

## 3. Prozesse mit kontinuierlicher Zeit

### 3.1 Einführung

Wir betrachten nun Markov-Ketten  $(X(t))_{t \in \mathbb{R}_0^+}$ .

Wie beim Übergang von der geometrischen zur Exponentialverteilung können wir uns auch hier einen Grenzprozess vorstellen.

Wie dort folgt, dass die Aufenthaltsdauer im Zustand 0 gemessen in Schritten der diskreten Markov-Kette geometrisch verteilt ist und im Grenzwert  $n \rightarrow \infty$  in eine kontinuierliche Zufallsvariable übergeht, die exponentialverteilt mit Parameter  $\lambda$  ist. Den Parameter  $\lambda$  bezeichnen wir auch als **Übergangsrate**.

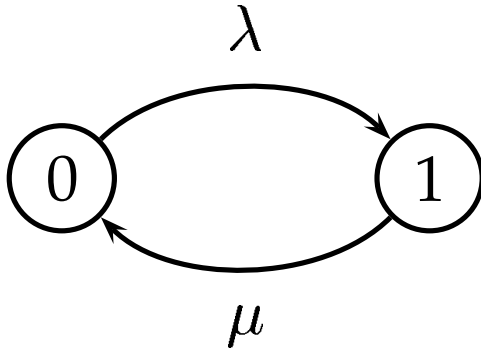


Abbildung: Markov-Kette mit kontinuierlicher Zeit

## Definition 153

Eine unendliche “Folge” von Zufallsvariablen  $X(t)$  ( $t \in \mathbb{R}_0^+$ ) mit Wertemenge  $S$  nennen wir (diskrete) **Markov-Kette mit kontinuierlicher Zeit**, wenn gilt:

- $S$  ist diskret, d.h. wir können ohne Einschränkung annehmen, dass  $S \subseteq \mathbb{N}_0$ .
- Die Zufallsvariablen erfüllen die **Markovbedingung**:  
Für alle  $n \in \mathbb{N}_0$  und beliebige Zeitpunkte  $0 \leq t_0 < t_1 < \dots < t_n < t$  und Zustände  $s, s_0, \dots, s_n \in S$  gilt

$$\Pr[X(t) = s \mid X(t_n) = s_n, \dots, X(t_0) = s_0] = \Pr[X(t) = s \mid X(t_n) = s_n]. \quad (13)$$

Eine Markov-Kette heißt **zeithomogen**, wenn für alle Zustände  $i, j \in S$  und für alle  $u, t \in \mathbb{R}_0^+$  gilt:

$$\Pr[X(t+u) = j \mid X(t) = i] = \Pr[X(u) = j \mid X(0) = i]$$

Die **Markov-Bedingung (13)** besagt anschaulich Folgendes: Wenn wir den Zustand des Systems zu einer Reihe von Zeitpunkten  $t_0 < t_1 < \dots < t_n$  kennen, so ist für das Verhalten nach dem Zeitpunkt  $t_n$  nur der Zustand zur Zeit  $t_n$  maßgebend. Anders formuliert heißt dies: Wenn wir den Zustand des Systems zur Zeit  $t_n$  kennen, so besitzen wir bereits die gesamte relevante Information, um Wahrscheinlichkeiten für das zukünftige Verhalten zu berechnen. Die „Geschichte“ des Systems, d.h. der „Weg“, auf dem der Zustand zur Zeit  $t_n$  erreicht wurde, spielt dabei keine Rolle. Eine Markov-Kette mit kontinuierlicher Zeit ist also ebenso wie eine Markov-Kette mit diskreter Zeit **gedächtnislos**.

Wie schon bei diskreten Markov-Ketten werden wir uns auch bei Markov-Ketten mit kontinuierlicher Zeit auf zeithomogene Markov-Ketten beschränken und diese Eigenschaft im Folgenden stillschweigend voraussetzen.

## Gedächtnislosigkeit der Aufenthaltsdauer

Sei  $Y$  die Aufenthaltsdauer in einem bestimmten Zustand, in dem sich die Markov-Kette zur Zeit  $t = 0$  befindet. Es gilt:

$$\begin{aligned}\Pr[Y \geq t] &= \Pr[X(t') = 0 \text{ für alle } 0 < t' < t \mid X(0) = 0] \\ &= \Pr[X(t' + u) = 0 \text{ für alle } 0 < t' < t \mid X(u) = 0] \\ &= \Pr[X(t' + u) = 0 \text{ für alle } 0 < t' < t \mid X(t'') = 0 \text{ f. a. } 0 \leq t'' \leq u] \\ &= \Pr[X(t') = 0 \text{ für alle } 0 < t' < t + u \mid X(t'') = 0 \text{ f. a. } 0 \leq t'' \leq u] \\ &= \Pr[Y \geq t + u \mid Y \geq u].\end{aligned}$$

Die Aufenthaltsdauer  $Y$  erfüllt also die Bedingung der Gedächtnislosigkeit und muss daher nach Satz 105 exponentialverteilt sein.

## Bestimmung der Aufenthaltswahrscheinlichkeiten

Wie zuvor bei Markov-Ketten mit diskreter Zeit interessieren wir uns auch bei kontinuierlichen Markov-Ketten für die Wahrscheinlichkeit, mit der sich das System zur Zeit  $t$  in einem bestimmten Zustand befindet. Dazu gehen wir von einer **Startverteilung**  $q(0)$  mit  $q_i(0) := \Pr[X(0) = i]$  für alle  $i \in S$  aus und definieren die **Aufenthaltswahrscheinlichkeit**  $q_i(t)$  im Zustand  $i$  zum Zeitpunkt  $t$  durch  $q_i(t) := \Pr[X(t) = i]$ .

Zur Bestimmung dieser Wahrscheinlichkeiten verwenden wir zum einen die soeben gezeigte Tatsache, dass die Aufenthaltsdauer in jedem Zustand  $i$  exponentialverteilt sein muss.

Weiter bezeichnen wir mit  $\nu_{ij}$  die **Übergangsrate** vom Zustand  $i$  in den Zustand  $j$ , sowie  $\nu_i := \sum_{j \in S} \nu_{ij}$ .

Wir betrachten nun ein kleines Zeitintervall  $d t$ . Dann ergibt sich die Änderung der Aufenthaltswahrscheinlichkeit in diesem Zeitintervall als Summe aller „zufließenden“ abzüglich aller „abfließenden“ Wahrscheinlichkeiten. Für alle Zustände  $i \in S$  gilt

$$\underbrace{d q_i(t)}_{\text{Änderung}} = \left( \underbrace{\sum_j q_j(t) \cdot \nu_{ji}}_{\text{Zufluss}} - \underbrace{q_i(t) \nu_i}_{\text{Abfluss}} \right) \cdot d t. \quad (14)$$

Das Lösen des Differentialgleichungssystems (14) ist meist sehr aufwändig. Wir werden es im Folgenden durch Betrachtung des Grenzwertes für  $t \rightarrow \infty$  zu gewöhnlichen linearen Gleichungen vereinfachen.

### Definition 154

Zustand  $j$  ist von  $i$  aus erreichbar, wenn es ein  $t \geq 0$  gibt mit

$$\Pr[X(t) = j \mid X(0) = i] > 0 .$$

Eine Markov-Kette, in der je zwei Zustände  $i$  und  $j$  untereinander erreichbar sind, heißt irreduzibel.



## Satz 155

*Für irreduzible kontinuierliche Markov-Ketten existieren die Grenzwerte*

$$\pi_i = \lim_{t \rightarrow \infty} q_i(t)$$

*für alle  $i \in S$ , und ihre Werte sind unabhängig vom Startzustand.*

Ohne Beweis.

Wenn für  $t \rightarrow \infty$  Konvergenz erfolgt, so gilt

$$\lim_{t \rightarrow \infty} \frac{d q_i(t)}{d t} = 0,$$

da sich  $q_i(t)$  für genügend große  $t$  „so gut wie nicht mehr“ ändert. Diese Gleichung setzen wir in die Differentialgleichungen (14) ein und erhalten

$$0 = \sum_j \pi_j \nu_{ji} - \pi_i \nu_i$$

für alle  $i \in S$ .

Dieses Gleichungssystem hat immer die triviale Lösung  $\pi_i = 0$  für alle  $i \in S$ . Wir suchen jedoch eine Wahrscheinlichkeitsverteilung, und  $\pi$  muss deshalb zusätzlich die Normierungsbedingung  $\sum_{i \in S} \pi_i = 1$  erfüllen. Bei Markov-Ketten mit endlicher Zustandsmenge  $S$  führt dieses Verfahren immer zum Ziel. Wenn  $S$  jedoch unendlich ist, gibt es Fälle, in denen  $\pi_1 = \pi_2 = \dots = 0$  die einzige Lösung darstellt und wir somit keine gültige Wahrscheinlichkeitsverteilung erhalten.

## 3.2 Warteschlangen

Für ein System mit  $m$  Servern und einer gemeinsamen Warteschlange hat sich die Bezeichnung  $X/Y/m$ -Warteschlange eingebürgert. Dabei ersetzt man  $X$  und  $Y$  durch Buchstaben, die jeweils für eine bestimmte Verteilung stehen. Beispielsweise bezeichnet „D“ eine feste Dauer (von engl. *deterministic*), „M“ die Exponentialverteilung (das M kommt von *memoryless*, dem englischen Wort für gedächtnislos) und „G“ eine beliebige Verteilung (von engl. *general*).  $X$  gibt die Verteilung der Zeit zwischen zwei ankommenden Jobs an, während  $Y$  für die Verteilung der eigentlichen Bearbeitungszeit eines Jobs auf dem Server steht (ohne Wartezeit).

### 3.2.1 M/M/1-Warteschlangen

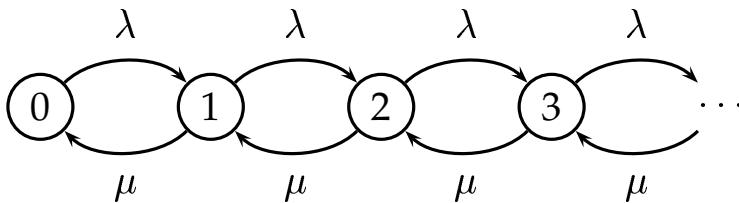


Abbildung: Modellierung einer M/M/1-Warteschlange

Diese Markov-Kette ist irreduzibel, und im Gleichgewichtszustand gelten die Gleichungen

$$\begin{aligned}0 &= \lambda\pi_{k-1} + \mu\pi_{k+1} - (\lambda + \mu)\pi_k \text{ für alle } k \geq 1 \\0 &= \mu\pi_1 - \lambda\pi_0.\end{aligned}$$

Wir definieren die **Verkehrsdichte**  $\rho := \frac{\lambda}{\mu}$  und erhalten:

$$\pi_k = \rho\pi_{k-1} = \dots = \rho^k\pi_0.$$

Damit:

$$1 = \sum_{i=0}^{\infty} \pi_i = \pi_0 \cdot \sum_{i=0}^{\infty} \rho^i = \pi_0 \cdot \frac{1}{1 - \rho} \Rightarrow \pi_0 = 1 - \rho.$$

Dabei haben wir angenommen, dass  $\rho < 1$  ist. Für  $\rho \geq 1$  konvergiert das System nicht. Da in diesem Fall  $\lambda \geq \mu$  gilt, kommen die Jobs schneller an, als sie abgearbeitet werden können. Intuitiv folgt daraus, dass die Warteschlange immer größer wird.

Für  $\rho < 1$  erhalten wir als Endergebnis

$$\pi_k = (1 - \rho)\rho^k \text{ für alle } k \in \mathbb{N}_0.$$

Aus diesem Resultat können wir einige interessante Schlussfolgerungen ziehen. Zunächst betrachten wir die Zufallsvariable

$N :=$  Anzahl der Jobs im System (wartend + in Bearbeitung).

Für  $N$  gilt (die Berechnung von  $\mathbb{E}[N]$  und  $\text{Var}[N]$  erfolgt mit den schon bei der geometrischen Verteilung in Abschnitt 3 verwendeten Summenformeln)

$$\mathbb{E}[N] = \sum_{k \geq 0} k \cdot \pi_k = \frac{\rho}{1 - \rho} \quad \text{und} \quad \text{Var}[N] = \frac{\rho}{(1 - \rho)^2}. \quad (15)$$



Abbildung 4 zeigt  $\mathbb{E}[N]$  als Funktion von  $\rho$ . Man erkennt, wie das System für  $\rho \rightarrow 1$  divergiert.

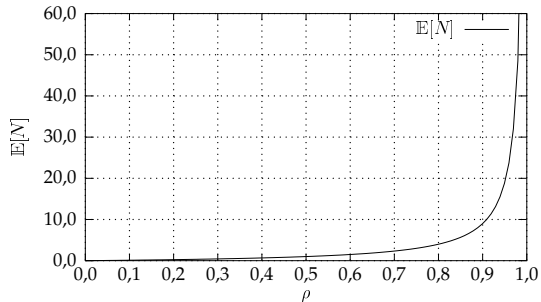


Abbildung: Mittlere Anzahl der Jobs in einer M/M/1-Warteschlange

Für eine weitergehende Analyse der Leistung des Systems definieren wir für den  $i$ -ten Job (bezüglich der Reihenfolge, mit der die Jobs im System ankommen):

$$R_i := \text{Antwortzeit (Gesamtverweildauer im System)}.$$

Der Wert von  $R_i$  hängt natürlich vom Zustand des Systems zur Ankunftszeit des Jobs ab. Betrachten wir das System jedoch im Gleichgewichtszustand, so können wir den Index  $i$  auch weglassen und einfach von der Antwortzeit  $R$  sprechen.

Bei der Berechnung von  $R$  hilft uns der folgende Satz.

### Theorem 156

*(Formel von Little) Für Warteschlangen-Systeme mit mittlerer Ankunftsrate  $\lambda$ , bei denen die Erwartungswerte  $\mathbb{E}[N]$  und  $\mathbb{E}[R]$  existieren, gilt*

$$\mathbb{E}[N] = \lambda \cdot \mathbb{E}[R].$$

*Hierbei werden keine weiteren Annahmen über die Verteilung der Ankunfts- und Bearbeitungszeiten getroffen.*

## Beweis:

[(Skizze)]Wir beobachten das System über einen (langen) Zeitraum (siehe Abbildung 5). In einer Zeitspanne der Länge  $t_0$  seien  $n(t_0)$  Anforderungen eingetroffen.  $N(t)$  gibt die Anzahl der Jobs an, die sich zum Zeitpunkt  $t$  im System befinden. Nun betrachten wir die beiden Größen

$$\sum_{i=1}^{n(t_0)} R_i \quad \text{und} \quad \int_0^{t_0} N(t) dt.$$

Beide Größen messen „ungefähr“ die in Abbildung 5 grau gefärbte Fläche.

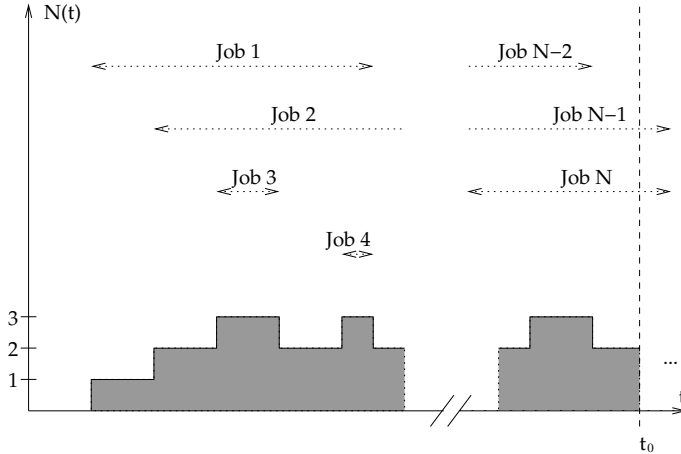


Abbildung: Graphik zum Beweis des Satzes von Little

## Beweis (Forts.):

Die rechte Größe misst sogar genau diese Fläche, bei der Summe wird hingegen bei den Jobs, die zur Zeit  $t_0$  noch im System sind, die gesamte Aufenthaltsdauer gezählt, statt nur der Anteil bis zum Zeitpunkt  $t_0$ . Für große  $t_0$  ist der Unterschied dieser beiden Größen aber vernachlässigbar. Führt man daher den Grenzübergang  $t_0 \rightarrow \infty$  durch und normiert beide Größen mit  $1/n(t_0)$ , erhält man

$$\begin{aligned} \lim_{t_0 \rightarrow \infty} \frac{1}{n(t_0)} \sum_{i=1}^{n(t_0)} R_i &= \lim_{t_0 \rightarrow \infty} \frac{1}{n(t_0)} \int_0^{t_0} N(t) \, dt \\ &= \lim_{t_0 \rightarrow \infty} \frac{t_0}{n(t_0)} \cdot \frac{1}{t_0} \int_0^{t_0} N(t) \, dt. \end{aligned}$$

## Beweis (Forts.):

Mit

$$\bar{R}(t_0) := \frac{1}{n(t_0)} \sum_{i=1}^{n(t_0)} R_i, \quad \bar{N}(t_0) := \frac{1}{t_0} \int_0^{t_0} N(t) \, dt$$

und  $\bar{\lambda}(t_0) := \frac{n(t_0)}{t_0}$  erhalten wir daraus wegen

$$\lambda = \lim_{t_0 \rightarrow \infty} \bar{\lambda}(t_0) = \lim_{t_0 \rightarrow \infty} \frac{n(t_0)}{t_0},$$

$$\mathbb{E}[R] = \lim_{t_0 \rightarrow \infty} \bar{R}(t_0) = \lim_{t_0 \rightarrow \infty} \frac{1}{n(t_0)} \sum_{i=1}^{n(t_0)} R_i \quad \text{und}$$

$$\mathbb{E}[N] = \lim_{t_0 \rightarrow \infty} \bar{N}(t_0) = \lim_{t_0 \rightarrow \infty} \frac{1}{t_0} \int_0^{t_0} N(t) \, dt$$

sofort die Behauptung. □

Bei der Berechnung von  $\mathbb{E}[R]$  haben wir verwendet, dass sich für lange Beobachtungszeiträume die relative Häufigkeit immer mehr dem Erwartungswert annähert. Man vergleiche dies mit dem Gesetz der großen Zahlen, Satz 63. Bei den Zufallsvariablen  $R_i$  ist allerdings die Unabhängigkeit nicht gesichert und ein formal korrekter Beweis von  $\mathbb{E}[R] = \lim_{t_0 \rightarrow \infty} \bar{R}(t_0)$  würde deshalb aufwändiger.  $\mathbb{E}[N] = \lim_{t_0 \rightarrow \infty} \bar{N}(t_0)$  gilt aufgrund ähnlicher Überlegungen. Die obige Argumentation ist zweifellos ein wenig informell, sie sollte jedoch ausreichen, um die Hintergründe des Satzes zu verdeutlichen.

Mit Satz 156 ist die Berechnung von  $\mathbb{E}[R]$  für die Markov-Kette aus Abbildung 3 kein Problem mehr. Aus (15) folgt

$$\mathbb{E}[R] = \frac{\mathbb{E}[N]}{\lambda} = \frac{\rho}{\lambda(1 - \rho)}. \quad (16)$$

Manchmal sieht man statt  $R$  auch die leicht abgewandelte Größe

$$W := (\text{reine}) \text{Wartezeit.}$$

Wegen der Linearität des Erwartungswerts ist die Berechnung von  $\mathbb{E}[W]$  für M/M/1-Warteschlangen kein Problem:

$$\mathbb{E}[W] = \mathbb{E}[R] - \frac{1}{\mu} = \frac{\rho}{\mu(1 - \rho)}. \quad (17)$$