrow \infty} F(t) = 1 - P(X_i = \infty) < 1$$

(defekte Verteilungsfunktion auf \mathbb{R}_+). Dadurch ist $\overline{\mathbb{R}}_+$ als Zustandsraum des Prozesses $(X_i : i \in \mathbb{N}_+)$ motiviert. \square

5.4 Definition

Bei den in Definition 5.2 festgelegten Größen sei auch $X_1 \sim G$, $X_i \sim F$, $i \geq 2$, zugelassen. Dann wird die Folge

$$S_0 \equiv 0, \quad S_n = \sum_{i=1}^n X_i, \quad n \geq 1,$$

als verzögerter Erneuerungsprozeß bezeichnet. G heißt Verzögerungsdauerverteilung. Wir nehmen stets an: $G(\mathbb{R}_+) = 1$. Falls $G = F$ ist und dies betont werden soll, wird $S = (S_n : n \in \mathbb{N})$ als unverzögerter Erneuerungsprozeß bezeichnet.

Die Voraussetzung $F(0) = 0$ kann abgeschwächt werden zu $F(0) < 1$, d. h. ein neu eingesetztes Element fällt mit positiver Wahrscheinlichkeit sofort aus bzw. in der Interpretation als Ankunftsstrom treten Gruppenankünfte auf mit geometrisch verteilter Gruppengröße.

Wir verzichten auf diese Verallgemeinerung, um einfachere Rechnungen zu erhalten. Im folgenden seien stets die Bezeichnungen aus den Definitionen 5.2 und 5.4 sowie aus Beispiel 5.3 vorausgesetzt — auch als Modellannahmen für den jeweils anstehenden Problemfall. Dazu die folgende zentrale Begriffsbildung der Erneuerungstheorie:

5.5 Definition

Für den durch Definition 5.2 definierten Zählprozeß heißt

$$M : [0, \infty) \rightarrow \mathbb{R}_+, \quad t \mapsto E(N_t),$$

Erneuerungsfunktion (Erwartungswertfunktion).

Theoretisch ist die Erneuerungsfunktion einfach zu bestimmen, in der Praxis ergibt sich als großes Problem die numerische Auswertung der vorliegenden Ausdrücke.

5.6 Lemma

a) Die Verteilungsfunktion $F^{(k)}$ von S_k , $k \in \mathbb{N}$, ist gegeben als:

$$S_0 \sim F^{(0)} = 1_{[0, \infty)}; \quad S_1 \sim F^{(1)} = G; \quad S_k \sim F^{(k)} = G * F^{(k-1)*}, \quad k \geq 2,$$

wobei F^{n*} die n -fache Faltung von F bedeutet.

b) Die Zähldichte der eindimensionalen Randverteilung $\mathcal{L}(N_t)$ von $N = (N_t : t \in [0, \infty))$ ist gegeben durch

$$P(N_t = n) = \begin{cases} 1 - G(t) & \text{für } n = 0, \\ G * F^{(n-1)*}(t) - G * F^{n*}(t) & \text{für } n \geq 1. \end{cases}$$

c) Für $t \in [0, \infty)$ gilt $M(t) = \sum_{k=1}^{\infty} G * F^{(k-1)*}(t)$.

d) Für jedes $t \in [0, \infty)$ ist $M(t) < \infty$.

Beweis: Grundlegend sind die Mengengleichungen

$$\{N_t = n\} = \{S_n \leq t < S_{n+1}\} \quad \text{für alle } t \in [0, \infty), n \in \mathbb{N} \quad (1)$$

und

$$\{N_t \geq n\} = \{S_n \leq t\} \quad \text{für alle } t \in [0, \infty), n \in \mathbb{N}. \quad (2)$$

b) folgt damit aus $P(N_t = n) = P(\{N_t \geq n\} - \{N_t \geq n+1\})$.

c) folgt über $M(t) = E(N_t) = \sum_{n=0}^{\infty} nP(N_t = n) = \sum_{n=1}^{\infty} P(N_t \geq n)$ aus (2).

d) Aus der Faltungsdefinition folgt

$$F^{(m+n)*}(t) = \int_{[0,t]} F^{m*}(t-x)F^{n*}(dx) \leq \int_{[0,t]} F^{m*}(t)F^{n*}(dx) = F^{m*}(t)F^{n*}(t)$$

und $F^{n*}(t)$ ist nicht wachsend in $n \in \mathbb{N}$ für jedes $t \geq 0$. Für festes $t \in \mathbb{R}_+$ gibt es ein $h \in \mathbb{N}$ mit

$$P(X_2 + \dots + X_{h+1} > t) \geq P\left(X_i > \frac{t}{h} : i = 2, \dots, h+1\right) = \left(P\left(X_2 > \frac{t}{h}\right)\right)^h > 0.$$

(Sonst wäre $P(X_2 = 0) = F(0) = 1$.) Also ist $F^{h*}(t) < 1$ und $F^{mh*}(t) \leq (F^{h*}(t))^m$, $m \in \mathbb{N}$, konvergiert geometrisch. Mit

$$F^{mh*}(t) \geq F^{(mh+k)*}(t) \geq F^{(m+1)h*}(t), \quad k = 1, 2, \dots, h-1,$$

folgt dann die Konvergenz von $M(t)$. ■

Erneuerungstheorie kann auch als die Theorie regenerativer Phänomene oder rekurrenter Ereignisse angesehen werden. Wir haben derartige Zusammenhänge bei der Untersuchung Markovscher Ketten ausgenutzt. Der folgende Satz zeigt, daß essentiell nicht die Markov-Eigenschaft, sondern die daraus folgende regenerative Struktur des Prozesses ist.

5.7 Definition

$Y = (Y_t : t \geq 0)$ sei ein stochastischer Prozeß mit diskretem Zustandsraum $(E, \mathcal{P}(E))$, $Y_t : (\Omega, \mathcal{F}, P) \rightarrow (E, \mathcal{P}(E))$, $t \geq 0$.

Y heißt regenerativer Prozeß, falls es eine zufällige Zeit $X_1 : (\Omega, \mathcal{F}, P) \rightarrow (\mathbb{R}_+, \mathbb{B}_+)$ gibt mit den Eigenschaften:

- i) $P(X_1 \in \mathbb{R}_+) = 1$, $P(X_1 = 0) = 0$;
- ii) Der Post- X_1 -Prozeß $(Y(X_1 + t) : t \geq 0)$ ist stochastisch unabhängig von $(Y(s) : s \leq X_1)$;
- iii) Der Post- X_1 -Prozeß $(Y(X_1 + t) : t \geq 0)$ besitzt die gleiche Verteilung wie $(Y(s) : s \geq 0)$.

Für den regenerativen Prozeß Y wird dann X_1 als (erster) Regenerationszeitpunkt bezeichnet.

5.8 Korollar

Mit den Annahmen und Bezeichnungen aus Definition 5.7 folgt aus der Existenz einer unendlichen Folge S_1, S_2, \dots von Regenerationzeiten, für die gilt:

$$S_1 = X_1, \quad S_n - S_{n-1} =: X_n \sim X_1, \quad n \geq 2.$$

Die X_1, X_2, \dots sind unabhängig identisch verteilt, so daß mit $S_0 \equiv 0$ die Folge $(S_n : n \geq 0)$ ein (eingebetteter) Erneuerungsprozeß ist.

5.9 Definition

Sei $(S_n : n \in \mathbb{N})$ der gemäß Korollar 5.8 in den regenerativen Prozeß $(Y_t : t \geq 0)$ eingebettete Erneuerungsprozeß. Dann werden die Prozesse $(Y_t : S_n \leq t < S_{n+1})$, $n \geq 0$, als Zyklen bezeichnet.

5.10 Satz

Gegeben sei ein regenerativer Prozeß gemäß Definition 5.7. Die Verteilung der Regenerationszeit sei nichtarithmetisch. Die Pfade von Y seien rechtsstetig und es gelte $E(X_1) = \mu < \infty$. Dann existieren die Limiten

$$\lim_{t \rightarrow \infty} P(Y_{t+s} = k \mid X_1 = s) = \pi_k, \quad k \in E,$$

und definieren ein Wahrscheinlichkeitsmaß auf E , d. h. es ist $\pi_k \geq 0$ für $k \in E$ und $\sum_{k \in E} \pi_k = 1$.

Beweis: Es sei $F(\bullet)$ die Verteilungsfunktion von X_1 . Dann gilt:

$$\begin{aligned} P(Y_t = k) &= P(Y_t = k, X_1 > t) + P(Y_t = k, X_1 \leq t) \\ &= P(Y_t = k, X_1 > t) + \int_{x \in [0, t]} dP(X_1 \leq x) P(Y_t = k \mid X_1 = x). \end{aligned}$$

Die Regenerationseigenschaft vereinfacht dies zu

$$P(Y_t = k) = P(Y_t = k, X_1 > t) + \int_{x \in [0, t]} P(Y_{t-x} = k) F(dx), \quad k \in E, \quad t \geq 0,$$

und diese Gleichungen haben die Struktur

$$Z(t) = z(t) + \int_{[0, t]} Z(t-x) F(dx), \quad t \geq 0,$$

die als Erneuerungsgleichung bezeichnet wird. Dabei ist die Funktion

$$z(t) := P(Y_t = k, X_1 > t)$$

(bei festem $k \in E$) „direkt Riemann-integrierbar“. Damit ist über eine Version des sogenannten „Erneuerungstheorems“ bekannt, daß für die Lösung $(Z(t) : t \geq 0)$ der Erneuerungsgleichungen gilt:

$$Z(t) \xrightarrow{t \rightarrow \infty} \mu^{-1} \int_{\mathbb{R}_+} z(s) ds,$$

wobei $\mu = E(X_1)$ ist. Also haben wir

$$P(Y_t = k) \xrightarrow{t \rightarrow \infty} \mu^{-1} \int_0^\infty P(Y_x = k, X_1 > x) dx =: \pi_k \geq 0.$$

Weiterhin gilt:

$$\sum_{k \in E} \pi_k = \mu^{-1} \int_0^\infty \sum_{k \in E} P(Y_x = k \wedge X_1 > x) dx = \mu^{-1} \int_0^\infty (1 - F(x)) dx = 1. \quad \blacksquare$$

Satz 5.10 fordert, daß der regenerative Prozeß Y im „gerade erneuerten Zustand“ gestartet wird. Für die Existenz der Grenzwahrscheinlichkeiten ist dies unwesentlich. Analog zur Definition 5.4 (gemäß Beispiel 5.3 c)) eines verzögerten Erneuerungsprozesses können wir auch einen verzögerten regenerativen Prozeß definieren, wobei Definition 5.7 dann abgeschwächt wird, so daß die Zeit (bis zur ersten „Regeneration“ eine andere Verteilung hat als die eigentliche Regenerationszeitverteilung. Dann ist der Prozeß

$$\tilde{Y} := (Y_t : t \geq \tilde{X}) = (Y_{s+\tilde{X}} : s \geq 0)$$

ein regenerativer Prozeß gemäß Definition 5.7. (Für genauere Ausführungen siehe [Wol89], S. 89.)

Die im Beweis von Satz 5.10 verwendete Methode wird im allgemeinen kurz als „Erneuerungsargument“ bezeichnet. Die daraus entstehende Aufgabe ist: Löse die Erneuerungsgleichung und untersuche die Eigenschaften der Lösung für möglichst allgemeine Funktionen $z(t)$, $t \geq 0$. Die Darstellung folgt [Koh77], [Fel68] und [Fel71]; [Wol89] und [Asm87] enthalten detailliertere weitere Informationen und zum Teil modernere Darstellungen.

Um die Schreibweise zu vereinfachen, führen wir eine Verallgemeinerung der Faltungsoperation ein:

5.11 Definition

Sei $F : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ eine maßdefinierende Funktion mit $F(t) = 0$ für alle $t \leq 0$ sowie $g : \mathbb{R} \rightarrow \mathbb{R}$ eine meßbare Funktion, die auf $(-\infty, 0)$ verschwindet. Dann ist die Faltung $g * F : \mathbb{R} \rightarrow \overline{\mathbb{R}}$ definiert als:

$$(g * F)(y) = \int_{[0, y]} g(y-x) F(dx) \quad \text{für } y \geq 0 \quad \text{und} \quad (g * F)(y) = 0 \quad \text{für } y < 0.$$

5.12 Lemma

Seien F , F_1 und F_2 maßdefinierende Funktionen sowie g , g_1 und g_2 meßbare Funktionen, die den Bedingungen aus Definition 5.11 genügen. Dann gilt:

$$(g_1 + g_2) * F = g_1 * F + g_2 * F, \quad g * (F_1 + F_2) = g * F_1 + g * F_2.$$

5.13 Satz

Sei F eine (möglicherweise defekte) Verteilungsfunktion auf $[0, \infty)$ mit $F(0) = 0$ und $z : \mathbb{R} \rightarrow \mathbb{R}$ eine auf endlichen Intervallen beschränkte, meßbare, auf $(-\infty, 0)$ verschwindende Funktion. Dann hat die Erneuerungsgleichung

$$Z(t) = z(t) + \int_{[0,t]} Z(t-x)F(dx), \quad t \geq 0, \quad (1)$$

die Lösung

$$Z(t) = z(t) + \int_{[0,t]} z(t-x)U(dx), \quad t \geq 0, \quad (2)$$

wobei

$$U(t) = \sum_{k=1}^{\infty} F^{k*}(t), \quad t \in \mathbb{R}_+,$$

die Erneuerungsfunktion von F ist. Außerdem ist (2) die einzige Lösung von (1), die auf jedem endlichen Intervall beschränkt ist und auf $(-\infty, 0)$ verschwindet.

Beweis: (2) kann als

$$Z = z + z * U = z + \left(z * F + \sum_{k=2}^{\infty} z * F^{k*} \right) = z + \left(z * F + \left(\sum_{k=1}^{\infty} z * F^{k*} \right) * F \right) = z + (z + z * U) * F$$

geschrieben werden, wobei die Assoziativ- und Distributivgesetze für Faltungen von Maßen bzw. nach Lemma 5.12 verwendet wurden und der Satz von der monotonen Konvergenz. (2) löst also (1), verschwindet auf $(-\infty, 0)$, und es gilt:

$$\sup_{0 \leq t \leq s} |Z(t)| \leq \sup_{0 \leq t \leq s} |z(t)| + \int_{[0,s]} \sup_{0 \leq t \leq s} |z(t)| U(dx) = \sup_{0 \leq t \leq s} |z(t)| (1 + U(s)) < \infty$$

nach Lemma 5.6 d) und Voraussetzung für alle $s \geq 0$.

Gäbe es eine weitere Lösung Z' mit den angegebenen Eigenschaften, so hätte die Differenz $V = Z - Z'$ die gleichen Endlichkeitseigenschaften und es würde gelten:

$$V * F = Z * F - Z' * F = (z + Z * F) - (z + Z' * F) = Z - Z' = V$$

und entsprechend

$$V * F^{k*} = V \quad \text{für alle } k \geq 1.$$

Das impliziert

$$|V(t)| = \left| \int_{[0,t]} V(t-x)F^{k*}(dx) \right| \leq \sup_{0 \leq s \leq t} |V(s)| F^{k*}(t)$$

und es gilt

$$F^{k*}(t) \xrightarrow{k \rightarrow \infty} 0 \quad \text{für alle } t,$$

d. h. $V(t) = 0$ für alle $t \in \mathbb{R}$. ■

5.14 Anmerkung

Zählt man im (verzögerten oder unverzögerten) Erneuerungsprozeß auch den Zeitpunkt $S_0 \equiv 0$ als Erneuerungs Augenblick, so ist $\tilde{N}(t)$ zu definieren als zugehöriger Zählprozeß

$$\tilde{N}(t) = 1 + \sum_{k=1}^{\infty} 1_{(S_k \leq t)}, \quad t \geq 0,$$

mit Erneuerungsfunktion

$$\widetilde{M}(t) = 1 + \sum_{k=1}^{\infty} F^{k*}(t) = \sum_{k=0}^{\infty} F^{k*}(t), \quad t \geq 0.$$

Letzteres ist eine Definition, die auch für eine Fortsetzung auf \mathbb{R} geeignet ist. Diese Konvention wird bei Feller ([Fel71]) eingeführt und liefert als Lösung der Erneuerungsgleichung

$$Z(t) = \int_{[0,t]} z(t-y)\widetilde{M}(dy), \quad t \in \mathbb{R},$$

was gerade (2) in Satz 5.13 ist. □

5.15 Korollar

Für die Erneuerungsfunktion im (verzögerten) Erneuerungsprozeß gilt: $M(t)$, $t \geq 0$, erfüllt die Erneuerungsgleichung

$$M(t) = G(t) + \int_{[0,t]} M(t-x)F(dx), \quad t \geq 0.$$

Beweis: Einsetzen von $M(t) = \sum_{k=1}^{\infty} G * F^{(k-1)*}(t)$. ■

5.16 Beispiel (Alterserneuerungspolitiken)

Eine der einfachsten Strategien, die bei Ausfall eines Elementes entstehenden — als hoch angenommenen — Kosten zu senken, besteht darin, nach Erreichen eines bestimmten Alters das Element auf jeden Fall zu ersetzen. Durch den möglichen Planungsvorlauf entstehen dann geringere Kosten als beim zufälligen überraschenden Ausfall.

Bezeichnen wir mit $T > 0$ den fixierten Zeitpunkt der präventiven Erneuerung, so ist das Ziel der im folgenden skizzierten mathematischen Modellierung des Problems die Konstruktion eines optimalen T , welches dann eine „optimale Alterserneuerungspolitik“ beschreibt. Dazu müssen geeignete Kostenfunktionen und Optimalitätsbegriffe festgelegt werden. Es seien dazu gegeben

- $c_1 \hat{=}$ Ersatzkosten, falls das Element vor T ausfällt,
- $c_2 \hat{=}$ Ersatzkosten, falls das Element zur Zeit T präventiv ersetzt wird,
- $0 \leq \varrho \hat{=}$ Diskontierungsrate der anfallenden Kosten, d. h. Kosten der Höhe C , die zur Zeit t anfallen, sind zur Zeit 0 mit $C \cdot \exp(-\varrho t)$ anzusetzen,
- $c(t) \hat{=}$ Erwartungswert der auf den Zeitpunkt 0 diskontierten Kosten, die in $[0, t]$ entstehen.

Wir versuchen im folgenden, $c(t)$, $t \geq 0$, als Funktion der (als konstant angenommenen) Kosten und von T darzustellen, und hoffen danach, in T die erhaltenen Kosten $c(t)$ für alle $t \geq 0$ zu minimieren.

a) Herleitung einer Erneuerungsgleichung für $c(\bullet)$:

Sei F die Lebensdauervertiefungsfunktion der verwendeten Elemente, einschließlich des ersten. Sei $T > 0$ der Zeitpunkt präventiver Erneuerungen, falls das Element das Alter T erreicht. Die tatsächliche Lebensdauervertiefungsfunktion (Arbeitsdauervertiefungsfunktion) der Elemente ist dann $\tilde{F}(\bullet)$, gegeben durch

$$\tilde{F}(x) = \begin{cases} F(x) & \text{für } x < T, \\ 1 & \text{für } x \geq T. \end{cases}$$

Es gilt dann für $t \geq 0$:

i) Findet zur Zeit $x \leq t$ die erste Erneuerung statt, so ist der Wert dieser Kosten zur Zeit 0 gerade

$$L(x) := \begin{cases} c_1 \cdot e^{-\varrho x} & \text{falls } x < T, \\ c_2 \cdot e^{-\varrho x} & \text{falls } x = T. \end{cases}$$

- ii) Die weiteren, nach x bis t entstehenden Kosten diskontieren wir auf den Zeitpunkt x . Da nach der Erneuerung die stochastische Entwicklung des betrachteten Prozesses identisch der des Gesamtprozesses ist, sind dies Kosten mit dem Wert $c(t-x)$ zur Zeit x . Diskontieren wir auf den Zeitpunkt 0, so ergibt sich für die weiteren Kosten in $(x, t]$: $c(t-x) \cdot e^{-\rho x}$.
- iii) Die gesamten in $[0, t]$ entstehenden Kosten, diskontiert auf den Zeitpunkt 0, sind also

$$c(t) = \int_{[0,t]} L(x) \tilde{F}(dx) + \int_{[0,t]} c(t-x) e^{-\rho x} \tilde{F}(dx), \quad t \geq 0. \quad (1)$$

Ist $G(\bullet)$ die maßdefinierende Funktion, die über die \tilde{F} -Dichte $e^{-\rho x}$ festgelegt ist, so kann mit

$$z(t) := \int_{[0,t]} L(x) \tilde{F}(dx)$$

die Gleichung (1) als Erneuerungsgleichung

$$c(t) = z(t) + \int_{[0,t]} c(t-x) G(dx), \quad t \geq 0, \quad (1')$$

geschrieben werden, deren Lösung in Satz 5.13 (2) angegeben ist.

- b) Um die von T abhängige Lösung von (1') in T zu minimieren, muß man also im allgemeinen die Erneuerungsfunktion $U(\bullet)$ explizit angeben. Da dies in vielen Fällen nicht möglich ist, sucht man nach Näherungen vor allem für große Werte t in der Form, daß

$$\lim_{t \rightarrow \infty} (U(t+h) - U(t))$$

untersucht wird, also die mittlere Anzahl der Erneuerungen in $(t, t+h]$. □

Wir kommen auf das Beispiel wieder zurück, werden aber zunächst weitere technische Hilfsmittel bereitstellen müssen.

5.17 Definition

Ein Erneuerungsprozeß heißt *transient*, wenn für die Erneuerungsverteilungsfunktion F gilt:

$$\lim_{t \rightarrow \infty} F(t) =: F(\infty) < 1.$$

Ansonsten heißt der Prozeß *rekurrent*.

5.18 Beispiel

Sei $X = (X_n : n \in \mathbb{N})$ eine homogene Markov-Kette mit diskretem Zustandsraum E . X werde mit Wahrscheinlichkeit 1 im Zustand $i \in E$ gestartet. Sei $\varrho_n(i)$, $n \in \mathbb{N}_+$, die Folge der Zwischeneintrittszeiten („Rückkehrzeiten“) zwischen Eintritten von X in i und $\tau_0(i) \equiv 0$, $\tau_n(i)$, $n \in \mathbb{N}_+$, die Folge der Eintrittszeiten von X in i .

Nach Lemma 3.32 ist $\tau(i) = (\tau_n(i) : n \in \mathbb{N})$ ein Erneuerungsprozeß mit Lebensdauern $(\varrho_n(i) : n \in \mathbb{N}_+)$. $\tau(i)$ ist nach Definition 3.34 genau dann transient, falls i ein transienter Zustand von X ist; $\tau(i)$ ist genau dann rekurrent, falls i rekurrenter Zustand von X ist. □

5.19 Satz

Gegeben sei ein transienter, gewöhnlicher Erneuerungsprozeß. Dann finden mit Wahrscheinlichkeit 1 nur endlich viele Erneuerungen statt und die erwartete Anzahl der Erneuerungen $U(t)$ in $[0, t]$ konvergiert gegen

$$U(\infty) := \frac{F(\infty)}{1 - F(\infty)} < \infty.$$

Beweis:

- i) Mit Wahrscheinlichkeit $F(\infty)^r (1 - F(\infty))$ finden genau $r = 0, 1, 2, \dots$ Erneuerungen statt. Die Anzahl der Erneuerungen ist also verteilt gemäß einer (nicht defekten!) auf \mathbb{N} konzentrierten, geometrischen Verteilung mit „Erfolgswahrscheinlichkeit“ $1 - F(\infty)$ und Erwartungswert $F(\infty)(1 - F(\infty))^{-1}$.

- ii) Wegen $\{N_t \geq k\} = \{S_k \leq t\}$ ist $F^{k*}(\infty) := \lim_{t \rightarrow \infty} F^{k*}(t)$ die Wahrscheinlichkeit, daß insgesamt mindestens k Erneuerungen stattfinden. Weiter gilt $F^{k*}(\infty) = (F(\infty))^k$ und damit

$$U(t) = \sum_{k=1}^{\infty} F^{k*}(t) \longrightarrow \frac{F(\infty)}{1 - F(\infty)} < \infty. \quad \blacksquare$$

5.20 Korollar

- a) Gegeben sei ein transienter, unverzögerter Erneuerungsprozeß mit Lebensdauerverteilungsfunktion $F(t)$, $t \geq 0$, d. h. es gilt $F(\infty) < 1$. Es sei $z : (\mathbb{R}_+, \mathbb{B}_+) \rightarrow (\mathbb{R}_+, \mathbb{B}_+)$ eine beschränkte, meßbare Funktion, deren Grenzwert $z(\infty) := \lim_{t \rightarrow \infty} z(t)$ existiert. Dann gilt:

Die Lösung $Z(\bullet)$ der Erneuerungsgleichung

$$Z(t) = z(t) + \int_{[0,t]} Z(t-x)F(dx), \quad t \geq 0,$$

konvergiert für $t \rightarrow \infty$, und es ist

$$\lim_{t \rightarrow \infty} Z(t) = \frac{z(\infty)}{1 - F(\infty)} < \infty.$$

- b) Für $t \rightarrow \infty$ konvergiert die Erneuerungsfunktion in einem verzögerten, transienten Erneuerungsprozeß und es gilt

$$\lim_{t \rightarrow \infty} M(t) = \frac{G(\infty)}{1 - F(\infty)} < \infty.$$

Beweis:

- a) Nach Satz 5.13, Gleichung (2), haben wir als Lösung der Erneuerungsgleichung

$$Z(t) = z(t) + \int_{[0,t]} z(t-x)U(dx), \quad t \geq 0,$$

zu untersuchen für $t \rightarrow \infty$. Wir zerlegen das Integral:

$$\int_{[0,t]} z(t-x)U(dx) = \int_{[0,t']} z(t-x)U(dx) + \int_{(t',t]} z(t-x)U(dx), \quad t' \leq t,$$

und wählen t, t' geeignet. Nach Voraussetzung gibt es zu $\varepsilon > 0$ ein $t' \in \mathbb{R}_+$, so daß $U(\infty) - U(t') < \varepsilon$ gilt, und ein $t > t'$, so daß für alle $x \in [0, t']$ gilt: $|z(\infty) - z(t-x)| < \varepsilon$. Es folgt

$$\begin{aligned} & \left| \int_{[0,t]} z(t-x)U(dx) - \frac{z(\infty)F(\infty)}{1 - F(\infty)} \right| \\ &= \left| \int_{(t',t]} (z(t-x) - z(\infty))U(dx) + \int_{[0,t']} (z(t-x) - z(\infty))U(dx) \right. \\ & \quad \left. + \int_{[0,t]} z(\infty)U(dx) - \frac{z(\infty)F(\infty)}{1 - F(\infty)} \right| \\ &\leq \int_{(t',\infty]} |z(t-x) - z(\infty)|U(dx) + \int_{[0,t']} |z(t-x) - z(\infty)|U(dx) \\ & \quad + z(\infty) \left| U(t) - \frac{F(\infty)}{1 - F(\infty)} \right| \\ &\leq 2 \sup_{s \geq 0} z(s) \cdot \varepsilon + \varepsilon U(\infty) + z(\infty) \cdot \varepsilon \\ &\xrightarrow{t \rightarrow \infty} 0. \end{aligned}$$

Es folgt die Behauptung, da *varepsilon* beliebig klein gewählt werden kann.

b) Nach Korollar 5.15 erfüllt die Erneuerungsfunktion $M(t)$, $t \geq 0$, eine Erneuerungsgleichung mit

$$z(t) = G(t) \xrightarrow{t \rightarrow \infty} G(\infty). \quad \blacksquare$$

Es sei angemerkt, daß wir stets $G(\infty) = 1$ ansetzen; Korollar 5.20 b) gilt auch ohne diese Annahme und ist auch dann sehr anschaulich zu interpretieren.

5.21 Beispiel (Alterserneuerungspolitiken bei positiver Diskontierung: Fortsetzung von Beispiel 5.16)

Die Erneuerungsgleichung für die diskontierten Kosten für die (zufälligen oder geplanten) Ersetzungen war

$$c(t) = z(t) + \int_{[0,t]} c(t-x)G(dx), \quad t \geq 0,$$

wobei

$$G(x) = \int_{[0,x]} e^{-ey} \bar{F}(dy), \quad x \geq 0,$$

ist. $G(\bullet)$ ist also eine defekte Verteilungsfunktion mit $G(\infty) < 1$. Außerdem existiert der Grenzwert

$$\lim_{t \rightarrow \infty} z(t) = \lim_{\substack{t \rightarrow \infty \\ T \leq t < \infty}} \left(c_1 \int_{[0,T]} e^{-ex} F(dx) + c_2(1 - F(T-)) \right) = z(T).$$

Nach Korollar 5.20 a) gilt also

$$c(\infty) := \lim_{t \rightarrow \infty} c(t) = \frac{z(T)}{1 - G(T)},$$

denn es ist $G(\infty) = G(T)$. Für eine vorgegebene Lebensdauerverteilungsfunktion $F(\bullet)$ haben wir also zur Lösung unseres Optimierungsproblems ein T_0 zu bestimmen, für das gilt:

$$\frac{z(T_0)}{1 - G(T_0)} = \inf_{T \in \mathbb{R}_+} \frac{z(T)}{1 - G(T)}.$$

Dabei berücksichtigen wir, daß es möglich sei kann, optimal dadurch zu handeln, daß als Politik gewählt wird: „Keine vorzeitigen Erneuerungen“, d. h. $T_0 = \infty$. Dies ist sicher der Fall, wenn $c_2 \geq c_1$ gilt. \square

Die Untersuchung des asymptotischen Verhaltens von Lösungen der Erneuerungsgleichung bzw. von Erneuerungsfunktionen in Satz 5.19 und Korollar 5.20 geschah für arithmetische und nicht-arithmetische Lebensdauerverteilungen gleichzeitig, sofern der Erneuerungsprozeß transient ist. Die dort einfach erhaltenen Ergebnisse können, wie in Beispiel 5.21 gezeigt, durch geeignete Modifikationen der Prozesse auch zur Untersuchung rekurrenter Erneuerungsprozesse verwendet werden. Das dort durchgeführte Diskontieren ist ein typisches Vorgehen.

Im rekurrenten Fall ist keine Möglichkeit bekannt, arithmetische und nicht-arithmetische Lebensdauerverteilungen gemeinsam zu behandeln. Dabei können wir uns für Sätze über arithmetische Lebensdauerverteilungsfunktionen auf Verteilungsfunktionen, die auf \mathbb{N} konzentriert sind, beschränken. (Dies wird in der Regel getan.)

5.22 Definition

Gegeben sei ein verzögerter Erneuerungsprozeß $S = (S_k : k \in \mathbb{N})$ mit Lebensdauern $(X_i : i \in \mathbb{N}_+)$, deren Verteilungen auf \mathbb{N} konzentriert sind. Es seien

$$g(\infty) = g = P(X_1 \in \mathbb{N}) \leq 1 \quad \text{und} \quad \mu = \mathbb{E} X_2 = \sum_{k \in \mathbb{N}} k \cdot P(X_2 = k) \leq \infty.$$

Der Erneuerungsprozeß heißt *periodisch*, wenn es eine ganze Zahl $d > 1$ gibt derart, daß

$$P(X_2 = k) = 0, \quad P(X_1 = k) = 0$$

gilt, falls $k \notin d \cdot \mathbb{N}$ ist. Die größte ganze Zahl d , welche diese Eigenschaft besitzt, heißt Periode von S . Ein nichtperiodischer Prozeß heißt aperiodisch (mit „Periode 1“).

Weiter sei die Folge von Zufallsvariablen $Z = (Z_k : k \in \mathbb{N}_+)$ definiert als Abbildung

$$Z_k : (\Omega, \mathcal{F}, P) \rightarrow (\{0, 1\}, \mathcal{P}\{0, 1\}), \quad k \in \mathbb{N}_+,$$

mit

$$Z_k = 1 \iff N(k-) < N(k),$$

d. h. zur Zeit k findet eine Erneuerung statt.

5.23 Satz (Erneuerungstheorem von Erdős, Feller, Pollard)³

In einem rekurrenten, verzögerten Erneuerungsprozeß mit Periode $d \geq 1$ gilt

$$\lim_{k \rightarrow \infty} P(Z_{k \cdot d} = 1) \rightarrow \frac{g \cdot d}{\mu}, \quad \text{und} \quad P(Z_l = 1) = 0 \quad \text{für } l \notin d \cdot \mathbb{N}.$$

Beweis: Für den Fall $g = 1$ kann das Ergebnis direkt aus Satz 4.16 b), Gleichung (3), abgelesen werden. Falls $g < 1$ ist, wird auf X_1 bedingt.

Formal muß eine geeignete Markov-Kette definiert werden, bei der die jeweilige Erneuerung aus dem aktuellen Zustand abgelesen werden kann. Die mittlere Rückkehrzeit in diesen ausgezeichneten Zustand (z. B. 0) ist dann in der Terminologie von Kapitel 3 gerade $m_{00} = \mu$. Die geeignete Markov-Kette erhalten wir durch eine Modifikation aus Beispiel 3.10: Wir wählen eine „linksstetige“ Version des Restlebensdauerprozesses. Nimmt dieser Wert 0 an, findet eine Erneuerung statt. ■

5.24 Anmerkung

Für $n \in \mathbb{N}$ besagt der Erneuerungssatz für diskrete Verteilungen bei rekurrenten, aperiodischen Prozessen, daß

$$m_n := P(Z_n = 1) = E(Z_n) = M(n) - M(n-1) \xrightarrow{n \rightarrow \infty} \frac{1}{\mu}$$

gilt, was für $h \geq 1$, $h \in \mathbb{N}$,

$$M(n+h) - M(n) = \sum_{i=1}^h m_{n+i} \xrightarrow{n \rightarrow \infty} \frac{h}{\mu}$$

impliziert. Daraus folgt im allgemeinen nicht

$$M(t+h) - M(t) \xrightarrow{t \rightarrow \infty} \frac{h}{\mu}, \quad t, h \in \mathbb{R}_+. \quad (*)$$

Dies zeigt noch einmal, daß der Fall arithmetischer Verteilungen in der Erneuerungstheorie ein Sonderfall ist, denn eben die Aussage (*) wird das Erneuerungstheorem für nicht-arithmetische Verteilungen werden. □

Unser Ziel ist, den Erneuerungssatz entsprechend Satz 5.23 für nicht-arithmetische Verteilungen zu formulieren und zu beweisen. Wir benötigen eine Reihe von vorbereitenden Aussagen, deren erste noch für allgemeine Lebensdauerverteilungsfunktion gilt:

5.25 Lemma

Die Erneuerungsfunktion $M(\bullet)$ erfüllt

$$M(t+h) - M(t) \leq 1 + M(h) \quad \text{für } t > 0, h > 0,$$

d. h. ihre Zuwächse sind gleichmäßig beschränkt für jede feste Intervalllänge.

³Siehe [Fel68], S. 312–313, oder [Koh77], S. 51.

Beweis (für den Fall des unverzögerten Erneuerungsprozesses):

$$\begin{aligned} M(t+h) - M(t) &= \mathbb{E}(N_{t+h} - N_t) = \sum_{n=1}^{\infty} P(N_{t+h} - N_t \geq n) = \sum_{n=0}^{\infty} P(N_{t+h} - N_t > n) \\ &\stackrel{(*)_1}{\leq} \sum_{n=0}^{\infty} P(S_n \leq h) = 1 + M(h), \quad h, t \geq 0. \end{aligned}$$

Dabei folgt die Ungleichung $(*_1)$ summandenweise:

$$P(N_{t+h} - N_t > n) \leq P(S_n \leq h)$$

mit den folgenden Umformungen:

$$\begin{aligned} P(S_n > h) &= P\left(\sum_{i=2}^{n+1} X_i > h\right) \stackrel{(*)_2}{=} P\left(\sum_{i=N_t+2}^{N_t+n+1} X_i > h\right) \\ &= P\left(\sum_{i=1}^{N_t+n+1} X_i > t+h\right) \quad (\text{wegen } S_{N_t+1} > t) \\ &\stackrel{(*)_3}{=} P(S_{N_t+n+1} > t+h) = P(N_{t+h} < N_t + n + 1) \\ &\leq P(N_{t+h} - N_t < n + 1) = P(N_{t+h} - N_t \leq n), \end{aligned}$$

wobei an der Stelle $(*_2)$ benutzt wurde, daß die X_i unabhängig identisch verteilt sind:

$$P\left(\sum_{i=N_t+2}^{N_t+n+1} X_i > h\right) = \sum_{k=0}^{\infty} P\left(\sum_{i=N_t+2}^{N_t+n+1} X_i > h \mid \{N_t = k\}\right) P(N_t = k), \quad \text{und} \quad \{N_t = k\} = \{S_k \leq t < S_{k+1}\},$$

und bei $(*_3)$ Gleichung (2) aus dem Beweis von Lemma 5.6 zum Einsatz kam. ■

5.26 Lemma

Sei $z : \mathbb{R} \rightarrow \mathbb{R}$ stetig und beschränkt. Gilt dann

$$z(x) = \int_{[0, \infty)} z(x-s)F(ds), \quad x \in \mathbb{R},$$

so ist $z(\bullet)$ eine Konstante.

(Beweisskizze siehe [Fel71], S. 364, 382; der Satz gilt auch auf allgemeinen lokalkompakten Gruppen und ist als das Choquet–Deny–Theorem bekannt.)

Es gibt zwei — anscheinend völlig unterschiedliche — Arten, die Aussagen des „Erneuerungssatzes“ zu formulieren. Die erste Version haben wir im diskreten Fall in Satz 5.23 formuliert. Im nicht–arithmetischen Fall haben wir

5.27 Satz (Erneuerungstheorem, 1. Version)

Sei $U(t)$ die Erneuerungsfunktion eines rekurrenten, unverzögerten Erneuerungsprozesses mit nicht–arithmetischer Verteilungsfunktion F und $\mathbb{E}(F) = \mu \leq \infty$. Dann gilt für $h > 0$:

$$U(t+h) - U(t) \xrightarrow{t \rightarrow \infty} \frac{h}{\mu}.$$

Beweis: $g : \mathbb{R} \rightarrow \mathbb{R}$ sei stetig, beschränkt und verschwinde außerhalb von $[0, h]$. Es sei

$$\Phi(t) = g(t) + \int_{[0, t]} g(t-s)U(ds) = g(t) + \int_{[t-h, t]} g(t-s)U(ds), \quad t > 0.$$

Durch direkte Abschätzung folgt aus Lemma 5.25, daß $\Phi(\bullet)$ beschränkt und gleichmäßig stetig ist. Außerdem ist $\Phi(\bullet)$ die Lösung der Erneuerungsgleichung

$$\Phi(s) = g(s) + \int_{[0,s]} \Phi(s-x)F(dx), \quad s \geq 0.$$

Diese Gleichung integrieren wir über $[0, t]$, $t \geq h$:

$$\int_{[0,t]} g(s) ds = \int_{[0,t]} \Phi(s) ds - \int_{[0,t]} ds \int_{[0,s]} F(dx)\Phi(s-x)$$

(Satz von Fubini)

$$= \int_{[0,t]} \Phi(s) ds - \int_{[0,t]} F(dx) \int_{[x,t]} \Phi(s-x) ds$$

(zwei Substitutionen)

$$= \int_{[0,t]} \Phi(t-s) ds - \int_{(0,t]} F(dx) \int_{(0,t-x]} \Phi(s) ds$$

(Partielle Integration)⁴

$$\begin{aligned} &= \int_{[0,t]} \Phi(t-s) ds - \left(F(t) \cdot \int_{(0,t-t]} \Phi(s) ds - F(0) \cdot \int_{(0,t]} \Phi(s) ds \right. \\ &\quad \left. - (-1) \int_{(0,t]} \Phi(t-x)F(x) dx \right) \\ &= \int_{[0,t]} \Phi(t-s)(1-F(s)) ds. \end{aligned}$$

Da g außerhalb von $[0, h]$ verschwindet, haben wir damit

$$\int_{[0,\infty)} g(s) ds = \int_{[0,h]} g(s) ds = \int_{[0,t]} \Phi(t-s)(1-F(s)) ds. \quad (1)$$

Damit haben wir die Voraussetzung geschaffen, um das asymptotische Verhalten von $\Phi(\bullet)$ zu untersuchen. Für große t ($> h$) gilt aber:

$$\Phi(t) = \int_{[t-h,t]} g(t-s)U(ds).$$

Ist jetzt $g = 1_{[a,b]}$, so erhalten wir bei geeigneter Wahl von a, b und der Interpretation von $U(\bullet)$ als maßdefinierender Funktion gerade $U(t, t+h]$ und dessen asymptotisches Verhalten. Wir haben dann noch Indikatorfunktionen durch stetige $g(\bullet)$, die außerhalb von $[0, h]$ verschwinden, zu approximieren. ([Koh77], S. 59, ausführlich.)

Sei also $\bar{m} = \limsup_{t \rightarrow \infty} \Phi(t)$ und $(t_n : n \in \mathbb{N})$ eine unbeschränkt wachsende Folge mit $\lim_{n \rightarrow \infty} \Phi(t_n) = \bar{m}$. Sei

$$z_n(x) = \begin{cases} \Phi(t_n + x) & \text{falls } -t_n < x < \infty, \\ 0 & \text{falls } x \leq t_n \text{ gilt.} \end{cases}$$

Die Menge von Funktionen $\{z_n : n \in \mathbb{N}\}$ ist gleichgradig stetig, d. h. zu jedem $\varepsilon > 0$ gibt es ein $\delta > 0$ derart, daß $|x' - x''| < \delta$ schon $|z_n(x') - z_n(x'')| < \varepsilon$ für jedes n impliziert, und gleichgradig beschränkt, d. h. es gibt ein $C \geq 0$ mit $|z_n(x)| < C$ für alle n und für alle x (folgt aus der gleichmäßigen Stetigkeit von Φ und der Beschränktheit von Φ).

⁴(siehe z. B.: [Fel71], S. 150; [Koh77], S. 31): Sei $u : (\mathbb{R}, \mathbb{B}) \rightarrow (\mathbb{R}, \mathbb{B})$ beschränkt und stetig differenzierbar. Dann gilt für $-\infty < a \leq b < \infty$:

$$\int_{(a,b]} u(x)F(dx) = F(b)u(b) - F(a)u(a) - \int_{(a,b]} u'(x)F(x) dx.$$

Für die Folge $(z_n : n \in \mathbb{N})$ gilt⁵: Es gibt eine Teilfolge $(z_{n_k} : k \in \mathbb{N}) \subseteq (z_n : n \in \mathbb{N})$, die gegen eine stetige Funktion $z(\bullet)$ konvergiert, und die Konvergenz ist gleichmäßig auf endlichen Intervallen (siehe dazu außerdem [Fel71], S. 270).

Da $(z_n : n \in \mathbb{N})$ sogar gleichmäßig gleichgradig stetig ist (siehe [Die71], S. 145, Aufgaben 5 und 6), ist z ebenfalls gleichmäßig stetig. Auswerten der Erneuerungsgleichung für $\Phi(t_n + x)$, $x > -t_n$, und Grenzübergang $n \rightarrow \infty$ zeigen, daß $z(\bullet)$ die Gleichung

$$z(x) = \int_{[0, \infty)} z(x-s)F(ds), \quad -\infty < x < \infty,$$

erfüllt ($g(t_n + x) = 0$ für n hinreichend groß; $\{z_n(\bullet)\}$ gleichgradig beschränkt \Rightarrow majorisierte Konvergenz).

Nach Lemma 5.26 ist z also eine Konstante:

$$z(x) = z(0) = \lim_{n \rightarrow \infty} z_n(0) = \lim_{n \rightarrow \infty} \Phi(t_n) = \bar{m}.$$

Aus (1) folgt damit

$$\int_{[0, h]} g(s) ds = \lim_{n \rightarrow \infty} \int_{[0, t_n]} \Phi(t_n - s)(1 - F(s)) ds = \lim_{n \rightarrow \infty} \int_{[0, \infty)} z_n(-s)(1 - F(s)) ds = \bar{m}\mu.$$

Analog zeigt man mit $\underline{m} = \liminf_{t \rightarrow \infty} \Phi(t)$:

$$\underline{m} = \frac{1}{\mu} \int_{[0, \infty)} g(s) ds = \bar{m} = \lim_{t \rightarrow \infty} \Phi(t) = \lim_{t \rightarrow \infty} \int_{[t-h, t]} g(t-s)U(ds)$$

für jede stetige Funktion $g(\bullet)$, die außerhalb von $[0, h]$ verschwindet.

Es bleibt die Approximation vom Indikator explizit durchzuführen. Dabei erhalten wir jeweils die Konvergenz

$$\frac{1}{\mu} \int_{[0, h]} g(s) ds \longrightarrow \frac{1}{\mu}(b-a). \quad \blacksquare$$

Neben der Behauptung über das asymptotische Verhalten der Erneuerungsfunktion $M(\bullet)$ haben wir im Beweis anscheinend eine weitaus stärkere Aussage gezeigt:

Ist $\Phi(\bullet)$ die Lösung der Erneuerungsgleichung

$$\Phi(t) = g(t) + \int_{[0, t]} \Phi(t-s)F(ds), \quad t \geq 0, \quad (*)$$

wobei g stetig und auf $[0, h]$ konzentriert ist, so gilt

$$\lim_{t \rightarrow \infty} \Phi(t) = \frac{1}{\mu} \int_{[0, \infty)} g(s) ds.$$

Die Aussage des Erneuerungstheorems ist dann die entsprechende Behauptung für Funktionen

$$g(s) = 1_{[t, t+h]}(s), \quad s \in \mathbb{R}.$$

Es stellt sich die Frage, für welche anderen Funktionen g die Lösung der Erneuerungsgleichung Φ das asymptotische Verhalten $(*)$ zeigt. Die Antwort wird in der zweiten Version des Erneuerungstheorems gegeben und führt auf die Klasse der direkt Riemann-integrierbaren Funktionen.

5.28 Definition

$z : \mathbb{R}_+ \rightarrow \mathbb{R}$ sei eine beliebige Funktion. Für $k = 1, 2, \dots$ und $h > 0$ seien

$$\underline{u}_k(h) = \inf_{(k-1)h \leq x < kh} z(x), \quad \bar{u}_k(h) = \sup_{(k-1)h \leq x < kh} z(x);$$

⁵siehe [GF68], S. 127; dort wird von „gleichartiger“ Stetigkeit gesprochen

die Reihen

$$\underline{\sigma}(h) = h \cdot \sum_{k=1}^{\infty} \underline{u}_k(h), \quad \bar{\sigma}(h) = h \cdot \sum_{k=1}^{\infty} \bar{u}_k(h) \quad (1)$$

heißen Unter- bzw. Obersumme zur Spanne h . Falls für alle $h > 0$ die Reihen in (1) absolut konvergieren und falls für $h \rightarrow 0$

$$-\infty < \lim_{h \downarrow 0} \underline{\sigma}(h) = \lim_{h \downarrow 0} \bar{\sigma}(h) < \infty \quad (2)$$

gilt, heißt z direkt Riemann-integrierbar und

$$\int_{[0, \infty)} z(t) dt = \lim_{h \downarrow 0} \underline{\sigma}(h) = \lim_{h \downarrow 0} \bar{\sigma}(h)$$

heißt direktes Riemann-Integral.

5.29 Anmerkung

- a) Das übliche (uneigentliche) Riemann-Integral wird definiert über

$$\lim_{a \uparrow \infty} \int_{[0, a]} z(t) dt =: \int_{[0, \infty)} z(t) dt,$$

falls dieser Grenzwert existiert. Dabei werden also Unter- und Obersummen nur über $[0, a]$, $a < \infty$, gebildet, so daß bei deren Berechnung noch kein Grenzwert gebildet werden muß wie bei der direkten Riemann-Integration.

- b) Ist z (direkt) Riemann-integrierbar über jedem endlichen Intervall $[0, a]$ und gibt es ein $h > 0$, so daß $\bar{\sigma}(h) < \infty$ und $\underline{\sigma}(h) > -\infty$ ist, so ist z direkt Riemann-integrierbar über $[0, \infty)$ ([Fel71], S. 362).
- c) Es gibt Riemann-integrierbare Funktionen, die nicht direkt Riemann-integrierbar sind ([Fel71], S. 363): $z : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ sei gegeben durch

i) $z(n) = a_n \in \mathbb{R}_+$ für $n = 1, 2, \dots$

- ii) Sei $0 < h_n < \frac{1}{2}$: auf $[n - h_n, n]$ und $[n, n + h_n]$ sei z linear und verschwinde außerhalb von

$$\bigcup_{n \in \mathbb{N}_+} (n - h_n, n + h_n).$$

z ist also genau dann Riemann-integrierbar, wenn $\sum_{n \in \mathbb{N}_+} h_n a_n < \infty$ gilt. Divergiert jetzt $a_n \uparrow \infty$ für $n \uparrow \infty$, so ist z nicht direkt Riemann-integrierbar.

- d) Analog kann eine Riemann-integrierbare Funktion $z : \mathbb{R}_+ \rightarrow \mathbb{R}$ definiert werden mit $\limsup_{n \uparrow \infty} a_n = \infty$ und $\liminf_{n \uparrow \infty} a_n = -\infty$, deren Schwankung also über jede Grenze geht.
- e) Aus direkter Riemann-Integrierbarkeit folgt Riemann-Integrierbarkeit und die Integrale sind identisch.
- f) Ein häufig auftretender Spezialfall: z sei auf \mathbb{R}_+ monoton, verschwinde auf $(-\infty, 0)$ und es gelte $\int_{[0, \infty)} |z(t)| dt < \infty$. Dann ist z direkt Riemann-integrierbar ([Koh77], S. 63).
- g) „ z ist direkt Riemann-integrierbar“ impliziert wegen (2) aus Definition 5.28, daß z beschränkt ist. (Sonst wären $|\underline{\sigma}(h)|$ oder $|\bar{\sigma}(h)|$ oder beide nicht endlich.) \square

Wir können jetzt (gemäß [Fel71], S. 363) formulieren, was vor Satz 5.27 angekündigt war:

5.30 Satz (Erneuerungstheorem, 2. Version)

Sei $z : (\mathbb{R}, \mathbb{B}) \rightarrow (\mathbb{R}, \mathbb{B})$ meßbar und direkt Riemann-integrierbar. $F(\bullet)$ sei eine nicht-arithmetische Verteilungsfunktion auf $(\mathbb{R}_+, \mathbb{B}_+)$. Dann gilt für die Lösung $Z(\bullet)$ der Erneuerungsgleichung:

$$Z(t) = z(t) + \int_{[0,t]} Z(t-y)F(dy), \quad t \geq 0, \quad \lim_{t \rightarrow \infty} Z(t) = \frac{1}{\mu} \int_{[0,\infty)} z(s) ds,$$

wobei $\mu \leq \infty$ der Mittelwert von F ist.

Beweis:

i) Sei $z = 1_{[a,b]}$; dann gilt

$$Z(t) = U(t-a) - U(t-b) \longrightarrow \frac{b-a}{\mu}$$

nach der ersten Version des Erneuerungstheorems.

ii) Ist z eine Treppenfunktion, die nur endlich viele Werte annimmt:

$$z(t) = \sum_{k=1}^s c_k 1_{[a_k, b_k)}(t), \quad b_k - a_k = h_k,$$

so folgt direkt aus i) schon

$$\lim_{t \rightarrow \infty} Z(t) = \frac{1}{\mu} \sum_{k=1}^s c_k h_k = \frac{1}{\mu} \int_{[0,\infty)} z(s) ds.$$

iii) Ist z eine Treppenfunktion, die abzählbar viele Werte annimmt, von der Form

$$z(t) = \sum_{k=1}^{\infty} c_k 1_{[(k-1)h, kh)}(t), \quad t \geq 0,$$

mit

$$\infty > \int_{[0,\infty)} |z(t)| dt = \sum_{k=1}^{\infty} |c_k| h, \quad (1)$$

so schließt man wie folgt:

Sei $Z_k(h; t)$ die Lösung der Erneuerungsgleichung für die Funktion $1_{[(k-1)h, kh)}(t)$. Dann gilt nach Lemma 5.25

$$Z_k(h; t) = U(t - (k-1)h) - U(t - kh) \leq 1 + U(h), \quad t \geq 0, \quad (2)$$

so daß die Reihe

$$Z(h; t) = \sum_{k=1}^{\infty} c_k Z_k(h; t), \quad t \geq 0,$$

gleichmäßig in t konvergiert. Aus (1) und (2) folgt, daß $Z(h; \bullet)$ die Lösung der Erneuerungsgleichung zu $z(h; \bullet)$ ist. Dann gilt

$$\lim_{t \rightarrow \infty} Z(h; t) = \lim_{t \rightarrow \infty} \sum_{k=1}^{\infty} c_k Z_k(h; t) = \sum_{k=1}^{\infty} c_k \frac{h}{\mu} = \frac{1}{\mu} \int_{[0,\infty)} z(s) ds.$$

Die Vertauschung von Summe und Grenzwert ist wegen der gleichmäßigen Konvergenz der Reihe möglich.

iv) Sei z beliebig und direkt Riemann-integrierbar. Für $h > 0$ sei

$$\underline{z}(h; t) = \sum_{k=1}^{\infty} \underline{u}_k(h) \cdot 1_{[(k-1)h, kh)}(t), \quad t \geq 0, \quad \bar{z}(h; t) = \sum_{k=1}^{\infty} \bar{u}_k(h) \cdot 1_{[(k-1)h, kh)}(t), \quad t \geq 0.$$

Mit dem in iii) definierten $Z_k(h; t)$ sind nach iii)

$$\underline{Z}(h; t) = \sum_{k=1}^{\infty} \underline{u}_k(h) \cdot Z_k(h; t), \quad t \geq 0, \quad \overline{Z}(h; t) = \sum_{k=1}^{\infty} \overline{u}_k(h) \cdot Z_k(h; t), \quad t \geq 0,$$

die Lösungen der Erneuerungsgleichung zu $\underline{z}(h; \bullet)$ bzw. $\overline{z}(h; \bullet)$. Es folgt

$$\lim_{t \rightarrow \infty} \underline{Z}(h; t) = \lim_{t \rightarrow \infty} \sum_{k=1}^{\infty} \underline{u}_k(h) \cdot Z_k(h; t) = \sum_{k=1}^{\infty} \underline{u}_k(h) \lim_{t \rightarrow \infty} Z_k(h; t) = \sum_{k=1}^{\infty} \underline{u}_k(h) \frac{h}{\mu} = \frac{1}{\mu} \underline{\sigma}(h),$$

bzw.

$$\lim_{t \rightarrow \infty} \overline{Z}(h; t) = \frac{1}{\mu} \overline{\sigma}(h).$$

Dabei können Grenzwert und Summe vertauscht werden, da die absolute Konvergenz der Reihen $\underline{\sigma}(h)$ und $\overline{\sigma}(h)$ (der Forderung (1) in iii) entsprechend) und (2) die gleichmäßige Konvergenz in t der Reihen $\underline{Z}(h; \bullet)$ und $\overline{Z}(h; \bullet)$ implizieren. Aus $\underline{Z}(h; t) \leq Z(t) \leq \overline{Z}(h; t)$ (nach Definition und Darstellung als Lösung der Erneuerungsgleichung) folgt

$$\frac{\underline{\sigma}(h)}{\mu} \leq \liminf_{t \rightarrow \infty} Z(t) \leq \limsup_{t \rightarrow \infty} Z(t) \leq \frac{\overline{\sigma}(h)}{\mu}, \quad h > 0.$$

Die Riemann-Integrierbarkeit mit $h \rightarrow 0$ liefert die Behauptung. ■

Für den arithmetischen Fall haben wir:

5.31 Satz (Erneuerungstheorem, 2. Version)

Mit den Voraussetzungen und Bezeichnungen aus Satz 5.30 sei F (statt nicht-arithmetisch) arithmetisch mit Spanne $\lambda > 0$. Dann gilt:

$$\lim_{n \rightarrow \infty} Z(x + n\lambda) = \frac{\lambda}{\mu} \sum_{j=1}^{\infty} z(x + j\lambda), \quad x \geq 0.$$

5.32 Anmerkung

Die Aussage des Erneuerungstheorems in der ersten Version für den nicht-arithmetischen Fall kann man wie folgt interpretieren.

$M(t)$, $t \geq 0$, ist eine isotone, nicht-negative Funktion, also eine maßdefinierende Funktion. Bezeichnen wir ebenfalls mit M das davon auf $(\mathbb{R}_+, \mathbb{B}_+)$ erzeugte „Erneuerungsmaß“, d. h.

$$M((a, b]) := M(b) - M(a) \quad \text{für alle } a, b \in \mathbb{R} \text{ mit } a \leq b,$$

so hat das Maß M asymptotisch die λ^1 -Dichte μ^{-1} . □

5.33 Beispiel (Alterserneuerungspolitik: Fortsetzung der Beispiele 5.16 und 5.21)

Werden die anfallenden Kosten für die zukünftigen Erneuerungen nicht diskontiert, d. h. ist im Formalismus von Beispiel 5.16 $\rho = 0$, so ist $z(\bullet)$ wegen

$$z(t) = \int_{[0, t]} L(x) \overline{F}(dx) \xrightarrow{t \rightarrow \infty} c_1 \cdot F(T-) + c_2 \cdot (1 - F(T-)) = z(T)$$

nicht direkt Riemann-integrierbar. Die erwarteten (nicht-diskontierten) Gesamtkosten in $[0, \infty)$ sind unendlich groß. Man kann also das Erneuerungstheorem nicht zur Berechnung der entstehenden Kosten verwenden, um daraus Erneuerungsstrategien abzuleiten.

Ein anderes Kriterium, das man in diesem Fall einsetzen kann, ist die Frage nach den erwarteten durchschnittlichen Kosten über einen unendlich langen Zeitraum; zu untersuchen wäre also $\lim_{t \rightarrow \infty} \frac{c(t)}{t}$, wobei $c(t)$

die erwarteten Kosten in $[0, t]$ sind. Um das unbeschränkte Wachstum von $c(\bullet)$ auszugleichen, untersuchen wir statt $c(\bullet)$ die Funktion

$$d(t) = z(T)(\bar{U}(t) + 1) - c(t), \quad t \geq 0,$$

wobei $\bar{U}(\bullet)$ die Erneuerungsfunktion zu $\bar{F}(\bullet)$ ist. Nun erfüllt $\bar{U}(t)$ die Erneuerungsgleichung

$$\bar{U}(t) = \bar{F}(t) + \int_{[0,t]} \bar{U}(t-x)\bar{F}(dx) = \int_{[0,t]} (1 + \bar{U}(t-x))\bar{F}(dx), \quad t \geq 0,$$

während

$$c(t) = z(t) + \int_{[0,t]} c(t-x)\bar{F}(dx), \quad t \geq 0,$$

gilt. Daraus folgt

$$d(t) = z(T) - z(t) + \int_{[0,t]} d(t-x)\bar{F}(dx), \quad t \geq 0,$$

und $z(T) - z(t)$ ist monoton fallend für $t \rightarrow \infty$:

$$z(T) - z(t) = \begin{cases} c_1(F(T-) - F(t)) + c_2(1 - F(T-)), & \text{für } t < T, \\ 0, & \text{für } t \geq T. \end{cases}$$

Weiter ist für $T < \infty$

$$\int_{[0,\infty)} (z(T) - z(t)) dt = T(c_1F(T-) + c_2(1 - F(T-))) - \int_{[0,\infty)} c_1F(t) dt < \infty.$$

Auf $z(T) - z(t)$, $t \geq 0$, kann man also Anmerkung 5.29 f) anwenden: $z(T) - z(t)$ ist monoton und es gilt

$$\int_{[0,\infty)} |z(T) - z(t)| dt < \infty \quad \text{für } T \in [0, \infty);$$

also folgt die direkte Riemann-Integrierbarkeit von $z(T) - z(t)$, $t \geq 0$. Die Anwendung der zweiten Version des Erneuerungstheorems liefert für $T < \infty$

$$\lim_{t \rightarrow \infty} d(t) = \frac{1}{\bar{\mu}} \left(T(c_1F(T-) + c_2(1 - F(T-))) - \int_{[0,\infty)} c_1F(t) dt \right) < \infty,$$

wobei

$$\bar{\mu} = \int_{[0,\infty)} x\bar{F}(dx) = \int_{[0,T-]} xF(dx) + T(1 - F(T-))$$

der Erwartungswert von $\bar{F}(\bullet)$ ist. Aus der Definition von $d(\bullet)$ erhält man

$$\lim_{t \rightarrow \infty} \frac{c(t)}{t} = \lim_{t \rightarrow \infty} \frac{z(T)(\bar{U}(t) + 1) - d(t)}{t} = \lim_{t \rightarrow \infty} \frac{z(T)\bar{U}(t)}{t} + \underbrace{\lim_{t \rightarrow \infty} \frac{z(T) - d(t)}{t}}_{\rightarrow 0, \text{ da } \lim_{t \rightarrow \infty} d(t) < \infty} = \lim_{t \rightarrow \infty} \frac{z(T)}{\bar{\mu}} \cdot \bar{\mu} \frac{\bar{U}(t)}{t}.$$

Kennt man also $\lim_{t \rightarrow \infty} \frac{\bar{U}(t)}{t} = \frac{1}{\bar{\mu}}$ (siehe Korollar 5.35), so folgt

$$\lim_{t \rightarrow \infty} \frac{c(t)}{t} = \frac{z(T)}{\bar{\mu}} = \frac{c_1F(T-) + c_2(1 - F(T-))}{\bar{\mu}}. \quad (1)$$

Auf demselben Weg bestimmt man für $T = \infty$, d. h. keine vorzeitige Erneuerung, die erwarteten asymptotischen Durchschnittskosten als

$$\lim_{t \rightarrow \infty} \frac{c(t)}{t} = \frac{c_1}{\mu}, \quad (2)$$

so daß auch unter diesem Kriterium nach T optimiert werden kann.

Zwei Spezialfälle für $T < \infty$:

- $c_1 = 0$, $c_2 = 1$: $c(t)$ zählt die erwartete Anzahl der „geplanten“, vorzeitigen Erneuerungen.
- $c_1 = 1$, $c_2 = 0$: $c(t)$ zählt die erwartete Anzahl der „ungeplanten“, überraschenden Erneuerungen. \square

5.34 Anmerkung

Die in den Gleichungen (1) und (2) von Beispiel 5.33 angegebenen Grenzwerte werden auch als „erwartete Durchschnittskosten pro Zeiteinheit“ bezeichnet. In der dort angegebenen Formulierung als Grenzwert über einen „unendlichen Planungshorizont“ ist dies korrekt. 5.33 (1), (2) sagen aber nicht, daß in einem endlichen Intervall der Länge 1 diese Kosten im Mittel entstehen. Dies würde gelten, wenn sich das System schon eingeschwungen hat. \square

5.35 Korollar

Die Verteilungsfunktion F habe endlichen Erwartungswert μ . Dann gilt für die Erneuerungsfunktion

$$\frac{U(t)}{t} \xrightarrow{t \rightarrow \infty} \frac{1}{\mu}.$$

Beweis: Sei $a_n := U(n) - U(n-1)$; dann gilt

- für F nicht-arithmetisch: Aus Satz 5.27 folgt $a_n \rightarrow \mu^{-1}$.
- für F arithmetisch mit Spanne 1: Aus Satz 5.23 folgt $a_n \rightarrow \mu^{-1}$.

Es folgt

$$\frac{1}{n}U(n) = \frac{1}{n} \left(\sum_{k=1}^n a_k \right) \rightarrow \frac{1}{\mu}.$$

Für beliebiges $t \in \mathbb{R}$ gilt

$$\frac{\lceil t \rceil}{t} \cdot \frac{U(\lceil t \rceil)}{\lceil t \rceil} \leq \frac{U(t)}{t} \leq \frac{\lceil t+1 \rceil}{t} \cdot \frac{U(\lceil t+1 \rceil)}{\lceil t+1 \rceil}. \quad \blacksquare$$

5.36 Definition

Ein nicht-verzögerter Erneuerungsprozeß $(S_k : k \in \mathbb{N})$ mit Erneuerungsverteilungsfunktion $F(t) = 1 - e^{-\lambda t}$, $t \geq 0$, heißt Poisson-Prozeß, der zugehörige Zählprozeß $(N(t) : t \geq 0)$ wird als Poisson-Zählprozeß mit Parameter λ bezeichnet.

In den meisten Büchern wird auch der Poisson-Zählprozeß kurz als Poisson-Prozeß bezeichnet, wobei dann auch noch zwischen $(S_k : k \in \mathbb{N})$ und $(N_t : t \in \mathbb{R}_+)$ nicht unterschieden wird. Dies ist insofern auch für allgemeine (un)verzögerte Erneuerungsprozesse gerechtfertigt, da die Pfade von S und N wechselseitig sich eindeutig bestimmen.

Die Bedeutung des Poisson-Prozesses in Theorie und Praxis kann kaum unterschätzt werden. Gründe dafür sind:

- Der Prozeß ist analytisch einfach zu untersuchen und man kann für viele interessierende Größen geschlossene Ausdrücke angeben.
- Der Poisson-Prozeß tritt auf als Grenzprozeß in Sätzen über die Konvergenz von Folgen stochastischer Prozesse: Ist ein Zählprozeß darstellbar als Überlagerung (also Summe) einer sehr großen Anzahl unabhängiger Zählprozesse, von denen jeder einzelne nur unbedeutend zum Gesamtprozeß beiträgt, so ist der Überlagerungsprozeß ein Poisson-Prozeß. (Man beachte die Analogie zur Theorie der Meßfehlerverteilung, die zur Normalverteilung als Meßfehlerverteilung führt.) Eine ausführliche, sehr viel weiter gehende Darstellung findet man in [GK74], S. 112 ff.
- Der Poisson-Prozeß tritt tatsächlich häufig in der Praxis auf; erklärbar ist dies durch den eben skizzierten Satz: der Ankunftsstrom von Gesprächen in einer Telefonzentrale ist die Überlagerung der Prozesse, die die Gesprächswünsche der einzelnen Teilnehmer zählen; der Ankunftsstrom von Aufträgen an einem Zentralrechner mit einer großen Anzahl von Terminals ist die Überlagerung der Ankunftsströme, die von den einzelnen Terminals erzeugt werden.
- Selbst wenn man weiß, daß ein Zählprozeß tatsächlich kein Poisson-Prozeß ist, verwendet man in der Modellierung komplexer Systeme häufig eine Poisson-Approximation, weil nur unter dieser Zusatzannahme noch vernünftig handhabbare Ergebnisse zu erreichen sind.

- Teilchenströme beim radioaktiven Zerfall können als (fast ideale) Poisson-Prozesse angesehen werden, solange sich die Masse der radioaktiven Probe nicht (wesentlich) ändert.

Verallgemeinerungen des Poisson-Prozesses, die zumindest einen Teil der „schönen“ Eigenschaften des Poisson-Prozesses erhalten, findet man z. B. in [GK74].

5.37 Korollar

Für einen Poisson-Zählprozeß mit Parameter λ gilt:

- $N_t \sim \pi(\lambda t)$ für $t \geq 0$;
- $U(t) = \lambda t$ für $t \geq 0$.

Beweis: Nach Lemma 5.6 b) gilt

$$\begin{aligned} P(N_t = n) &= F^{n*}(t) - F^{(n+1)*}(t) \\ &= \left(1 - \sum_{k=1}^{n-1} e^{-\lambda t} \frac{(\lambda t)^k}{k!}\right) - \left(1 - \sum_{k=1}^n e^{-\lambda t} \frac{(\lambda t)^k}{k!}\right) = e^{-\lambda t} \frac{(\lambda t)^n}{n!} = \pi(\lambda t)(n). \quad \blacksquare \end{aligned}$$

Die im Erneuerungssatz gemachte Grenzwertaussage lautet für den Poisson-Prozeß mit Parameter λ (d. h. $E(F) = \lambda^{-1}$):

$$U(t+h) - U(t) \rightarrow \frac{h}{\lambda^{-1}}, \quad t \geq 0, \quad h > 0.$$

Die Aussage 5.37 b) zeigt, daß in diesem Fall schon die entsprechende endliche Aussage gilt, mit anderen Worten: das Erneuerungsmaß im Poisson-Prozeß mit Parameter λ hat die λ^1 -Dichte λ . Üblicherweise bezeichnet man z. B. in einem Teilchenstrom die mittlere Anzahl von Teilchen pro (infinitesimale) Zeit als Intensität des Teilchenstromes. Einen Poisson-Prozeß mit Parameter λ bezeichnet man deshalb auch als Poisson-Prozeß mit Intensität λ .

5.38 Beispiel (Wartezeitparadoxon, Inspektionsparadoxon)

Die Folge von Ankunfts Augenblicken an einer Bushaltestelle sei für die Busse ein Poisson-Prozeß mit Parameter $\lambda > 0$. Ein Passagier komme zur Zeit $t \geq 0$ an der Haltestelle an (ohne Kenntnis über die vergangenen und zukünftigen Ankunftszeiten). Sei W_t die Wartezeit bis zur Ankunft des nächsten Busses. Gesucht ist $E(W_t)$. Es gibt die Behauptungen:

- Da der Ankunftszeitpunkt t unabhängig vom Ankunftsstrom gewählt war, ist die mittlere Wartezeit aus Symmetriegründen gerade die Hälfte des Erwartungswertes der zur Zeit t laufenden Zwischenankunftszeit.
- Die Zwischenankunftszeiten sind gedächtnislos. Also ist die Wartezeit nach t verteilt wie die Zwischenankunftszeit.

Augenscheinlich bedeutet II) $E(W_t) = \lambda^{-1}$, während I) offensichtlich $E(W_t) = \frac{1}{2}\lambda^{-1}$ impliziert. Da das hier auftretende Problem auch in der Praxis erst verschwindet, wenn die mittlere Zwischenankunftszeit unterhalb der Rechengenauigkeit des verwendeten Rechners liegt, bietet sich eine theoretische Analyse des Problems an. Es wird sich herausstellen, daß die Behauptungen in I) und II) beide korrekt sind. \square

5.39 Definition

Gegeben sei ein Erneuerungsprozeß $(S_k : k \in \mathbb{N})$ mit Zählprozeß $(N_t : t \in \mathbb{R})$. Für jedes feste $t \in \mathbb{R}_+$ sind dann die Zufallsvariablen

$$\begin{aligned} S_{N(t)+1} - t &: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}_+, \overline{\mathbb{B}}_+) \quad (\text{Restlebensdauer zur Zeit } t, \text{ Vorwärtsrekurrenzzzeit}) \\ t - S_{N(t)} &: (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}_+, \mathbb{B}_+) \quad (\text{Alter zur Zeit } t, \text{ Rückwärtsrekurrenzzzeit}) \\ S_{N(t)+1} - S_{N(t)} &: (\Omega, \mathcal{A}, P) \rightarrow (\overline{\mathbb{R}}_+, \overline{\mathbb{B}}_+) \quad (\text{inspizierte Lebensdauer zur Zeit } t) \end{aligned}$$

eindeutig (punktweise) definiert.

5.40 Korollar

Gegeben sei ein unverzögerter Erneuerungsprozeß. Mit den Bezeichnungen aus Definition 5.39 sei

$$\begin{aligned} R_z(t) &:= P(S_{N(t)+1} - t > z), & 0 \leq z; \\ A_z(t) &:= P(t - S_{N(t)} \geq z), & 0 \leq z \leq t; \\ L_z(t) &:= P(S_{N(t)+1} - S_{N(t)} > z), & 0 \leq z. \end{aligned}$$

Dann gilt:

- a) $A_z(t) = R_z(t - z)$ für $z < t$.
 b) $R_z(t)$ genügt der Erneuerungsgleichung

$$R_z(t) = 1 - F(t + z) + \int_{[0,t]} R_z(t - x)F(dx), \quad z \geq 0.$$

- c) $L_z(t)$ genügt der Erneuerungsgleichung

$$L_z(t) = 1 - F(\max(z, t)) + \int_{[0,t]} L_z(t - x)F(dx), \quad z \geq 0.$$

Beweis:

- a) Es gilt

$$R_z(t) = P(S_{N(t)+1} > t + z) = P(\text{keine Erneuerung in } (t, t + z])$$

und

$$A_z(t) = P(S_{N(t)} \leq t - z) = P(\text{keine Erneuerung in } (t - z, t]) = R_z(t - z), \quad \text{falls } z \leq t \text{ ist.}$$

- b) Das Ereignis „keine Erneuerung in $(t, t + z]$ “ wird disjunkt zerlegt in „keine Erneuerung in $(t, t + z]$ \wedge keine Erneuerung in $[0, t]$ “ und „keine Erneuerung in $(t, t + z]$ \wedge mindestens eine Erneuerung in $[0, t]$ “. Das zweite Ereignis wird weiter zerlegt über die Eintrittszeit der ersten Erneuerung, die nach $F(\bullet)$ verteilt ist. Es folgt

$$\begin{aligned} R_z(t) &= P(\text{keine Erneuerung in } [0, t + z]) \\ &\quad + \int_{[0,t]} P(\text{keine Erneuerung in } (t, t + z] \mid \text{erste Erneuerung zur Zeit } x)F(dx) \\ &= 1 - F(t + z) + \int_{[0,t]} R_z(t - x)F(dx). \end{aligned}$$

- c) Wir zerlegen wieder:

$$\begin{aligned} P(S_{N(t)+1} - S_{N(t)} > z) &= P(S_{N(t)+1} - S_{N(t)} > z \wedge \text{keine Erneuerung bis } t) \\ &\quad + P(S_{N(t)+1} - S_{N(t)} > z \wedge \text{mindestens eine Erneuerung bis } t) \\ &= P(S_{0+1} - S_0 > z \wedge S_1 > t) \\ &\quad + \int_{[0,t]} P(S_{N(t)+1} - S_{N(t)} > z \mid S_1 = x)F(dx) \\ &= 1 - F(\max(z, t)) + \int_{[0,t]} P(S_{N(t-x)+1} - S_{N(t-x)} > z)F(dx) \quad \blacksquare \end{aligned}$$

5.41 Korollar

Die Restlebensdauerverteilungsfunktion im Poisson-Prozeß mit Parameter $\lambda > 0$ ist $\exp(\lambda)(\bullet)$, also gleich der Lebensdauerverteilungsfunktion. Die Altersverteilungsfunktion ist gegeben durch

$$\begin{cases} 1 - A_z(t) = 1 - e^{-\lambda z}, & \text{falls } z \leq t; \\ 1 & \text{sonst;} \end{cases}$$

und die Zufallsvariable $S_{N(t)+1} - S_{N(t)}$ hat die λ^1 -Dichte

$$\begin{cases} \lambda^2 x e^{-\lambda x}, & \text{falls } 0 < x \leq t; \\ \lambda(1 + \lambda t) e^{-\lambda x}, & \text{falls } t < x. \end{cases}$$

Beweis:

- i) Die Erneuerungsgleichung für $R_z(t)$ wird durch $e^{-\lambda z}$ gelöst, unabhängig von t . Dies impliziert über Korollar 5.40 a) die Aussage über die Altersverteilung.
- ii) Für $x \geq 0$ ist für festes $t > 0$:

$$\begin{aligned} & P(S_{N(t)+1} - S_{N(t)} \leq x) \\ &= \sum_{n=0}^{\infty} P(\underbrace{S_n \leq t < S_{n+1}}_{\Leftrightarrow N_t=n}, X_{N(t)+1} \leq x) \\ &= P(0 \leq t < X_1, X_1 \leq x) \\ &\quad + \sum_{n=1}^{\infty} \int_{[0, \infty)} P(X_{n+1} \leq x, S_n \leq t < S_n + X_{n+1} \mid S_n = y) dP(S_n \leq y) \\ &= P(t < X_1 \leq x) + \sum_{n=1}^{\infty} \int_{[0, t]} P(X_{n+1} \leq x, t < S_n + X_{n+1} \mid S_n = y) dP(S_n \leq y) \\ &= P(t < X_1 \leq x) + \sum_{n=1}^{\infty} \int_{[0, t]} P(X_{n+1} \leq x, t < y + X_{n+1} \mid S_n = y) dP(S_n \leq y) \end{aligned}$$

($\{X_1, \dots, X_n\}$ und X_{n+1} sind unabhängig.)

$$= P(t < X_1 \leq x) + \sum_{n=1}^{\infty} \int_{[0, t]} P(t - y < X_{n+1} \leq x) \lambda \frac{(\lambda y)^{n-1}}{(n-1)!} e^{-\lambda y} \lambda^1(dy).$$

Da $P(t - y < X_{n+1} \leq x) = 0$, falls $t - y \geq x$, d. h. $y \leq t - x$, unterscheiden wir zwei Fälle:

a) $x \leq t$:

$$\begin{aligned} P(S_{N(t)+1} - S_{N(t)} \leq x) &= \int_{[t-x, t]} \lambda \sum_{n=1}^{\infty} \frac{(\lambda y)^{n-1}}{(n-1)!} e^{-\lambda y} (e^{-\lambda(t-y)} - e^{-\lambda x}) \lambda^1(dy) \\ &= 1 - e^{-\lambda x} - \lambda x e^{-\lambda x}. \end{aligned}$$

b) $x > t$:

$$\begin{aligned} P(S_{N(t)+1} - S_{N(t)} \leq x) &= e^{-\lambda t} - e^{-\lambda x} + \int_{[0, t]} \lambda \sum_{n=1}^{\infty} \frac{(\lambda y)^{n-1}}{(n-1)!} e^{-\lambda y} (e^{-\lambda(t-y)} - e^{-\lambda x}) \lambda^1(dy) \\ &= 1 - e^{-\lambda x} - \lambda t e^{-\lambda x}. \end{aligned}$$

Ableitung der beiden Ausdrücke ergibt die Dichte. ■

5.42 Beispiel

Es gilt

$$E(S_{N(t)+1} - S_{N(t)}) = \frac{2}{\lambda} - \frac{1}{\lambda} e^{-\lambda t},$$

was für großes t

$$E(S_{N(t)+1} - S_{N(t)}) \approx \frac{2}{\lambda}$$

entspricht, so daß die Argumentation I) in Beispiel 5.38 bezüglich der Symmetrie aufrechterhalten werden kann, trotzdem Korollar 5.41 gerade die Argumentation 5.38 II) bestätigte. Der Korrektursummand $\frac{1}{\lambda} e^{-\lambda t}$

geht zurück auf die unterschiedliche Darstellung der Dichte von $S_{N(t)+1} - S_{N(t)}$ für $x \leq t$ bzw. $x > t$. Diese wiederum ist begründet in der speziellen Rolle, die der Nullpunkt der Zeitachse spielt. Vorausgesetzt, daß wir eine beidseitig unendliche Zeitachse haben, erhalten wir $\frac{2}{\lambda}$ als erwartete Länge des inspizierten Intervalls.

Als anschauliche Begründung wird in der Regel angegeben: Bei der zufälligen Auswahl eines $t \in \mathbb{R}$ trifft man mit größerer Wahrscheinlichkeit längere Intervalle. \square

Während für den Poisson-Prozeß noch explizite Angaben über das Verhalten im Endlichen gemacht werden können, ist dies bei allgemeinen Lebensdauerverteilungsfunktionen nicht mehr möglich. Andererseits zeigt sich in diesen Fällen die Bedeutung der Erneuerungstheoreme, die Approximationen liefern für Prozesse, die schon lange laufen.

5.43 Satz

Gegeben sei ein rekurrenter (möglicherweise verzögerter) Erneuerungsprozeß mit Lebensdauerverteilungsfunktion F , deren Erwartungswert $\mu = E(F) < \infty$ erfülle. Dann sind (für $t \rightarrow \infty$) die asymptotischen Verteilungen von Restlebensdauer und Alter gegeben durch die Verteilungsfunktion

$$\frac{1}{\mu} \int_{[0,z]} (1 - F(x)) dx, \quad z \geq 0.$$

Die asymptotische Verteilung der inspizierten Lebensdauer ist durch die Verteilungsfunktion

$$\frac{1}{\mu} \int_{[0,z]} xF(dx), \quad z \geq 0,$$

gegeben.

Beweis:

- i) a) In der Erneuerungsgleichung von Korollar 5.40 b) ist $1 - F(t+z)$ in t fallend und integrierbar, also direkt Riemann-integrierbar. Das Erneuerungstheorem liefert in seiner zweiten Version also

$$\lim_{t \rightarrow \infty} R_z(t) =: R_z(\infty) = \frac{1}{\mu} \int_{[0,\infty)} (1 - F(x+z)) dx = \frac{1}{\mu} \int_{(z,\infty)} (1 - F(x)) dx.$$

Die Aussage über die Altersverteilung folgt aus Korollar 5.40 a).

- b) In der Erneuerungsgleichung von Korollar 5.40 c) ist $1 - F(\max(z,t))$ als Funktion von t direkt Riemann-integrierbar. Also folgt

$$\lim_{t \rightarrow \infty} L_z(t) =: L_z(\infty) = \frac{1}{\mu} \int_{[0,\infty)} (1 - F(\max(z,x))) dx = \frac{1}{\mu} \int_{(z,\infty)} xF(dx).$$

(Fallunterscheidung und partielle Integration; $b(1 - F(b)) \xrightarrow{b \rightarrow \infty} 0$ wegen $\mu < \infty$.)

- ii) Für den verzögerten Fall bedingt man auf den Zeitpunkt der ersten Erneuerung. \blacksquare

Im Poisson-Prozeß ist die Verteilung der Vorwärtsrekurrenzzeit unabhängig vom betrachteten Zeitpunkt, d. h. „stationär“. Für alle anderen unverzögerten Erneuerungsprozesse (mit nicht-arithmetischer Verteilung der Erneuerungsintervalle) gilt dies nicht. Andererseits haben wir in Satz 5.43 gezeigt, daß die Verteilung der Vorwärtsrekurrenzzeit im rekurrenten Erneuerungsprozeß schwach gegen eine Grenzverteilung konvergiert.

Die folgende anschauliche Überlegung führt auf eine notwendige und hinreichende Bedingung für die „Stationarität“ eines rekurrenten Erneuerungsprozesses: Zeitachse sei \mathbb{R} , so daß zur Zeit 0 das System eingeschungen ist, d. h. seinen Gleichgewichtszustand erreicht hat. Nach Satz 5.43 muß dann zur Zeit 0 die Verteilung der Vorwärtsrekurrenzzeit die Verteilungsfunktion

$$z \mapsto \frac{1}{\mu} \int_{[0,z]} (1 - F(x)) dx, \quad z \geq 0,$$

haben, und dies muß auch die Verteilungsfunktion der Restlebensdauer für den Prozeß auf \mathbb{R}_+ sein.

5.44 Definition

Ein verzögerter, rekurrenter Erneuerungsprozeß heißt stationär, wenn die Verteilung der Zeit bis zur ersten Erneuerung gleich der asymptotischen Vorwärtsrekurrenzzeitverteilung (des zugehörigen, unverzögerten Erneuerungsprozesses) ist.

5.45 Satz

Ein verzögerter Erneuerungsprozeß mit nicht-arithmetischer Erneuerungsverteilungsfunktion $F(\bullet)$ und $E(F) = \mu < \infty$ ist genau dann stationär, wenn das Erneuerungsmaß die λ^1 -Dichte μ^{-1} besitzt, d. h. wenn

$$M(t) = \frac{t}{\mu}, \quad t \geq 0,$$

gilt.

Beweis:

- i) Sei $M(t) = \frac{t}{\mu}$. Nach Korollar 5.15 ist $M(\bullet)$ die Lösung der Gleichung $M = G + M * F$, und nach Satz 5.13 ist diese Lösung unter den auf Intervallen beschränkten, auf $(-\infty, 0)$ verschwindenden Funktionen eindeutig bestimmt. Also gilt

$$M(t) = \begin{cases} \frac{t}{\mu}, & t \geq 0 \\ 0, & t < 0 \end{cases} \iff G(t) = \frac{t}{\mu} - \int_{[0,t]} \frac{t-x}{\mu} F(dx), \quad t \geq 0;$$

partielle Integration der rechten Seite führt direkt auf

$$G(t) = \frac{1}{\mu} \int_{[0,t]} (1 - F(x)) dx, \quad t \geq 0.$$

- ii) Der Prozeß sei stationär. Die Verzögerungsverteilung G sei die asymptotische Rekurrenzzeitverteilung. Zu zeigen ist:

$$M(t) = \sum_{k=1}^{\infty} \left(\left(\frac{1}{\mu} \int_{[0,(\bullet)]} (1 - F(x)) dx \right) * F^{(k-1)*} \right) (t) = \frac{t}{\mu}.$$

Wir untersuchen die Laplace-Stieltjes-Transformierte von $M(\bullet)$ (siehe [Hin72], S. 184; [Fel71], S. 429):

$$\widehat{M}(\theta) = \int_{[0,\infty)} e^{-t\theta} M(dt), \quad \theta \geq 0.$$

Ist \widehat{F} die Laplace-Stieltjes-Transformierte von F , so hat

$$G(\bullet) := \int_{[0,(\bullet)]} \frac{1 - F(x)}{\mu} dx$$

die Laplace-Stieltjes-Transformierte

$$\widehat{G}(\theta) = \frac{1 - \widehat{F}(\theta)}{\mu\theta}.$$

Es folgt:

$$\widehat{M}(\theta) = \sum_{k=1}^{\infty} \frac{1 - \widehat{F}(\theta)}{\mu\theta} \widehat{F}(\theta)^{k-1} = \frac{1}{\mu\theta}, \quad \theta \geq 0.$$

Da die Laplace-Stieltjes-Transformierte $\widehat{M}(\bullet)$ das zugehörige Maß eindeutig bestimmt und

$$\theta \mapsto \int_{[0,\infty)} e^{-\theta t} \frac{1}{\mu} dt = \frac{1}{\mu\theta}$$

die Laplace–Stieltjes–Transformierte des durch die maßdefinierende Funktion

$$t \mapsto \frac{t}{\mu}, \quad t \geq 0,$$

bestimmten Maßes ist, folgt die Behauptung. ■

Im Beweis ii) von Satz 5.45 finden wir ein Beispiel für die Nützlichkeit von Integraltransformationen (Laplace–Stieltjes–Transformation, Laplacetransformation, Fouriertransformation bzw. Charakteristische Funktion, erzeugende Funktion usw.). Für die Anwendung von Integraltransformationen in der Erneuerungstheorie siehe z. B.: [Fel71], S. 466, [Cox66].

Unter Benutzung von Satz 5.13 kann man ii) analog zu i) zeigen, indem direkt nachgewiesen wird, daß

$$t \mapsto \frac{t}{\mu}, \quad t \geq 0,$$

die Erneuerungsgleichung

$$M(t) = \frac{1}{\mu} \int_{[0,t]} (1 - F(x)) dx + \int_{[0,t]} M(t-x)F(dx), \quad t \geq 0,$$

löst.

5.46 Korollar

Mit den Bezeichnungen aus Satz 5.45 ist ein verzögerter, rekurrenter Erneuerungsprozeß mit nicht–arithmetischer Erneuerungsverteilungsfunktion genau dann stationär, wenn die eindimensionalen Randverteilungen des Prozesses $(S_{N_t+1} - t : t \in \mathbb{R}_+)$ von t unabhängig sind und die λ^1 –Dichte

$$\frac{1}{\mu}(1 - F(x)), \quad x \geq 0,$$

auf \mathbb{R}_+ haben.

Beweis: Siehe [Coh82], S. 111. (Im Buch von Cohen findet man eine sehr ausführliche Diskussion von Alters– und Restlebenszeitprozessen.) ■

5.47 Anmerkung

Die asymptotische (und stationäre) Vorwärtsrekurrenzzzeit im rekurrenten Erneuerungsprozeß hat genau dann ein r –tes Moment ($r \geq 1$), wenn die Erneuerungsintervalle ein $(r + 1)$ –tes Moment haben. □

Im Korollar 5.37 war gezeigt:

$$\frac{U(t)}{t} \xrightarrow{t \rightarrow \infty} \frac{1}{\mu}.$$

Eine Konvergenzabschätzung gibt der folgende Satz.

5.48 Satz

Sei F nicht–arithmetisch mit Erwartungswert μ und Varianz σ^2 . Dann gilt

$$0 \leq U(t) - \frac{t - \mu}{\mu} \xrightarrow{t \rightarrow \infty} \frac{\sigma^2 + \mu^2}{2\mu^2}.$$

Beweis: Sei

$$z(t) := \frac{1}{\mu} \int_{[t,\infty)} (1 - F(x)) dx, \quad t \geq 0,$$

und $Z(t)$, $t \geq 0$, die Lösung der Erneuerungsgleichung $Z = z + Z * F$. Die eindeutig bestimmte, auf Intervallen beschränkte und auf $(-\infty, 0)$ verschwindende Lösung ist

$$Z(t) = U(t) - \frac{t - \mu}{\mu}, \quad t \geq 0; \quad Z(t) = 0, \quad t < 0.$$

Aus der zweiten Version des Erneuerungstheorems folgt

$$\lim_{t \rightarrow \infty} Z(t) = \frac{1}{\mu} \int_{[0, \infty)} z(s) ds = \frac{1}{\mu} \cdot \frac{1}{2\mu} \int_0^\infty y^2 F(dy) = \frac{\sigma^2 + \mu^2}{2\mu^2}.$$

(Partielle Integration nach der auf Seite 68 angegebenen Formel.) ■

Von großer Bedeutung in praktischen Aufgabenstellungen ist die Untersuchung sogenannter kumulativer Prozesse. Ein Beispiel trat in Beispiel 5.33 auf, als $c(t)$ die erwarteten (aufsummierten) Kosten in $[0, t]$ unter gewissen Alterserneuerungspolitiken waren.

5.49 Satz

Sei $\{(Y_i, X_i) : i = 1, 2, \dots\}$ eine Folge von unabhängig identisch verteilten \mathbb{R}^2 -Zufallsvektoren. Die Verteilung der X_i sei auf \mathbb{R}_+ konzentriert mit Verteilungsfunktion $F(\bullet)$ und $E(F) < \infty$. Die Y_i mögen ebenfalls endlichen Erwartungswert besitzen.

Sei $(N_t : t \geq 0)$ der von der Folge $(X_i : i = 1, 2, \dots)$ definierte Zählprozeß. Dann gilt für

$$Z(t) = \sum_{i=1}^{N_t+1} Y_i, \quad t \geq 0,$$

den von $\{(Y_i, X_i) : i = 1, 2, \dots\}$ erzeugten kumulativen Prozeß:

$$\lim_{t \rightarrow \infty} \frac{E(Z(t))}{t} = \frac{E(Y_1)}{E(X_1)}.$$

Beweis: Sieht man die Y_i als Auszahlungen an, die beim Beginn des i -ten Erneuerungsintervalls $[S_{i-1}, S_i)$ fällig werden, $i = 1, 2, \dots$, so erhält man im Mittel zur Zeit $x < t$ die erste Erneuerung $E(Y_1)$; findet dann zur Zeit $x < t$ die erste Erneuerung statt, so kann man noch als Auszahlung bis zum Zeitpunkt t gerade $E(Z(t-x))$ erwarten. Insgesamt in $[0, t]$ also

$$E(Z(t)) = E(Y_1) + \int_{[0, t]} E(Z(t-x)) F(dx), \quad t \geq 0.$$

Die Lösung dieser Erneuerungsgleichung ist

$$Z(t) = E(Y_1)(1 + U(t)), \quad t \geq 0,$$

wobei $U(t)$ die zu $F(\bullet)$ gehörige Erneuerungsfunktion ist. Aus Korollar 5.35 folgt die Behauptung. ■

5.50 Anmerkung

Ist die Verteilung von X_i , $i = 1, 2, \dots$, in Satz 5.49 arithmetisch auf \mathbb{N}_0 konzentriert, so ist es einfacher, mit

$$\lim_{n \rightarrow \infty} \frac{E(Z(n))}{n} = \frac{E(Y_1)}{E(X_1)}$$

zu arbeiten. □

Literatur

- [Asm87] Søren Asmussen. *Applied Probability and Queues*. John Wiley & Sons, Inc., Chichester – New York – Brisbane – Toronto – Singapore, 1987.
- [Bac00] L. Bachelier. Théorie de la spéculation. *Annales de l'École normale supérieure*, 17(3):21–86, 1900.
- [Bau74] Heinz Bauer. *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*. Walter de Gruyter & Co., Berlin – New York, second edition, 1974.
- [Ber23] S. N. Bernstein. Démonstration mathématique de la loi d'hérédité de Mendel. *Comptes Rendus Acad. Sci.*, 177:528–531, 1923.
- [BN84] Konrad Behnen and Georg Neuhaus. *Grundkurs Stochastik*. Teubner Studienbücher: Mathematik. Teubner, Stuttgart, 1984.
- [BP78] Richard E. Barlow and Frank Proschan. *Statistische Theorie der Zuverlässigkeit: wahrscheinlichkeitstheoretische Modelle*. Verlag Harry Deutsch, Frankfurt am Main, 1978.
- [Chu67] Kai Lai Chung. *Markov Chains with Stationary Transition Probabilities*. Springer-Verlag, Berlin – Heidelberg – New York, 1967.
- [Chu82] Kai Lai Chung. *Lectures from Markov Processes to Brownian Motion*. Springer-Verlag, New York – Heidelberg – Berlin, 1982.
- [Çin75] Erhan Çinlar. *Introduction to Stochastic Processes*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1975.
- [Coh82] J. W. Cohen. *The Single Server Queue*. North-Holland Publishing Company, Amsterdam – London, second edition, 1982.
- [Cox66] David Roxbee Cox. *Erneuerungstheorie*. R. Oldenbourg Verlag, München – Wien, 1966.
- [DB73] W. Dück and M. Bliefernich. *Operationsforschung: Mathematische Grundlagen, Methoden und Modelle*. VEB Deutscher Verlag der Wissenschaften, Berlin, 1973.
- [Die71] Jean Dieudonné. *Grundzüge der modernen Analysis*. Friedrich Vieweg Sohn Verlagsgesellschaft mbH, Braunschweig, 1971.
- [Ein05] Alfred Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17(4):548–560, 1905.
- [EKKK74] H. Ehrig, K.-D. Kiermeier, H.-J. Kreowski, and W. Kühnel. *Universal Theory of Automata*. Teubner, Stuttgart, 1974.
- [Fel68] William Feller. *An Introduction to Probability Theory and Its Applications*, volume I. John Wiley & Sons, Inc., New York – London – Sidney, third edition, 1968.
- [Fel71] William Feller. *An Introduction to Probability Theory and Its Applications*, volume II. John Wiley & Sons, Inc., New York – Chichester – Brisbane – Toronto, second edition, 1971.
- [Fre83] David Freedman. *Markov Chains*. Springer-Verlag, Berlin – Heidelberg – New York, 1983. Nachdruck des 1971 bei Holden-Day, San Francisco, erschienenen Buches.
- [GF68] Hans Grauert and Wolfgang Fischer. *Differential- und Integralrechnung II*. Springer-Verlag, Berlin – Heidelberg – New York, 1968.
- [GK74] B. W. Gnedenko and I. N. Kowalenko. *Einführung in die Bedienungstheorie*. Akademie-Verlag GmbH, Berlin, 1974.

- [Gra81] W. K. Grassmann. *Stochastic Systems for Management*. North Holland, 1981.
- [GS80] Iosif Il'ich Gihman and Anatoliĭ Vladimirovich Skorohod. *The Theory of Stochastic Processes*, volume I. Springer-Verlag, Berlin – Heidelberg – New York, Nachdruck 1980.
- [Hak78] H. Haken. *Synergetics, An Introduction (Non-equilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology)*. Springer-Verlag, Berlin – Heidelberg – New York, second edition, 1978.
- [Hin72] Karl Hinderer. *Grundbegriffe der Wahrscheinlichkeitstheorie*. Springer-Verlag, Berlin – Heidelberg – New York, 1972.
- [Hun83] Jeffrey J. Hunter. *Mathematical Techniques of Applied Probability*, volume II: *Discrete Time Models: Techniques and Applications*. Academic Press, New York – London – Paris – San Diego – San Francisco – São Paulo – Sydney – Tokyo – Toronto, 1983.
- [JR91] B. H. Juang and L. R. Rabiner. Hidden Markov models for speech recognition. *Technometrics*, 33(3):251–272, 1991.
- [Koh77] Jürg Kohlas. *Stochastische Methoden des Operations Research*. Teubner Studienbücher: Mathematik, 40. B. G. Teubner, Stuttgart, 1977.
- [KS83] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer-Verlag, New York – Berlin – Heidelberg – Tokyo, 1983. Nachdruck des 1960 bei Van Nostrand, Princeton, erschienenen Buches.
- [KSK76] John G. Kemeny, J. Laurie Snell, and Anthony W. Knapp. *Denumerable Markov Chains*. Springer-Verlag, New York – Heidelberg – Berlin, 1976. Nachdruck des 1966 bei Van Nostrand, Princeton, erschienenen Buches.
- [KT75] Samuel Karlin and Howard M. Taylor. *A First Course in Stochastic Processes*. Academic Press, New York – San Francisco – London, second edition, 1975.
- [KT81] Samuel Karlin and Howard M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981.
- [Lin91] T. Lindvall. W. Doeblin, 1915 – 1940. *Ann. Prob.* 19, pages 929–934, 1991.
- [Mai74] L. E. Maistrov. *Probability Theory, A Historical Sketch*. Academic Press, New York – London, 1974.
- [Mar13] A. A. Markov. Versuche einer statistischen Untersuchung über den Text des Romans ‚Eugen Onegin‘ zur Beleuchtung des Zusammenhangs der Kettenversuche. *Mitteilungen der Petersburger Akademie der Wissenschaften*, 7(6):153–162, 1913.
- [Sch81] Christoph Schneeweiß. *Modellierung industrieller Lagerhaltungssysteme: Einführung und Fallstudien*. Springer-Verlag, Berlin – Heidelberg – New York, 1981.
- [Sch88] I. Schneider, editor. *Die Entwicklung der Wahrscheinlichkeitstheorie von den Anfängen bis 1933*. Wissenschaftliche Buchgesellschaft, Darmstadt, 1988.
- [Sen80] E. Seneta. *Non-negative Matrices – An Introduction to Theory and Applications*. Springer-Verlag, 1980.
- [Tho85] H. Thorrisson. Vorlesungsmanuskript. Stanford University, 1985.
- [Unb90] Rolf Unbehauen. *Systemtheorie; Grundlagen für Ingenieure*. R. Oldenbourg Verlag, München – Wien, 5. auflage edition, 1990.
- [Wol89] R. W. Wolff. *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Inc., Englewood Cliffs, N. J., 1989.