

Stochastische Diskrete Ereignissysteme



Roger Wattenhofer

4. Stochastische diskrete Ereignissysteme

- 4.1 Grundbegriffe der Wahrscheinlichkeitsrechnung
- 4.2 Stochastische Prozesse in diskreter Zeit
 - ▶ Markov-Ketten in diskreter Zeit, stationäre Verteilung
 - ▶ Random Walk und PageRank
- 4.3 Stochastische Prozesse in kontinuierlicher Zeit
 - ▶ Markov-Ketten in kontinuierlicher Zeit
 - ▶ Warteschlangen

4/2

Literatur zu Kapitel 4 der Vorlesung

- ▶ Schickinger, Steger: **Diskrete Strukturen. Band 2: Wahrscheinlichkeitstheorie und Statistik.** Springer, Berlin, 2001.
[Kapitel 1–2: Grundlagen, Kapitel 4: Stochastische Prozesse]
- ▶ Bertsekas, Gallager: **Data Networks.** Second Edition. Prentice Hall, Upper Saddle River, NJ, 1992.
[Chapter 3: Delay Models in Data Networks]

4/3

Weitere Literatur

- ▶ Kleinrock: **Queueing Systems, Volume 1: Theory,** John Wiley & Sons, 1975.
- ▶ Kleinrock: **Queueing Systems, Volume 2: Computer Applications,** John Wiley & Sons, 1976.
- ▶ Gross, Harris: **Fundamentals of Queueing Theory,** Wiley, 1998.
- ▶ Tanner: **Practical Queueing Analysis,** McGraw-Hill, 1995.
- ▶ Nelson: **Probability, Stochastic Processes, and Queueing Theory,** Springer, 1995.
- ▶ ...

4/4

- ▶ **Bisher:** Das Verhalten von DES war deterministisch.
- ▶ **Jetzt:** Einbeziehung von "Unsicherheit" auf der Grundlage stochastischer Prozesse. Unsicherheit bezüglich der Funktion und des Zeitverhaltens.
- ▶ **Beispiele:**
 - ▶ Modellierung nur statistisch beschreibbarer Ereignisse, wie Telefonanrufe, Berechnungszeiten von Tasks.
 - ▶ Quantitative Analyse von Verkehrsprozessen (Zahl von Anrufen/Zeit, ...), Warteschlangen, Computernetzwerken, Rechnerarchitekturen.

4/5

- ▶ Menge Ω von Elementarereignissen
- ▶ $\Pr[\omega]$ bezeichnet die Wahrscheinlichkeit von $\omega \in \Omega$
- ▶ Es muss gelten: $0 \leq \Pr[\omega] \leq 1$ und $\sum_{\omega \in \Omega} \Pr[\omega] = 1$.
- ▶ **Wahrscheinlichkeitsraum:** Ω mit $\Pr[\omega]$ für alle $\omega \in \Omega$
- ▶ **diskret**, falls Ω endlich oder abzählbar, sonst **kontinuierlich**.
- ▶ **Ereignis:** Teilmenge von Ω
- ▶ Wahrscheinlichkeit von $E \subseteq \Omega$: $\Pr[E] = \sum_{\omega \in E} \Pr[\omega]$
- ▶ Ereignisse A und B heißen **unabhängig**, falls $\Pr[A \cap B] = \Pr[A] \cdot \Pr[B]$.

4/6

Beispiel 1

Zufallsexperiment: Werfen eines Würfels mit 6 Seiten

- ▶ $\Omega = \{\mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}, \mathbf{5}, \mathbf{6}\}$
- ▶ $\Pr[\mathbf{1}] = \Pr[\mathbf{2}] = \dots = \Pr[\mathbf{6}] = \frac{1}{6}$
- ▶ $E = \text{"gerade Zahl"} = \{\mathbf{2}, \mathbf{4}, \mathbf{6}\} \subseteq \Omega$
- ▶ $\Pr[E] = \Pr[\mathbf{2}] + \Pr[\mathbf{4}] + \Pr[\mathbf{6}] = \frac{1}{2}$
- ▶ $F = \text{"durch 3 teilbare Zahl"} = \{\mathbf{3}, \mathbf{6}\} \subseteq \Omega$
- ▶ $\Pr[F] = \Pr[\mathbf{3}] + \Pr[\mathbf{6}] = \frac{1}{3}$
- ▶ E und F sind unabhängig, da $\Pr[E \cap F] = \Pr[\mathbf{6}] = \frac{1}{6} = \Pr[E] \cdot \Pr[F]$.

4/7

Beispiel 2

Zufallsexperiment: Paketübertragung

- ▶ Jeder Übertragungsversuch gelingt mit W'keit p .
- ▶ Elementarereignis ω_i : Es braucht i Versuche bis zur ersten erfolgreichen Übertragung.
- ▶ $\Omega = \{\omega_1, \omega_2, \dots\}$ abzählbar unendlich
- ▶ $\Pr[\omega_1] = p, \Pr[\omega_2] = (1-p)p, \Pr[\omega_3] = (1-p)^2 p$
- ▶ Allgemein: $\Pr[\omega_i] = (1-p)^{i-1} p$
- ▶ Es gilt: $\sum_{i=1}^{\infty} (1-p)^{i-1} p = p \sum_{i=0}^{\infty} (1-p)^i = p \frac{1}{1-(1-p)} = 1$

4/8

Rechnen mit Wahrscheinlichkeiten

Seien $A, B \subseteq \Omega$ Ereignisse. Es gilt:

- ▶ $\Pr[\emptyset] = 0, \Pr[\Omega] = 1$
- ▶ $\Pr[\bar{A}] = 1 - \Pr[A]$, wobei $\bar{A} := \Omega \setminus A$
- ▶ $\Pr[A \cup B] \leq \Pr[A] + \Pr[B]$
- ▶ $\Pr[A \cup B] = \Pr[A] + \Pr[B] - \Pr[A \cap B]$

4/9

Bedingte Wahrscheinlichkeiten

Seien A, B Ereignisse mit $\Pr[B] > 0$.

Die **bedingte Wahrscheinlichkeit** von A gegeben B ist definiert durch:

$$\Pr[A | B] := \frac{\Pr[A \cap B]}{\Pr[B]}$$

Multiplikationssatz. Für Ereignisse A_1, \dots, A_n mit $\Pr[A_1 \cap \dots \cap A_n] > 0$ gilt:

$$\Pr[A_1 \cap \dots \cap A_n] = \Pr[A_1] \cdot \Pr[A_2 | A_1] \cdot \Pr[A_3 | A_1 \cap A_2] \cdot \dots \cdot \Pr[A_n | A_1 \cap \dots \cap A_{n-1}]$$

4/10

Totale Wahrscheinlichkeit

Satz (Satz der totalen Wahrscheinlichkeit)

Seien A_1, \dots, A_n paarweise disjunkte Ereignisse und $B \subseteq A_1 \cup \dots \cup A_n$. Dann folgt:

$$\Pr[B] = \sum_{i=1}^n \Pr[B | A_i] \cdot \Pr[A_i]$$

Bemerkung: Der Satz gilt analog für unendlich viele paarweise disjunkte Ereignisse A_1, A_2, \dots :

$$\Pr[B] = \sum_{i=1}^{\infty} \Pr[B | A_i] \cdot \Pr[A_i]$$

4/11

Zufallsvariablen

- ▶ Eine Abbildung $X : \Omega \rightarrow \mathbb{R}$ heisst **Zufallsvariable**. Wir schreiben W_X als Abkürzung für den Wertebereich $X(\Omega)$.
- ▶ Falls Ω diskret (endlich oder abzählbar unendlich) ist, heisst auch X diskret. Wir betrachten vorerst nur diskrete Zufallsvariablen.
- ▶ Die Funktionen $f_X : \mathbb{R} \rightarrow [0, 1]$ und $F_X : \mathbb{R} \rightarrow [0, 1]$ mit $f_X(x) = \Pr[X = x]$ und $F_X(x) = \Pr[X \leq x]$ heissen **Dichte(funktion)** und **Verteilung(sfunktion)** von X .
- ▶ **Erwartungswert:** $\mathbf{E}[X] := \sum_{x \in W_X} x \cdot \Pr[X = x]$ (falls die Summe konvergiert)
- ▶ **Varianz:** $\mathbf{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$ (falls $\mathbf{E}[X^2]$ und $\mathbf{E}[X]$ existieren)
- ▶ **Standardabweichung:** $\sigma(X) := \sqrt{\mathbf{Var}[X]}$

4/12

Rechnen mit Erwartungswert und Varianz

Mit $a, b \in \mathbb{R}$ gilt für die transformierte Zufallsvariable $a \cdot X + b$:

- ▶ $\mathbf{E}[a \cdot X + b] = a \cdot \mathbf{E}[X] + b$
- ▶ $\text{Var}[a \cdot X + b] = a^2 \cdot \text{Var}[X]$.

Satz (Linearität des Erwartungswertes)

Für Zufallsvariablen X_1, X_2, \dots, X_n und $X := a_1 X_1 + \dots + a_n X_n$ mit $a_1, \dots, a_n \in \mathbb{R}$ gilt:

$$\mathbf{E}[X] = a_1 \mathbf{E}[X_1] + \dots + a_n \mathbf{E}[X_n]$$

4/13

Unabhängigkeit von Zufallsvariablen

- ▶ Zufallsvariablen X_1, \dots, X_n heißen **unabhängig**, wenn für alle $(x_1, \dots, x_n) \in W_{X_1} \times \dots \times W_{X_n}$ gilt:

$$\Pr[X_1 = x_1, \dots, X_n = x_n] = \Pr[X_1 = x_1] \cdot \dots \cdot \Pr[X_n = x_n]$$

- ▶ Für unabhängige Zufallsvariablen X und Y gilt:

$$\mathbf{E}[X \cdot Y] = \mathbf{E}[X] \cdot \mathbf{E}[Y]$$

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

4/14

Beispiele diskreter Verteilungen (1)

- ▶ **Bernoulli-Verteilung** mit Erfolgsw'keit p :

$$\Pr[X = 1] = p, \quad \Pr[X = 0] = 1 - p$$

Es gilt $\mathbf{E}[X] = p$ und $\text{Var}[X] = p(1 - p)$.

- ▶ **Binomial-Verteilung** mit Parametern n und p :

$$\Pr[X = k] = \binom{n}{k} p^k (1 - p)^{n-k}, \quad \text{für } 0 \leq k \leq n$$

Es gilt $\mathbf{E}[X] = np$ und $\text{Var}[X] = np(1 - p)$.

- ▶ **Poisson-Verteilung** mit Parameter λ :

$$\Pr[X = k] = \frac{e^{-\lambda} \lambda^k}{k!}, \quad \text{für } k \in \mathbb{N}_0$$

Es gilt $\mathbf{E}[X] = \lambda$ und $\text{Var}[X] = \lambda$.

4/15

Beispiele diskreter Verteilungen (2)

- ▶ **Geometrische Verteilung** mit Parameter p :

$$\Pr[X = k] = (1 - p)^{k-1} p \quad \text{für } k \in \mathbb{N}$$

Der Erwartungswert ist

$$\mathbf{E}[X] = \sum_{k=1}^{\infty} k(1 - p)^{k-1} p = p \sum_{k=1}^{\infty} k(1 - p)^{k-1} = p \cdot \frac{1}{p^2} = \frac{1}{p}$$

Die Varianz ist $\frac{1-p}{p^2}$.

Anwendungsbeispiel: Paketübertragung mit Erfolgsw'keit p
 \Rightarrow Es sind im Mittel $\frac{1}{p}$ Versuche nötig, bis ein Paket erfolgreich übertragen werden kann.

4/16

Wichtige Wahrscheinlichkeitsschranken

- ▶ **Markov-Ungleichung:**

$$\Pr[|X| \geq k] \leq \frac{\mathbf{E}[|X|]}{k}$$

- ▶ **Chebyshev-Ungleichung:**

$$\Pr[|X - \mathbf{E}[X]| \geq k] \leq \frac{\text{Var}[X]}{k^2}$$

- ▶ **Chernoff-Ungleichung** für X_1, \dots, X_n Indikatorvariablen unabhängiger identischer Bernoulli-Experimente und $X = \sum X_j$:

$$\Pr[X \geq (1 + \delta) \cdot \mathbf{E}[X]] \leq \left(\frac{e^\delta}{(1 + \delta)^{(1 + \delta)}} \right)^{\mathbf{E}[X]}$$

$$\Pr[X \leq (1 - \delta) \cdot \mathbf{E}[X]] \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{(1 - \delta)}} \right)^{\mathbf{E}[X]}$$

4/17

4.2 Stochastische Prozesse in diskreter Zeit

- ▶ dynamisches System \sim zeitliche Folge von Zufallsexperimenten
- ▶ Zustand und Verhalten des Systems zur Zeit t wird als Zufallsvariable X_t modelliert. Wir betrachten nur Prozesse mit diskreten Zufallsvariablen X_t (zustandsdiskret).
- ▶ stochastischer Prozess: Folge von Zufallsvariablen $(X_t)_{t \in T}$
 - ▶ in diskreter Zeit: $T = \mathbb{N}_0$
 - ▶ in kontinuierlicher Zeit: $T = \mathbb{R}_0^+$
- ▶ Zufallsvariablen X_{t_1} und X_{t_2} können abhängig sein.
- ▶ **Markov-Prozesse:** Weiterer Ablauf ist nur vom aktuellen Zustand abhängig, nicht von der Vergangenheit.

4/18

Beispiel: Paketübertragung (1)

- ▶ Paketübertragung von Rechner A zu Rechner B.
- ▶ Jede Sekunde wird ein Paket übertragen.
- ▶ Zufallsvariablen X_t für $t \in \mathbb{N}_0$:

$$X_t = \begin{cases} 1 & \text{falls Übertragung zur Zeit } t \text{ erfolgreich} \\ 0 & \text{sonst} \end{cases}$$

- ▶ **Annahme:** Wahrscheinlichkeit für erfolgreiche Übertragung zur Zeit t ist nur abhängig vom Erfolg der Übertragung zur Zeit $t - 1$.

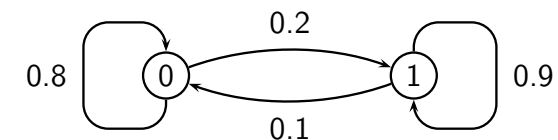
4/19

Beispiel: Paketübertragung (2)

- ▶ W'keit für Erfolg der Übertragung zur Zeit $t + 1$:

$$\begin{aligned} \Pr[X_{t+1} = 0 \mid X_t = 0] &= 0.8 & \Pr[X_{t+1} = 0 \mid X_t = 1] &= 0.1 \\ \Pr[X_{t+1} = 1 \mid X_t = 0] &= 0.2 & \Pr[X_{t+1} = 1 \mid X_t = 1] &= 0.9 \end{aligned}$$

- ▶ Graphische Veranschaulichung durch **Übergangsdiagramm**:



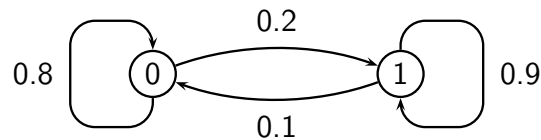
Kante von a nach b wird mit $\Pr[X_{t+1} = b \mid X_t = a]$ beschriftet.

Ablauf des Systems: **Random Walk** im Übergangsdiagramm.

4/20

Beispiel: Paketübertragung (3)

- ▶ Übergangsdiagramm



- ▶ Alternative Darstellung: **Übergangsmatrix**

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$$

Eintrag p_{ij} entspricht $\Pr[X_{t+1} = j \mid X_t = i]$.

4/21

Beispiel: Paketübertragung (4)

Entwicklung des Systems über die Zeit:

t	$\Pr[X_t = 0]$	$\Pr[X_t = 1]$	
0	0	1	Anfangszustand (vorgegeben)
1	0.1	0.9	
2	0.17	0.83	
3	0.219	0.781	$\Pr[X_3 = 1] = 0.17 \cdot 0.2 + 0.83 \cdot 0.9$
4	0.253	0.747	$= 0.781$
	\vdots		
1000	0.333	0.667	

4/22

Definition: Markov-Kette

Endliche Markov-Kette in diskreter Zeit

über der Zustandsmenge $S = \{0, 1, \dots, n-1\}$:

- ▶ **Folge von Zufallsvariablen** $(X_t)_{t \in \mathbb{N}_0}$ mit Wertemenge S
- ▶ **Startverteilung** $q_0 = (q_{00}, q_{01}, \dots, q_{0,n-1})$ mit $q_{0,i} \geq 0$ und $\sum_{i=0}^{n-1} q_{0,i} = 1$.
- ▶ **X_{t+1} hängt nur von X_t ab**, d.h. für alle $t > 0$ und alle $I \subseteq \{0, 1, \dots, t-1\}$ und $i, j, s_k \in S$ (für alle $k \in I$) gilt:

$$\Pr[X_{t+1} = j \mid X_t = i, \forall k \in I: X_k = s_k] = \Pr[X_{t+1} = j \mid X_t = i]$$

(Falls $S = \mathbb{N}_0$, dann **unendliche Markov-Kette in diskreter Zeit**.)

4/23

Zeithomogene Markov-Ketten

- ▶ Falls $\Pr[X_{t+1} = j \mid X_t = i]$ für alle $i, j \in S$ unabhängig von t ist, so heisst die Markov-Kette **(zeit)homogen**.
- ▶ Wir betrachten (fast) nur zeithomogene Markov-Ketten.
- ▶ Für zeithomogene Markov-Ketten sind die Werte

$$p_{ij} := \Pr[X_{t+1} = j \mid X_t = i]$$

eindeutig definiert und ergeben die Übergangsmatrix

$$P = (p_{i,j})_{0 \leq i,j < n}$$

4/24

Ablauf einer Markov-Kette

- ▶ Beobachtung einer Markov-Kette von Zeit 0 bis Zeit t_0 .
- ▶ Möglicher Ablauf: Zustände $x_0, x_1, x_2, \dots, x_{t_0}$.
- ▶ Wahrscheinlichkeit für diesen Ablauf (Musterpfad):

$$q_{0,x_0} \cdot \Pr[X_1 = x_1 | X_0 = x_0] \cdot \dots \cdot \Pr[X_{t_0} = x_{t_0} | X_{t_0-1} = x_{t_0-1}]$$

Beispiel

Paketübertragung mit $P = \begin{pmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}$ und $q_0 = (0.5, 0.5)$:

⇒ Wahrscheinlichkeit für Ablauf $(1, 1, 0, 0, 1, 0)$ ist:

$$0.5 \cdot 0.9 \cdot 0.1 \cdot 0.8 \cdot 0.2 \cdot 0.1 = 0.00072$$

4/25

Verweildauer

- ▶ Betrachte eine Markov-Kette, die zur Zeit t im Zustand i ist.
- ▶ Modelliere die Anzahl der Zeitschritte, die die Kette ab Zeit t im Zustand i bleibt, als Zufallsvariable V_i .
- ▶ Es gilt: $\Pr[V_i = k] = p_{ii}^{k-1}(1 - p_{ii})$
und $\Pr[V_i > k] = p_{ii}^k$.
- ▶ V_i ist also **geometrisch verteilt**.
- ▶ **Beachte:** Die Verweildauer ist unabhängig davon, wie lange die Kette schon im Zustand i war.

4/26

Rechnen mit der Übergangsmatrix (1)

- ▶ Startverteilung: q_0 (Zeilenvektor mit n Elementen)
- ▶ Wahrscheinlichkeitsverteilung auf den Zuständen zur Zeit t :
 $q_t = (q_{t,0}, \dots, q_{t,n-1})$ mit $q_{t,i} = \Pr[X_t = i]$
- ▶ Berechnung von q_{t+1} aus q_t :

$$q_{t+1,j} = \Pr[X_{t+1} = j] \\ = \sum_{i=0}^{n-1} \Pr[X_t = i] \cdot \Pr[X_{t+1} = j | X_t = i] = \sum_{i=0}^{n-1} q_{t,i} \cdot p_{ij}$$

Geschrieben als Multiplikation Vektor mit Matrix:

$$q_{t+1} = q_t \cdot P$$

4/27

Rechnen mit der Übergangsmatrix (2)

- ▶ Wir wissen also: $q_{t+1} = q_t \cdot P$.
- ▶ Dann muss gelten: $q_1 = q_0 \cdot P$
 $q_2 = q_1 \cdot P = q_0 \cdot P \cdot P = q_0 \cdot P^2$
 $q_3 = q_2 \cdot P = q_0 \cdot P^2 \cdot P = q_0 \cdot P^3$
 \vdots
 $q_t = q_0 \cdot P^t$

- ▶ Ebenso: $q_{t+k} = q_t \cdot P^k$ für alle $k \geq 0$
- ▶ Der Eintrag in Zeile i und Spalte j von P^k , bezeichnet mit $p_{ij}^{(k)} := (P^k)_{ij}$, gibt die Wahrscheinlichkeit an, in k Schritten von Zustand i nach Zustand j zu gelangen.

4/28

Chapman-Kolmogorov-Gleichungen

Kurze Exkursion: Zeitinhomogene Ketten

- ▶ $p_{ij}(k, n) := \Pr[X_n = j \mid X_k = i]$, $P(k, n) := (p_{ij}(k, n))_{i,j \in S}$.
- ▶ Betrachte Zeitpunkt u mit $k < u < n$:

$$\begin{aligned} p_{ij}(k, n) &= \Pr[X_n = j \mid X_k = i] \\ &= \sum_{r \in S} \Pr[X_n = j \mid X_u = r, X_k = i] \cdot \Pr[X_u = r \mid X_k = i] \\ &= \sum_{r \in S} \Pr[X_n = j \mid X_u = r] \cdot \Pr[X_u = r \mid X_k = i] \\ &= \sum_{r \in S} p_{rj}(u, n) \cdot p_{ir}(k, u) \end{aligned}$$

- ▶ $P(k, n) = P(k, u) \cdot P(u, n)$ für $k < u < n$.

4/29

Transientenanalyse

Typische Fragen:

- ▶ Wie gross ist die W'keit, nach k Schritten im Zustand j zu sein?
- ▶ Wie wahrscheinlich ist es, irgendwann von i nach j zu kommen?
- ▶ Wie viele Schritte benötigt die Kette im Mittel, um von i nach j zu gelangen?

Viele dieser Fragen können mit Hilfe der Gleichung

$$q_{t+k} = q_t \cdot P^k \text{ für alle } k \geq 0$$

beantwortet werden!

4/30

Beispiel (1)

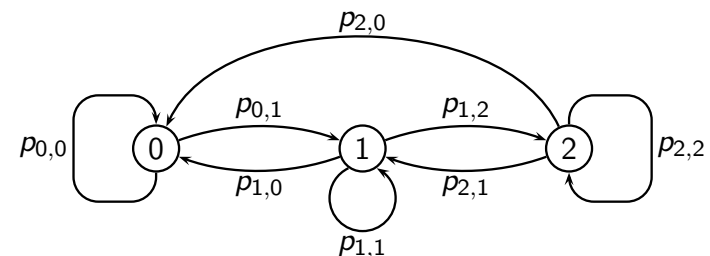
Rechnersystem aus 2 Computern

- ▶ In jedem Zeitschritt wird dem System höchstens ein Task übergeben. Dieses Ereignis tritt mit W'keit a auf.
- ▶ Der ankommende Task wird nur bearbeitet, wenn mindestens einer der beiden Prozessoren frei ist oder im selben Zeitschritt frei wird.
- ▶ Falls ein Prozessor belegt ist, beendet er den Task in jedem Zeitschritt mit W'keit b .

4/31

Beispiel (2)

Modellierung als Markov-Kette mit Zustandsmenge $S = \{0, 1, 2\}$:



Übergangsmatrix:

$$P = \begin{pmatrix} 1-a & a & 0 \\ b(1-a) & (1-a)(1-b) + ab & a(1-b) \\ b^2(1-a) & b^2a + 2b(1-b)(1-a) & (1-b)^2 + 2b(1-b)a \end{pmatrix}$$

4/32

Beispiel (3)

Für $a = 0.5$ und $b = 0.7$ ergibt sich:

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{pmatrix}$$

Sei $q_0 = (1, 0, 0)$.

Fragen und Antworten:

- ▶ **W'keit, dass System zur Zeit 3 leer ist?**

$$q_3 = (1, 0, 0) \cdot P^3 = (0.405875, 0.496625, 0.0975)$$

$$\Rightarrow \Pr[X_3 = 0] = 0.405875$$

4/33

Beispiel (4)

- ▶ **W'keit, dass System zur Zeit 2 und zur Zeit 3 leer ist?**

$$\begin{aligned} \Pr[X_3 = 0, X_2 = 0] &= \\ &= \Pr[X_2 = 0] \cdot \Pr[X_3 = 0 \mid X_2 = 0] \\ &= ((1, 0, 0) \cdot P^2)_0 \cdot p_{0,0} = 0.425 \cdot 0.5 = 0.2125 \end{aligned}$$

- ▶ **W'keit, dass zwischen Zeit 3 und 4 kein Task beendet wird?**

$$\begin{aligned} \Pr[\text{"keiner fertig zwischen 3 und 4"}] &= \\ &= \sum_{j=0}^2 \Pr[\text{"keiner fertig zwischen 3 und 4"} \mid X_3 = j] \cdot q_{3,j} \\ &= 1 \cdot q_{3,0} + (1 - b) \cdot q_{3,1} + (1 - b)^2 \cdot q_{3,2} \\ &= 1 \cdot 0.405875 + 0.3 \cdot 0.496625 + 0.09 \cdot 0.0975 \approx 0.564 \end{aligned}$$

4/34

Definitionen

- ▶ **Übergangszeit** T_{ij} von Zustand i nach j

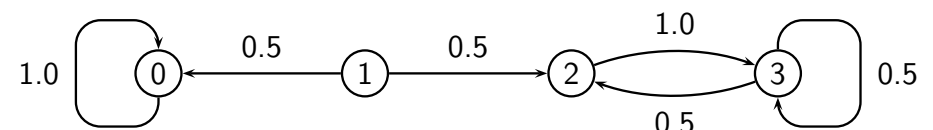
$$T_{ij} := \min\{n \geq 1 \mid X_n = j, \text{ wenn } X_0 = i\}$$

(falls Zustand j nie erreicht wird, setze $T_{ij} = \infty$)

- ▶ $h_{ij} := \mathbf{E}[T_{ij}]$ ist die **Hitting Time** von i nach j
- ▶ $f_{ij} := \Pr[T_{ij} < \infty]$ ist die **Ankunftswahrscheinlichkeit** von i nach j .
- ▶ $c_{ij} := h_{ij} + h_{ji}$ ist die **Commutate Time** zwischen i und j

4/35

Beispiel



- ▶ $T_{01} = T_{02} = T_{03} = \infty$
- ▶ T_{10} ist 1, falls $X_1 = 0$, und ∞ , falls $X_1 = 2$
 $\Rightarrow f_{10} = 0.5$ und $h_{10} = \mathbf{E}[T_{10}] = 0.5 \cdot 1 + 0.5 \cdot \infty = \infty$
- ▶ $h_{32} = 0.5 \cdot 1 + 0.5^2 \cdot 2 + 0.5^3 \cdot 3 + \dots = 0.5 \cdot \sum_{i=1}^{\infty} 0.5^{i-1} \cdot i = 0.5 \cdot \frac{1}{(1-0.5)^2} = 2$.

4/36

Berechnung der Hitting Time und Ankunfts-w'keit

Lemma

Für die **Hitting Time** gilt für alle $i, j \in S$

$$h_{ij} = 1 + \sum_{k:k \neq j} p_{ik} h_{kj},$$

falls h_{ij} und h_{kj} existieren.

Für die **Ankunftswahrscheinlichkeiten** gilt analog

$$f_{ij} = p_{ij} + \sum_{k:k \neq j} p_{ik} f_{kj}.$$

4/37

Beweis

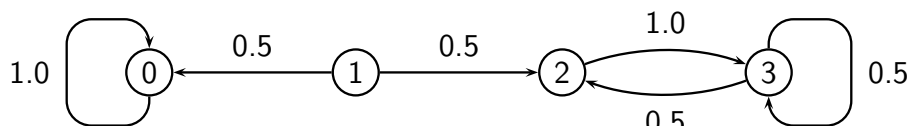
Beweis. (nur für $h_{ij} = 1 + \sum_{k:k \neq j} p_{ik} h_{kj}$)

$$\begin{aligned} h_{ij} &= \mathbf{E}[T_{ij}] = \sum_{k \in S} \mathbf{E}[T_{ij} \mid X_1 = k] \cdot p_{ik} \\ &= \mathbf{E}[T_{ij} \mid X_1 = j] \cdot p_{ij} + \sum_{k:k \neq j} \mathbf{E}[T_{ij} \mid X_1 = k] \cdot p_{ik} \\ &= 1 \cdot p_{ij} + \sum_{k:k \neq j} (1 + \mathbf{E}[T_{kj}]) \cdot p_{ik} \\ &= 1 + \sum_{k \neq j} \mathbf{E}[T_{kj}] \cdot p_{ik} = 1 + \sum_{k:k \neq j} p_{ik} h_{kj} \end{aligned}$$

□

4/38

Anwendung auf das Beispiel



Wende $h_{ij} = 1 + \sum_{k:k \neq j} p_{ik} h_{kj}$ auf $i, j \in \{2, 3\}$ an:

$$h_{22} = 1 + h_{32} \quad h_{32} = 1 + 0.5 \cdot h_{32}$$

$$h_{23} = 1 \quad h_{33} = 1 + 0.5 \cdot h_{23}$$

Lösen des Gleichungssystems liefert:

$$h_{23} = 1, \quad h_{33} = 1.5, \quad h_{32} = 2 \quad \text{und} \quad h_{22} = 3.$$

(Analog: $f_{22} = f_{33} = f_{23} = f_{32} = 1$)

4/39

Gambler's Ruin

- ▶ Zwei Spieler A und B spielen ein Spiel um m Franken.
- ▶ Spieler A hat a Franken, Spieler B hat $b = m - a$ Franken.
- ▶ In jeder Runde wird um 1 Franken gespielt. A gewinnt eine Runde mit W'keit p , B mit W'keit $q = 1 - p$.
- ▶ Wer gewinnt mit welcher Wahrscheinlichkeit?
- ▶ Wie lange dauert das Spiel?
- ▶ Was ändert sich, wenn B unendlich viel Geld hat? („ B ist die Bank“)

4/40

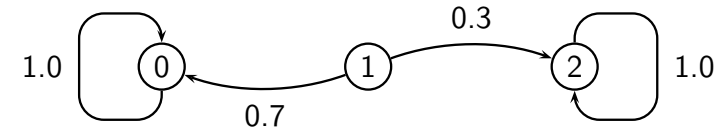
Stationäre Analyse

- ▶ Reale dynamische Systeme laufen oft über eine lange Zeit.
- ▶ Betrachte Verhalten für $t \rightarrow \infty$.
- ▶ Wahrscheinlichkeitsverteilung auf den Zuständen der Markov-Kette zur Zeit t ist $q_t = q_0 \cdot P^t$. **Konvergenz?**
- ▶ Intuitiv klar: Falls q_t für $t \rightarrow \infty$ gegen einen Vektor π konvergiert, so sollte π die Gleichung $\pi = \pi \cdot P$ erfüllen.
- ▶ **Definition.** Ein Zustandsvektor π mit $\pi_j \geq 0$ und $\sum_{j \in S} \pi_j = 1$ heisst **stationäre Verteilung** der Markov-Kette mit Übergangsmatrix P , falls $\pi = \pi \cdot P$.
- ▶ Die stationäre Verteilung π ist ein Eigenvektor von P zum Eigenwert 1.

4/41

(Nicht-)Eindeutigkeit der stat. Verteilung

- ▶ Übergangsdiagramm einer Beispiel-Markov-Kette:



- ▶ Übergangsmatrix: $P = \begin{pmatrix} 1 & 0 & 0 \\ 0.7 & 0 & 0.3 \\ 0 & 0 & 1 \end{pmatrix}$
- ▶ Diese Markov-Kette besitzt mehrere stationäre Verteilungen: zum Beispiel $(1, 0, 0)$ und $(0, 0, 1)$ und $(0.5, 0, 0.5)$
- ▶ Ursache: Zustände 0 und 2 sind "absorbierend".

4/42

Irreduzible Markov-Ketten

Definition. Eine Markov-Kette heisst **irreduzibel**, wenn es für alle Zustände $i, j \in S$ eine Zahl $n \in \mathbb{N}$ gibt, so dass $p_{ij}^{(n)} > 0$.

Satz

Eine irreduzible endliche Markov-Kette besitzt eine **eindeutige stationäre Verteilung** π und es gilt $\pi_j = 1/h_{jj}$ für alle $j \in S$.

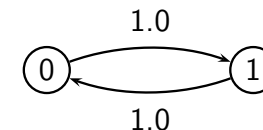
Intuition. Im Mittel wird Zustand j alle h_{jj} Schritte besucht.

Frage: Konvergiert eine irreduzible endliche Markov-Kette immer gegen ihre stationäre Verteilung?

4/43

Konvergenz?

Frage: Konvergiert eine irreduzible endliche Markov-Kette immer gegen ihre stationäre Verteilung? **NEIN!**



Diese Kette ist irreduzibel und endlich, aber der Zustandsvektor q_t konvergiert nicht unbedingt für $t \rightarrow \infty$:

$q_0 = (1, 0)$, $q_1 = (0, 1)$, $q_2 = (1, 0)$, $q_3 = (0, 1)$, ...

Ursache: Periodizität!

4/44

Aperiodische Markov-Ketten

- Die **Periode** eines Zustands $j \in S$ ist die grösste Zahl $\xi \in \mathbb{N}$, so dass gilt:

$$\{n \in \mathbb{N} \mid p_{jj}^{(n)} > 0\} \subseteq \{i \cdot \xi \mid i \in \mathbb{N}\}$$

- Ein Zustand mit Periode $\xi = 1$ heisst **aperiodisch**.
- Eine Markov-Kette heisst **aperiodisch**, wenn alle Zustände aperiodisch sind.
- Nützliche Testbedingung:** Zustand j ist aperiodisch, falls eine der beiden folgenden Bedingungen gilt:
 - $p_{jj} > 0$
 - $\exists n, m \in \mathbb{N} : p_{jj}^{(m)}, p_{jj}^{(n)} > 0$ und $\text{ggT}(m, n) = 1$

4/45

Ergodische Markov-Ketten

- Irreduzible, aperiodische Markov-Ketten heissen **ergodisch**.

Satz (Fundamentalsatz für ergodische Markov-Ketten)

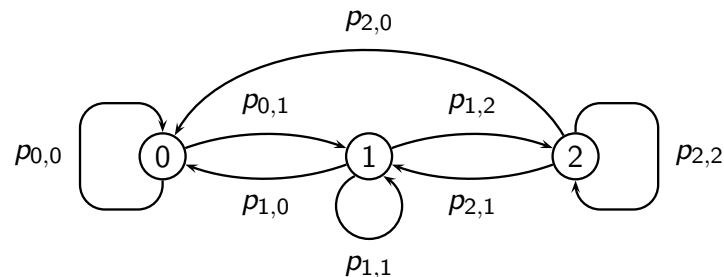
Für jede ergodische endliche Markov-Kette gilt unabhängig vom Startzustand

$$\lim_{t \rightarrow \infty} q_t = \pi,$$

wobei π die eindeutige stationäre Verteilung der Kette ist.

4/46

Beispiel: Rechensystem mit 2 Computern



$$\text{Übergangsmatrix } P = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0.35 & 0.5 & 0.15 \\ 0.245 & 0.455 & 0.3 \end{pmatrix}$$

- Kette ist aperiodisch und irreduzibel, also ergodisch.
- Aus $\pi = \pi P$ und $\pi_0 + \pi_1 + \pi_2 = 1$ erhält man die eindeutige stationäre Verteilung: $\pi = (0.399, 0.495, 0.106)$

4/47

Beispiel: Paging (1)

Modellierung eines Paging-Systems

- Hauptspeicher eines Rechners mit n logischen Seiten und $m < n$ physikalischen Seiten.
- Zugriff auf logische Seite σ , die nicht im physikalischen Hauptspeicher ist $\Rightarrow \sigma$ wird von Platte geladen, eine andere Seite wird aus dem physikalischen Hauptspeicher verdrängt.
- Zufallsvariable M_t gibt an, auf welche der n logischen Seiten zur Zeit t zugegriffen wird.
- Annahme:** M_t unabhängig von t und von Zugriffen in anderen Zeitschritten, also $\Pr[M_t = i] = \beta_i$ für $1 \leq i \leq n$, wobei $\sum_{i=1}^n \beta_i = 1$.

4/48

Beispiel: Paging (2)

- ▶ Betrachte Paging-Strategie LRU (least recently used): Die Seite, die am längsten nicht mehr zugegriffen wurde, wird verdrängt.
- ▶ Modell des Paging-Systems: **Markov-Kette** in diskreter Zeit.
- ▶ Zustand s_t zur Zeit t : Menge der m im phys. Hauptspeicher befindlichen logischen Seiten nach Zugriff zur Zeit $t - 1$.
- ▶ Betrachte Spezialfall $m = 2$: Zustand s_t zur Zeit t ist Paar $s_t = (i, j)$, wenn i und j die Seiten im physikalischen Speicher sind und zuletzt auf i zugegriffen wurde (zur Zeit $t - 1$).
- ▶ Wenn $s_t = (i, j)$, dann $s_{t+1} = \begin{cases} (i, j) & \text{falls } M_t = i \\ (j, i) & \text{falls } M_t = j \\ (k, i) & \text{falls } M_t = k \end{cases}$

4/49

Beispiel: Paging (3)

Für $m = 2$ und $n = 3$ erhalten wir folgende Übergangsmatrix P :

	(1, 2)	(2, 1)	(1, 3)	(3, 1)	(2, 3)	(3, 2)
(1, 2)	β_1	β_2	0	β_3	0	0
(2, 1)	β_1	β_2	0	0	0	β_3
(1, 3)	0	β_2	β_1	β_3	0	0
(3, 1)	0	0	β_1	β_3	β_2	0
(2, 3)	β_1	0	0	0	β_2	β_3
(3, 2)	0	0	β_1	0	β_2	β_3

4/50

Beispiel: Paging (4)

- ▶ Markov-Kette ist irreduzibel und aperiodisch, also ergodisch.
- ▶ Durch Lösen des Gleichungssystems $\pi = \pi \cdot P$, $\sum_{(i,j) \in S} \pi_{(i,j)} = 1$ erhält man:

$$\pi_{(i,j)} = \frac{\beta_i \beta_j}{1 - \beta_i}.$$
- ▶ Wahrscheinlichkeit, dass im Zustand (i, j) eine Seite nachgeladen werden muss, ist $1 - (\beta_i + \beta_j)$.
- ▶ Über lange Zeit ist in jedem Zeitschritt die Wahrscheinlichkeit, dass eine Seite nachgeladen werden muss, gegeben durch:

$$\sum_{(i,j) \in S} (1 - \beta_i - \beta_j) \frac{\beta_i \beta_j}{1 - \beta_i}$$

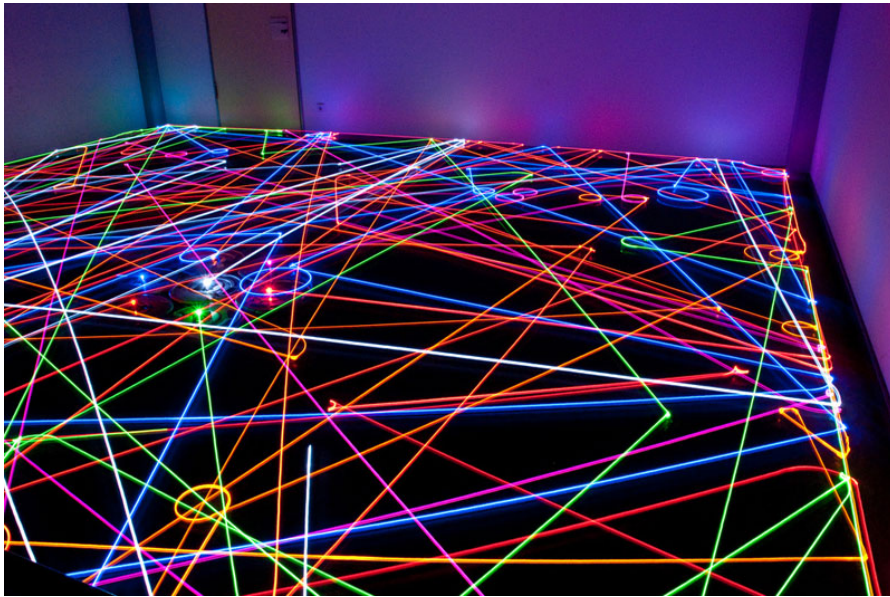
4/51

Stationäre Verteilung: Sichtweisen

1. Eigenvektor der Übergangsmatrix zum Eigenwert 1
 - ▶ numerisch anspruchsvoll
2. Wahrscheinlichkeitsverteilung iterativ von Runde zu Runde bestimmen
 - ▶ Wann kann man aufhören?
3. **Random Walk** im Graphen der Markov-Kette mit Histogramm der Zustände
 - ▶ Wann kann man aufhören?

4/52

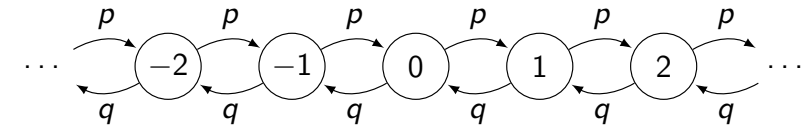
Roomba: „Random Walk“



4/53

Einfacher Random Walk in 1D

- ▶ Gambler's Ruin mit zwei unendlich reichen Spielern.
- ▶ Markov-Kette mit unendlich vielen Zuständen entsprechend \mathbb{Z} .
- ▶ Wechsel von Zustand z nach $z + 1$ mit W'keit $p = \frac{1}{2}$ und umgekehrt mit $q = 1 - p = \frac{1}{2}$



- ▶ Was ist die **Hitting Time** $h_{0,1}$ von Zustand 0 nach Zustand 1?

4/54

Einfacher Random Walk $\geq 1D$

- ▶ Markov-Kette mit Zuständen entsprechend \mathbb{Z}^d .

Lemma

Für einen Random Walk auf \mathbb{Z}^d für $d = 1, 2$ gilt $f_{0,0} = 1$.

Lemma

Für einen Random Walk auf \mathbb{Z}^3 gilt $f_{0,0} \approx 0.34$.

4/55

Random Walk in Graphen

- ▶ Gegeben ist ein **ungerichteter verbundener Graph** $G = (V, E)$ mit n Knoten und m Kanten, sowie ein Startknoten $s \in V$.
- ▶ In jedem Zeitschritt wechselt man **uniform** zu einem der Nachbarknoten.
- ▶ **Cover Time** $\text{cov}(s)$ = erwartete Anzahl Schritte bis ausgehend von s alle Knoten in V besucht sind.
- ▶ Verbindung zwischen **Cover Time** und **Hitting Time**:
 - ▶ Markov-Kette M mit Zuständen (v, S) , wobei v = aktueller Knoten und S = bereits besuchte Knoten
 - ▶ $\text{cov}(s) \approx h_{(s, \{s\}), (s, V)}$

4/56

Stationäre Verteilung von Random Walks

Lemma

Die stationäre Verteilung ist gegeben durch

$$\pi_u = \frac{d_u}{2m} \quad (d_u = \text{Grad von } u)$$

Beweis.

Annahme: Für alle j gilt $\pi_j = d_j/2m$. Dann folgt für beliebiges u

$$\pi_u = \sum_{v \in N(u)} \pi_v \cdot p_{v,u} = \sum_{v \in N(u)} \frac{d_v}{2m} \cdot \frac{1}{d_v} = \frac{d_u}{2m}.$$

Folglich ist $\pi_j = d_j/2m$ die eindeutige stationäre Verteilung. \square

- ▶ Damit gilt $h_{uu} = \frac{2m}{d_u}$.

4/57

Cover Time von Random Walks

Lemma

Für die **Cover Time** eines Graphen $G = (V, E)$ ausgehend von einem beliebigen Knoten $s \in V$ gilt $\text{cov}(s) < 4m(n-1)$.

Beweis.

- ▶ Für $(u, v) \in E$ folgt $h_{uv} < 2m$ aus $h_{uu} = \frac{2m}{d_u}$, denn

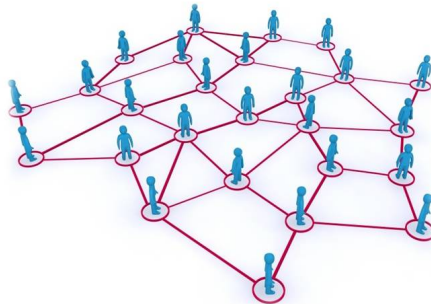
$$\frac{2m}{d_u} = h_{uu} = \frac{1}{d_u} \sum_{w \in N(u)} (h_{wu} + 1).$$

- ▶ Es existiert eine Traversierung der Knoten von G ausgehend von s , die $2n-2$ Kanten besucht (beliebigem Spannbaum mit Wurzel s doppelt ablaufen). Mit $h_{uv} < 2m$ für $(u, v) \in E$ folgt

$$\text{cov}(s) < (2n-2) \cdot 2m = 4m(n-1). \quad \square$$

4/58

Cover Time verschiedener Graphklassen



Wie schnell werden verschiedene Graph-Klassen exploriert?

- ▶ Clique \Rightarrow Coupon Collector
- ▶ Lineare Liste
- ▶ Baum
- ▶ Worst-Case?

4/59

Random Walks vs. elektrische Schaltkreise

- ▶ Kante e im Graphen $\hat{=}$ Widerstand mit $R_e = 1\Omega$
- ▶ Zwei Knoten u und v werden über eine 1V-Batterie verbunden.
- ▶ $R(u, v) =$ Ersatzwiderstand des Netzwerks $= \frac{1V}{I_{u,v}}$

Lemma

Für die **Commute Time** c_{uv} zwischen zwei Knoten u und v in einem Graphen mit m Kanten gilt

$$c_{uv} = 2m \cdot R(u, v).$$

4/60

Foster's Network Theorem

- ▶ **Hinzufügen/entfernen** einer Kante im Graph kann den Ersatzwiderstand nur **reduzieren** bzw. **erhöhen**

Satz (Foster's Network Theorem)

Für jeden verbundenen Graphen $G = (V, E)$ mit n Knoten gilt

$$\sum_{(u,v) \in E} R(u, v) = n - 1.$$

4/61

Anwendung: PageRank (1)

Wie rankt Google Webseiten?

- ▶ **Gesucht:** Vektor $\mathbf{v} \in \mathbb{R}^n$ für n Webseiten, so dass $v_i \geq 0$ die Wichtigkeit von Seite i misst.
- ▶ Annahme: „Link von Seite i nach j “ = „ i empfiehlt j “
- ▶ Ansatz: $A = (n \times n)$ -Matrix mit

$$a_{ij} = \begin{cases} 1 & \text{wenn Seite } i \text{ nach } j \text{ linkt} \\ 0 & \text{sonst} \end{cases}$$

- ▶ Ranking: $\mathbf{v} = \mathbf{1}^n \cdot A$ (\mathbf{v} und $\mathbf{1}^n$ wie üblich Zeilenvektor)

4/62

Anwendung: PageRank (2)

- ▶ **Verbesserung 1:** „Empfehlung“ von Seite i normalisieren durch Anzahl Links ausgehend von i .

$$d_i = \sum_j a_{ij}$$

- ▶ Neue Matrix P :

$$p_{ij} = \begin{cases} 1/d_i & \text{wenn Seite } i \text{ nach } j \text{ linkt} \\ 0 & \text{sonst} \end{cases}$$

- ▶ Ranking $\mathbf{v} = \mathbf{1}^n \cdot P$

4/63

Anwendung: PageRank (3)

- ▶ **Verbesserung 2:** Iterative Berechnung
- ▶ Gewicht der Empfehlung hängt von Wichtigkeit der empfehlenden Seite ab.

$$v_i = \sum_j p_{ji} \cdot v_j$$

- ▶ Ranking $\mathbf{v} = \mathbf{v} \cdot P$
- ▶ \mathbf{v} ist Eigenvektor zum Eigenwert 1 \Rightarrow stationäre Verteilung!

4/64

Anwendung: PageRank (4)

- ▶ **Verbesserung 3:** „Random Surfer“
- ▶ Annahme: In jedem Schritt wirft der User eine Münze mit W'keit für Kopf r und W'keit für Zahl $1 - r$.
- ▶ Wirft er Kopf, so wechselt er uniform zu einer verlinkten Seite. Wirft er Zahl, so wechselt er zufällig zu einer beliebigen aller n Seiten.
- ▶ Neue Matrix:

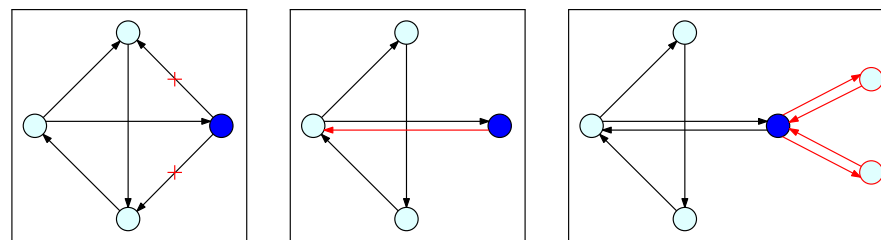
$$Q = r \cdot P + (1 - r) \cdot \begin{pmatrix} 1/n & \dots & 1/n \\ \vdots & \ddots & \vdots \\ 1/n & \dots & 1/n \end{pmatrix}^{n \times n}$$

- ▶ Die zugehörige Markov-Kette M_Q ist nun ergodisch.
- ▶ Ranking $\mathbf{v} = \mathbf{v} \cdot Q$ ist stationäre Verteilung von M_Q .

4/65

Google Bombing

- ▶ New Oxford American Dictionary, 2005:
„The activity of designing Internet links that will bias search engine results so as to create an inaccurate impression of the search target.“
- ▶ **Schritt 1:** Ausgehende Links entfernen.
- ▶ **Schritt 2:** $\pi_x = 1/h_{xx} \Rightarrow$ Auf Seiten linken, die „schnell“ zurücklinken.
- ▶ **Schritt 3:** Seiten erstellen, die direkt zurücklinken (Sybil-Angriff)



4/66

Vektor-Ketten

- ▶ Prozesse, bei denen X_t nicht nur von X_{t-1} abhängt, sondern auch von X_{t-2}, \dots, X_{t-k} , sind keine Markov-Prozesse.
- ▶ Sie können aber in Vektor-Ketten mit Markov-Eigenschaft umgewandelt werden.
- ▶ **Beispiel:** Der Prozess mit Zustandsmenge $S = \mathbb{Z}$ und $X_t = X_{t-1} - X_{t-2}$ erfüllt nicht die Markov-Bedingung.
- ▶ Der Vektorprozess mit Zustandsvektoren $Z_t = (X_t, X_{t-1})^T$ erfüllt die Markov-Bedingung:

$$Z_t = \begin{pmatrix} X_t \\ X_{t-1} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} X_{t-1} \\ X_{t-2} \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \cdot Z_{t-1}$$

4/67

4.3 Stoch. Prozesse in kontinuierlicher Zeit

Stochastische Prozesse in kontinuierlicher Zeit

- ▶ Oft müssen diskrete Ereignis-Systeme betrachtet werden, bei denen die Ereignisse zu beliebigen Zeitpunkten eintreten können (d.h. in kontinuierlicher Zeit).
- ▶ Im Weiteren:
 - ▶ Kontinuierliche Zufallsvariablen
 - ▶ Markov-Ketten in kontinuierlicher Zeit
 - ▶ Warteschlangen

4/68

Kontinuierliche Zufallsvariablen

- ▶ Einer kontinuierlichen Zufallsvariable X liegt der kontinuierliche Wahrscheinlichkeitsraum $\Omega = \mathbb{R}$ zugrunde.
- ▶ X ist definiert durch eine integrierbare **Dichte** (auch: Dichtefunktion) $f_X : \mathbb{R} \rightarrow \mathbb{R}_0^+$ mit der Eigenschaft

$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

- ▶ Jeder Dichte f_X kann eine **Verteilung** (auch: Verteilungsfunktion) F_X zugeordnet werden:

$$F_X(x) := \Pr[X \leq x] = \int_{-\infty}^x f_X(t) dt$$

- ▶ $\Pr[a < X \leq b] = \int_{(a,b]} f_X(x) dx = F_X(b) - F_X(a)$

4/69

Erwartungswert und Varianz

- ▶ Zur Berechnung von Erwartungswert und Varianz einer kontinuierlichen Zufallsvariable X ersetzen wir die Summen aus dem diskreten Fall durch Integrale.

- ▶ $\mathbf{E}[X] = \int_{-\infty}^{\infty} t \cdot f_X(t) dt,$
falls $\int_{-\infty}^{\infty} |t| \cdot f_X(t) dt$ endlich.

- ▶ $\mathbf{Var}[X] = \mathbf{E}[(X - \mathbf{E}[X])^2] = \int_{-\infty}^{\infty} (t - \mathbf{E}[X])^2 \cdot f_X(t) dt,$
falls $\mathbf{E}[(X - \mathbf{E}[X])^2]$ existiert.

- ▶ Kontinuierliche Zufallsvariablen X und Y heißen **unabhängig**, falls

$$\forall x, y \in \mathbb{R} : \Pr[X \leq x, Y \leq y] = \Pr[X \leq x] \cdot \Pr[Y \leq y].$$

4/70

Beispiele

Beispiele kontinuierlicher Verteilungen

- ▶ **Gleichverteilung** auf $[a, b]$:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{falls } a \leq x \leq b \\ 0, & \text{sonst} \end{cases}$$

$$\mathbf{E}[X] = \frac{a+b}{2}, \quad \mathbf{Var}[X] = \frac{(b-a)^2}{12}$$

- ▶ **Normalverteilung** mit Parametern μ und σ :

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\mathbf{E}[X] = \mu, \quad \mathbf{Var}[X] = \sigma^2$$

4/71

Exponentialverteilung (1)

Exponentialverteilung mit Parameter $\lambda > 0$

- ▶ Dichte $f_X(x) = \begin{cases} \lambda \cdot e^{-\lambda x}, & \text{falls } x \geq 0 \\ 0, & \text{sonst} \end{cases}$

- ▶ $\mathbf{E}[X] = \frac{1}{\lambda}, \quad \mathbf{Var}[X] = \frac{1}{\lambda^2}$

- ▶ Verteilungsfunktion $F_X(x) = \begin{cases} 1 - e^{-\lambda x}, & \text{falls } x \geq 0 \\ 0, & \text{sonst} \end{cases}$

- ▶ **Gutes Modell für**

- ▶ Dauer von Telefongesprächen
- ▶ Zwischenankunftszeiten von Anfragen
- ▶ Ausführungszeiten von Tasks

4/72

Exponentialverteilung (2)

Eigenschaften der Exponentialverteilung

- ▶ **Gedächtnislosigkeit:** Für alle $x, y > 0$ gilt:

$$\Pr[X > x + y \mid X > y] = \Pr[X > x]$$

- ▶ **Skalierung:** Falls X exponentialverteilt mit Parameter λ , so ist für $a > 0$ die Zufallsvariable $Y := aX$ exponentialverteilt mit Parameter λ/a .
- ▶ **Warteproblem:** Falls X_1, \dots, X_n unabhängig und exponentialverteilt mit Parametern $\lambda_1, \dots, \lambda_n$, dann ist $X := \min\{X_1, \dots, X_n\}$ exponentialverteilt mit dem Parameter $\lambda_1 + \dots + \lambda_n$.

4/73

Ereignissysteme in kontinuierlicher Zeit

Diskrete Ereignissysteme in kontinuierlicher Zeit

- ▶ In vielen Systemen ist es unnatürlich, Ereignisse nur zu diskreten Zeitpunkten zuzulassen:
 - ▶ Ankunft von Paketen in einem Router
 - ▶ Auftreten von Anfragen an einen Server
- ▶ Um stochastische Prozesse in kontinuierlicher Zeit zu modellieren, können wieder Markov-Ketten verwendet werden:
 - ▶ Zustandsübergänge nicht nur zu diskreten Zeitpunkten zulassen, sondern exponentialverteilte Aufenthaltsdauern annehmen!

4/74

Markov-Ketten in kontinuierlicher Zeit (1)

Endliche Markov-Kette in kontinuierlicher Zeit

über der Zustandsmenge $S = \{0, 1, \dots, n-1\}$:

- ▶ **Folge von Zufallsvariablen** $(X(t))_{t \in \mathbb{R}_0^+}$ mit Wertemenge S
- ▶ **Startverteilung** $q(0) = (q_0(0), q_1(0), \dots, q_{n-1}(0))$ mit $q_i(0) \geq 0$ und $\sum_{i=0}^{n-1} q_i(0) = 1$.
- ▶ **Markov-Bedingung:** Für alle $k \in \mathbb{N}_0$ und beliebige $0 \leq t_0 < t_1 < \dots < t_k < t$ und $s, s_0, \dots, s_k \in S$ gilt:

$$\begin{aligned} \Pr[X(t) = s \mid X(t_k) = s_k, X(t_{k-1}) = s_{k-1}, \dots, X(t_0) = s_0] \\ = \Pr[X(t) = s \mid X(t_k) = s_k] \end{aligned}$$

($S = \mathbb{N}_0 \Rightarrow$ **unendliche Markov-Kette in kontinuierlicher Zeit.**)

4/75

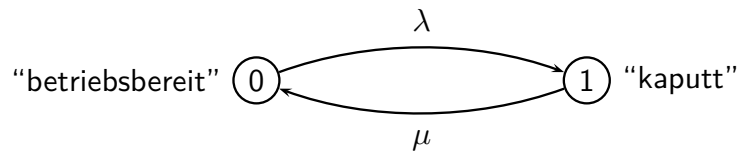
Markov-Ketten in kontinuierlicher Zeit (2)

- ▶ **Bemerkung:** Aus der Markov-Bedingung (Gedächtnislosigkeit) für die Markov-Kette kann man folgern, dass die Aufenthaltsdauern in den Zuständen exponentialverteilt sein müssen.
- ▶ Falls $\Pr[X(t+u) = j \mid X(t) = i] = \Pr[X(u) = j \mid X(0) = i]$ für alle $i, j \in S$ und $t, u \in \mathbb{R}_0^+$, so heisst die Markov-Kette **zeithomogen**.
- ▶ Wir betrachten im Folgenden ausschliesslich zeithomogene Markov-Ketten.

4/76

Beispiel

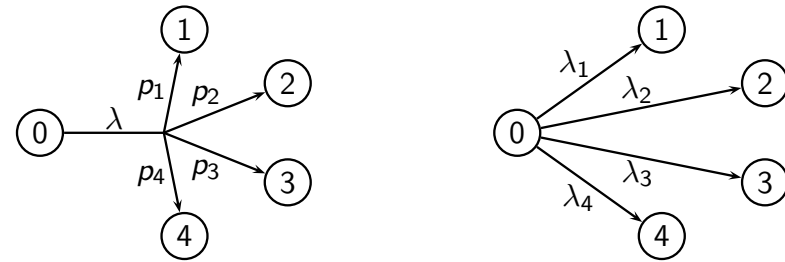
Beispiel einer Markov-Kette in kontinuierlicher Zeit:



- ▶ Aufenthaltsdauer in Zustand 0 ist exponentialverteilt mit Parameter λ .
- ▶ Aufenthaltsdauer in Zustand 1 ist exponentialverteilt mit Parameter μ .

4/77

Zustände mit mehreren Nachfolgern (1)



Gleichwertige Sichtweisen:

1. Zustand 0 hat Aufenthaltsdauer exponentialverteilt mit Parameter λ . Wenn Zustand 0 verlassen wird, so werden die Nachfolger mit Wahrscheinlichkeit p_1, p_2, p_3, p_4 ausgewählt, $p_1 + p_2 + p_3 + p_4 = 1$.
2. Es werden gleichzeitig vier Werte zufällig bestimmt gemäss Exponentialverteilungen mit Parametern λ_1 bis λ_4 , wobei $\lambda_i = \lambda \cdot p_i$ für $i = 1, 2, 3, 4$. Der kleinste Wert "gewinnt".

4/78

Zustände mit mehreren Nachfolgern (2)

Allgemein bedeutet das:

- ▶ Jeder Zustand $i \in S$ hat eine exponentialverteilte Aufenthaltsdauer mit Parameter ν_i .
- ▶ Wenn Zustand $i \in S$ verlassen wird, so wird mit Wahrscheinlichkeit p_{ij} der Nachfolgezustand $j \in S$ angenommen, wobei $p_{i,i} = 0$ und $\sum_{j \in S} p_{i,j} = 1$.
- ▶ Die **Übergangsrate** von Zustand i nach j ist als $\nu_{i,j} := \nu_i \cdot p_{i,j}$ definiert.
- ▶ Es gilt für $i \in S$: $\sum_{j \in S} \nu_{i,j} = \nu_i$.

4/79

Aufenthaltswahrscheinlichkeiten (1)

Bestimmung der Aufenthaltswahrscheinlichkeiten

- ▶ Startverteilung $q(0)$: $q_i(0) = \Pr[X(0) = i]$ für $i \in S$
- ▶ Verteilung zur Zeit t : $q_i(t) = \Pr[X(t) = i]$ für $i \in S$
- ▶ Die Änderung der Aufenthaltswahrscheinlichkeiten kann durch **Differentialgleichungen** für alle $i \in S$ beschrieben werden:

$$\underbrace{\frac{d}{dt} q_i(t)}_{\text{Änderung}} = \underbrace{\sum_{j:j \neq i} q_j(t) \cdot \nu_{j,i}}_{\text{Zufluss}} - \underbrace{q_i(t) \cdot \nu_i}_{\text{Abfluss}}$$

4/80

Aufenthaltswahrscheinlichkeiten (2)

$$\frac{d}{dt}q_i(t) = \sum_{j:j \neq i} q_j(t) \cdot \nu_{j,i} - q_i(t) \cdot \nu_i$$

- ▶ Lösung dieser Differentialgleichungen ist meist aufwändig.
- ▶ Betrachte Verhalten des Systems für $t \rightarrow \infty$.
- ▶ Falls die Aufenthaltswahrscheinlichkeiten gegen eine stationäre Verteilung konvergieren, so muss $\frac{d}{dt}q_i(t) = 0$ gelten.
- ▶ Man erhält für $t \rightarrow \infty$ ein lineares Gleichungssystem, das von einer stationären Verteilung π erfüllt werden muss:

$$0 = \sum_{j:j \neq i} \pi_j \cdot \nu_{j,i} - \pi_i \cdot \nu_i, \quad \text{für alle } i \in S$$

4/81

Irreduzible Markov-Ketten

- ▶ Ein 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 jeder Zustand von jedem anderen aus erreichbar ist, heisst **irreduzibel**.

Satz. Für irreduzible Markov-Ketten existieren die Grenzwerte

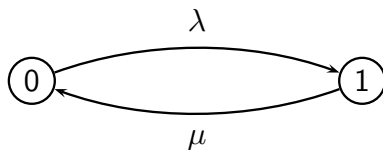
$$\pi_i := \lim_{t \rightarrow \infty} q_i(t)$$

für alle $i \in S$ und ihre Werte sind unabhängig von $q(0)$.

4/82

Berechnung der stationären Verteilung

Im Beispiel:



Gleichungssystem:

$$0 = \mu \cdot \pi_1 - \lambda \cdot \pi_0$$

$$0 = \lambda \cdot \pi_0 - \mu \cdot \pi_1$$

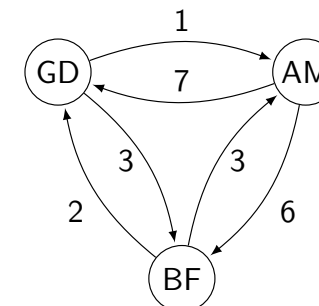
Zusammen mit $\pi_0 + \pi_1 = 1$ erhält man:

$$\pi_0 = \frac{\mu}{\lambda + \mu} \quad \pi_1 = \frac{\lambda}{\lambda + \mu}$$

4/83

Exkursion: Rank Aggregation

- ▶ Verschiedene User bewerten drei Filme relativ zueinander
(GD = Groundhog Day, AM = 12 Angry Men, BF = Back To The Future)

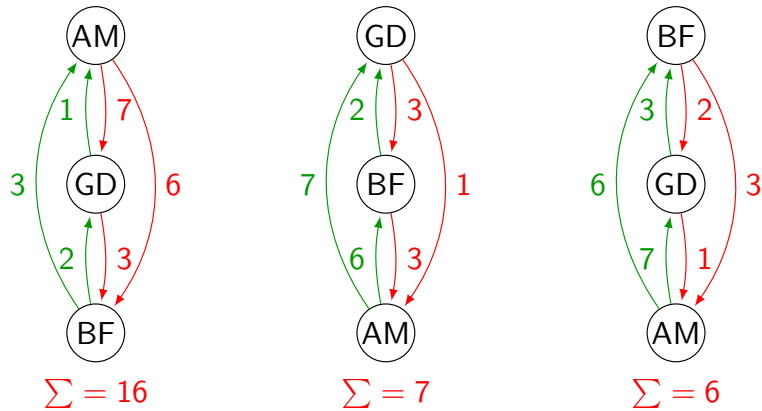


- ▶ Welcher Film ist der beste?

4/84

Feedback Arc Set Problem

- ▶ Filme so ordnen, dass „Gegenmeinungen“ minimiert werden.

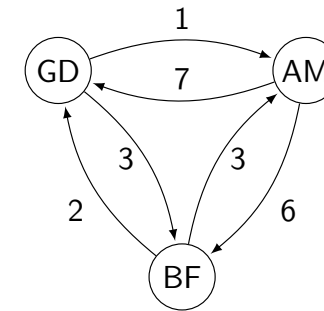


- ▶ **Back To The Future** ist der beste Film.
- ▶ Minimum Feedback Arc Set Problem ist **NP-schwer**.

4/85

Stationäre Verteilung

- ▶ Abstimmungsgraph als Markov-Kette interpretieren und stationäre Verteilung bestimmen.

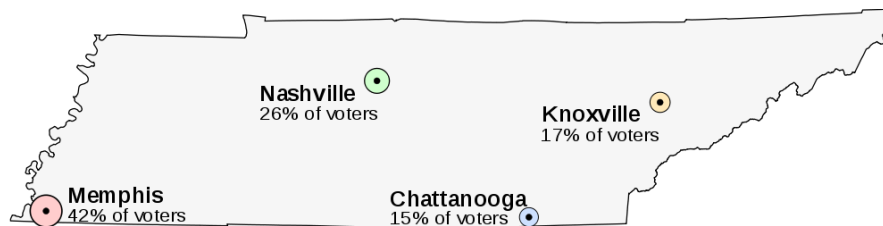


$$\pi(GD, BF, AM) = \left(\frac{47}{106}, \frac{45}{106}, \frac{14}{106} \right)$$

- ▶ **Groundhog Day** ist der beste Film.

4/86

Tennessees Hauptstadt



42% of voters (close to Memphis)	26% of voters (close to Nashville)	15% of voters (close to Chattanooga)	17% of voters (close to Knoxville)
1. Memphis	1. Nashville	1. Chattanooga	1. Knoxville
2. Nashville	2. Chattanooga	2. Knoxville	2. Chattanooga
3. Chattanooga	3. Knoxville	3. Nashville	3. Nashville
4. Knoxville	4. Memphis	4. Memphis	4. Memphis

Quelle: Wikipedia

4/87

Arrow-Theorem

Satz

Es existiert kein Rangordnungssystem, das die folgenden drei „Fairness“-Kriterien gleichzeitig erfüllt:

- ▶ Wenn jeder Wähler Alternative A gegenüber Alternative B bevorzugt, dann bevorzugt auch das System A gegenüber B.
- ▶ Für das Ranking von zwei Alternativen A und B sind ausschließlich die Präferenzen der Wähler bezüglich dieser beiden Alternativen relevant.
- ▶ Es gibt keinen Diktator, der allein über die Präferenzordnung der Gesellschaft entscheidet.

4/88

Warteschlangen (1)

Warteschlangentheorie

- ▶ Besonders wichtige Anwendung von Markov-Ketten mit kontinuierlicher Zeit.
- ▶ Systeme mit Servern, die Jobs abarbeiten
- ▶ **Ankunftszeiten** der Jobs und **Bearbeitungsdauern** auf den Servern werden als Zufallsvariablen modelliert.
- ▶ Jobs, die ankommen, wenn alle Server belegt sind, werden in eine Warteschlange eingefügt.
- ▶ Ein freiwerdender Server wählt einen neuen Job aus der Warteschlange zur Bearbeitung aus (hier: FCFS, "first come, first serve," aber andere Strategien denkbar).

4/89

Warteschlangen (2)

- ▶ Beispielanwendung: Paketverzögerung in Datennetzen (Pakete = Jobs), Antwortzeiten von Tasks in Rechenzentren, ...
 - ▶ Interessante Größen wie
 - ▶ durchschnittliche Anzahl Jobs im System
 - ▶ durchschnittliche Verzögerung (Antwortzeit, Systemzeit, Aufenthaltsdauer) der Jobs
- werden in Abhängigkeit von der **Ankunftsrate** (mittlere Anzahl ankommender Jobs pro Zeiteinheit) und den **Bearbeitungsdauern** analysiert, wobei das System über lange Zeit betrachtet wird.

4/90

Kendall-Notation $X/Y/m/\dots$

- ▶ X steht für die Verteilung der **Zwischenankunftszeiten** (Zeiten zwischen zwei ankommenden Jobs).
- ▶ Y steht für die Verteilung der reinen **Bearbeitungszeiten** (d.h. ohne Wartezeit) der Jobs auf dem Server.
- ▶ Die Zwischenankunftszeiten und Bearbeitungszeiten sind **unabhängige** Zufallsvariablen.
- ▶ m steht für die **Anzahl der Server**.
- ▶ Die Verteilungen für X und Y werden angegeben als:
 - ▶ **"D"** für feste Dauer (engl. *deterministic*)
 - ▶ **"M"** für exponentialverteilt (engl. *memoryless*)
 - ▶ **"G"** für beliebige Verteilung (engl. *general*)

4/91

Der Poisson-Prozess

- ▶ Im Fall von exponentialverteilten Zwischenankunftszeiten (mit Parameter λ) ist der Ankunftsprozess der Jobs ein **Poisson-Prozess mit Rate λ** .
- ▶ Die Anzahl ankommender Jobs in einem Intervall der Länge τ ist nämlich Poisson-verteilt mit Rate $\lambda\tau$:

$$\Pr[\alpha(t + \tau) - \alpha(t) = n] = e^{-\lambda\tau} \frac{(\lambda\tau)^n}{n!}, \quad \text{für } n = 0, 1, 2, \dots$$

$$\mathbf{E}[\alpha(t + \tau) - \alpha(t)] = \lambda \cdot \tau$$

- ▶ Poisson-Prozesse sind ein gutes Modell für die Ankunft von Paketen, Anfragen, Telefongesprächen, Jobs, etc., die von **vielen unabhängigen und ähnlichen Benutzern** erzeugt werden.

4/92

M/M/1-Warteschlangen

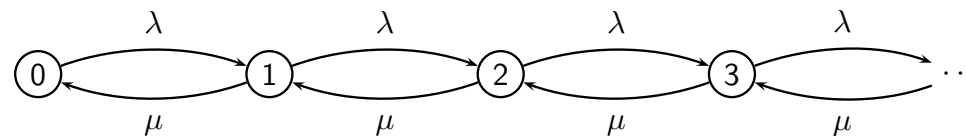
Die M/M/1-Warteschlange



- ▶ Zwischenankunftszeiten und Bearbeitungszeiten exponentialverteilt mit Parameter λ (Ankunftsrate) bzw. μ (Bedienrate).
- ▶ Definition: **Verkehrsdichte** $\rho := \frac{\lambda}{\mu}$
- ▶ Modellierung als Markov-Kette in kontinuierlicher Zeit:
 - ▶ Zustand: Anzahl Jobs im System (Warteschlange + Server)
 - ▶ Zustandsmenge $S = \mathbb{N}_0$
 - ▶ Übergangsrate von i nach $i + 1$ ist λ .
 - ▶ Übergangsrate von $i > 0$ nach $i - 1$ ist μ .

4/93

M/M/1: Stationäre Verteilung (1)



Gleichungssystem für stationäre Verteilung π :

$$0 = \mu \cdot \pi_1 - \lambda \pi_0$$

$$0 = \lambda \cdot \pi_{k-1} + \mu \cdot \pi_{k+1} - (\lambda + \mu)\pi_k \quad \text{für alle } k \geq 1$$

Umformen liefert:

$$\mu \cdot \pi_{k+1} - \lambda \cdot \pi_k = \mu \cdot \pi_k - \lambda \cdot \pi_{k-1} = \dots = \mu \cdot \pi_1 - \lambda \cdot \pi_0 = 0$$

$$\Rightarrow \mu \cdot \pi_k - \lambda \cdot \pi_{k-1} = 0 \Rightarrow \pi_k = \rho \cdot \pi_{k-1} \Rightarrow \pi_k = \rho^k \cdot \pi_0$$

4/94

M/M/1: Stationäre Verteilung (2)

Wir wissen: $\pi_k = \rho^k \cdot \pi_0$ für alle $k \geq 0$

- ▶ Falls $\rho \geq 1$, ist $\pi = (0, 0, \dots)$ die einzige Lösung. Das System **konvergiert nicht**, die Warteschlange wächst ins Unendliche.
- ▶ Falls $\rho < 1$, so rechnen wir:

$$1 = \sum_{k=0}^{\infty} \pi_k = \pi_0 \cdot \sum_{k=0}^{\infty} \rho^k = \pi_0 \cdot \frac{1}{1 - \rho} \Rightarrow \pi_0 = 1 - \rho$$

Das System **konvergiert** gegen eine stationäre Verteilung π mit $\pi_k = (1 - \rho)\rho^k$ für alle $k \geq 0$.

Die mittlere **Auslastung** des Servers ist $1 - \pi_0 = \rho$.

4/95

M/M/1: Anzahl Jobs im System (1)

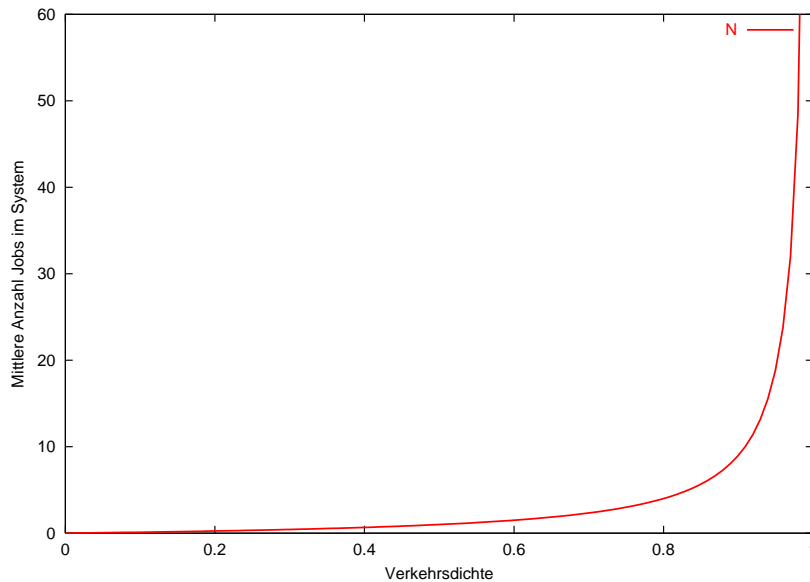
- ▶ Sei N der Erwartungswert der Anzahl der Jobs im System (Warteschlange + Server).
- ▶ In der stationären Verteilung ergibt sich:

$$\begin{aligned} N &= \sum_{k=0}^{\infty} k \cdot \pi_k = \sum_{k=0}^{\infty} k(1 - \rho)\rho^k = (1 - \rho)\rho \sum_{k=0}^{\infty} k\rho^{k-1} \\ &= (1 - \rho)\rho \frac{1}{(1 - \rho)^2} = \frac{\rho}{1 - \rho} = \frac{\lambda}{\mu - \lambda} \end{aligned}$$

Die Varianz der Anzahl Jobs im System ist $\frac{\rho}{(1 - \rho)^2}$.

4/96

M/M/1: Anzahl Jobs im System (2)



4/97

Little's Law – Definitionen

- ▶ $N(t) :=$ Anzahl Jobs im System (Warteschlange + Server) zur Zeit t .
- ▶ $\alpha(t) :=$ Anzahl Jobs, die in $[0, t]$ angekommen sind.
- ▶ $T_i :=$ Antwortzeit des i -ten Jobs (Wartezeit + Bearbeitungszeit).

Berechne **Durchschnittswerte** bis zur Zeit t :

$$N_t := \frac{1}{t} \int_0^t N(\tau) d\tau, \quad \lambda_t := \frac{\alpha(t)}{t}, \quad T_t := \frac{\sum_{i=1}^{\alpha(t)} T_i}{\alpha(t)}$$

Betrachte **Grenzwerte** für $t \rightarrow \infty$:

$$N := \lim_{t \rightarrow \infty} N_t, \quad \lambda := \lim_{t \rightarrow \infty} \lambda_t, \quad T := \lim_{t \rightarrow \infty} T_t,$$

4/98

Little's Law – Mittelwerte über die Zeit

Formel von Little

Falls die Grenzwerte

$$N := \lim_{t \rightarrow \infty} N_t, \quad \lambda := \lim_{t \rightarrow \infty} \lambda_t, \quad T := \lim_{t \rightarrow \infty} T_t,$$

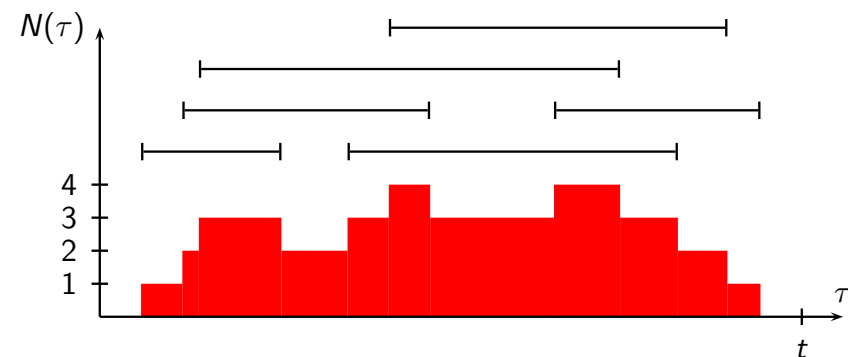
existieren und auch $\lim_{t \rightarrow \infty} \frac{\beta(t)}{t}$ existiert und gleich λ ist, wobei $\beta(t)$ die Anzahl der in $[0, t]$ beendeten Jobs ist, so gilt:

$$N = \lambda \cdot T$$

Bemerkung: Die Formel von Little gilt auch für andere Server-Strategien als FCFS.

4/99

Beweisidee zur Formel von Little



Annahme: $N(0) = 0$ und $N(t) = 0$ für unendlich viele, beliebig

grosse t . Dann gilt:

$$\underbrace{\frac{\alpha(t)}{t}}_{\xrightarrow{t \rightarrow \infty} \lambda} \cdot \underbrace{\frac{1}{\alpha(t)} \cdot \sum_{i=1}^{\alpha(t)} T_i}_{\xrightarrow{t \rightarrow \infty} T} = \underbrace{\frac{1}{t} \int_0^t N(\tau) d\tau}_{\xrightarrow{t \rightarrow \infty} N}$$

4/100

Little's Law – Stochastische Variante

Betrachte den Ablauf des Systems als stochastischen Prozess mit gegebener Startverteilung. Falls die Grenzwerte

$$\bar{N} = \lim_{t \rightarrow \infty} \mathbf{E}[N(t)], \quad \bar{T} = \lim_{i \rightarrow \infty} \mathbf{E}[T_i], \quad \lambda = \lim_{t \rightarrow \infty} \frac{\mathbf{E}[\alpha(t)]}{t}$$

existieren, so gilt:

$$\boxed{\bar{N} = \lambda \cdot \bar{T}}$$

Bemerkung 1: Die Formel von Little gilt für beliebige Verteilungen der Zwischenankunftszeiten und der Bearbeitungszeiten.

Bemerkung 2: Meist gilt $N = \bar{N}$ und $T = \bar{T}$ mit W'keit 1.

4/101

Little's Law und M/M/1-Warteschlangen

Mit $N = \frac{\rho}{1-\rho}$ und der Formel von Little erhalten wir:

$$\blacktriangleright T = \frac{1}{\lambda} N = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu - \lambda},$$

wobei T die mittlere **Antwortzeit** eines Jobs im Gleichgewichtszustand des Systems ist.

$$\blacktriangleright W = T - \frac{1}{\mu} = \frac{\rho}{\mu(1-\rho)} = \frac{\rho}{\mu - \lambda},$$

wobei W die mittlere **Wartezeit** (ohne Bearbeitungszeit) eines Jobs im Gleichgewichtszustand des Systems ist.

$\blacktriangleright N_Q = \lambda W = \frac{\rho^2}{1-\rho}$, wobei N_Q die mittlere Anzahl Jobs in der Warteschlange ist.

4/102

Anwendungen der Formel von Little (1)

Pakete in einem Datennetz

- ▶ Pakete werden an n Knoten in einem Netz mit Ankunftsraten $\lambda_1, \lambda_2, \dots, \lambda_n$ erzeugt.
- ▶ Jedes Paket wird im Netz zu seiner Zieladresse geleitet und dort aus dem Netz entfernt.
- ▶ Sei N die durchschnittliche Zahl von Paketen im Netz.
- ▶ Mit der Formel von Little lässt sich die durchschnittliche Paketverzögerung berechnen als:

$$T = \frac{N}{\sum_{i=1}^n \lambda_i}$$

4/103

Anwendungen der Formel von Little (2)

Ein geschlossenes Warteschlangensystem

- ▶ System mit K Servern und Platz für $N \geq K$ Jobs
- ▶ System sei immer voll ($N(t) = N$). Wenn ein Job abgearbeitet ist und das System verlässt, kommt sofort ein neuer Job an.
- ▶ Alle K Server bearbeiten durchgehend Jobs.
- ▶ Mittlere Bearbeitungszeit ist \bar{X} .

Bestimmung der **mittleren Antwortzeit** T :

- ▶ $N = \lambda T$ (Formel angewendet auf ganzes System)
- ▶ $K = \lambda \bar{X}$ (Formel angewendet auf K Server)
- ▶ $T = \frac{N}{\lambda} = \frac{N \cdot \bar{X}}{K}$

4/104

Anwendungen der Formel von Little (3)

Variante des Systems:

- ▶ Jobs kommen mit Rate λ an.
- ▶ Jobs werden abgewiesen, wenn bereits N Jobs im System sind.

Analyse des **Anteils abgewiesener Jobs**:

- ▶ \bar{K} = mittlere Anzahl aktiver Server
- ▶ β = Anteil abgewiesener Jobs
- ▶ Formel von Little $\Rightarrow \bar{K} = (1 - \beta)\lambda\bar{X}$
- ▶ Also: $\beta = 1 - \frac{\bar{K}}{\lambda\bar{X}} \geq 1 - \frac{K}{\lambda X}$ (untere Schranke für β)

4/105

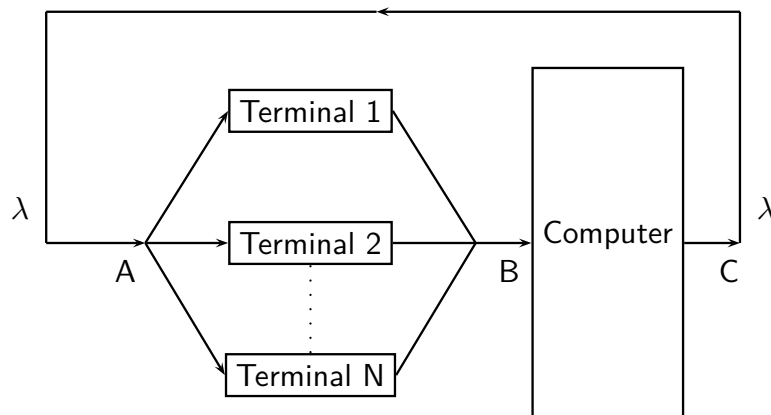
Durchsatzanalyse für Time-Sharing (1)

- ▶ System mit N Terminals, die mit einem Time-Sharing Computer verbunden sind.
- ▶ Benutzer an einem Terminal verhalten sich folgendermassen:
 - ① nachdenken (im Mittel R Sekunden)
 - ② an den Computer einen Job abschicken, der im Mittel Ausführungszeit P hat
 - ③ auf die Beendigung des Jobs warten
 - ④ System verlassen
- ▶ Im Computer werden die Jobs in einer Warteschlange eingereiht und von einer CPU abgearbeitet.
- ▶ **Ziel:** maximal erreichbaren Durchsatz λ abschätzen!

4/106

Durchsatzanalyse für Time-Sharing (2)

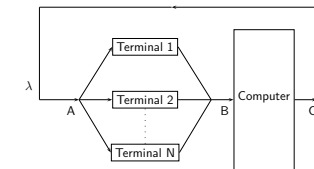
Schematische Darstellung des Systems:



Annahme: freiwerdende Terminals werden sofort wieder belegt
 \Rightarrow immer genau N Benutzer im System

4/107

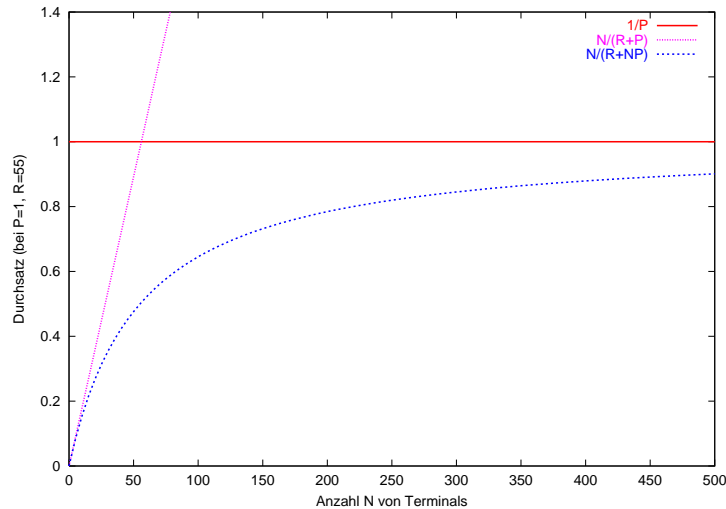
Durchsatzanalyse für Time-Sharing (3)



- ▶ Es gilt $\lambda = \frac{N}{T}$, wobei T = mittlere Aufenthaltszeit. (Formel von Little angewendet auf System zwischen A und C)
- ▶ Es gilt $T = R + D$, wobei $D \in [P, N \cdot P]$ die mittlere Zeit vom Abschicken eines Jobs bis zu seiner Erledigung ist.
- ▶ Somit gilt: $\frac{N}{R+NP} \leq \lambda \leq \frac{N}{R+P}$
- ▶ Klar: $\lambda \leq \frac{1}{P}$, da mittlere Ausführungszeit P ist.

4/108

Durchsatzanalyse für Time-Sharing (4)

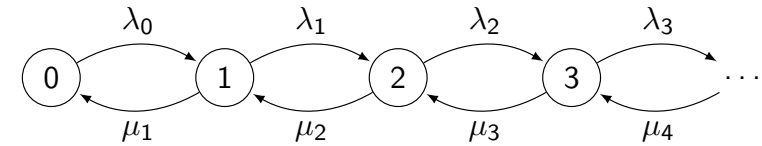


Maximal erzielbarer Durchsatz erfüllt $\frac{N}{R+NP} \leq \lambda \leq \min \left\{ \frac{N}{R+P}, \frac{1}{P} \right\}$.

4/109

Birth-and-Death Prozesse

- Verallgemeinerung der Markov-Kette der M/M/1-Warteschlange:



- Gleichungssystem für den Gleichgewichtszustand:

$$0 = \lambda_{k-1}\pi_{k-1} + \mu_{k+1}\pi_{k+1} - (\lambda_k + \mu_k)\pi_k \text{ für } k \geq 1,$$

$$0 = \mu_1\pi_1 - \lambda_0\pi_0$$

- Auflösen liefert $\pi_k = \pi_0 \cdot \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}$ für $k \geq 1$.

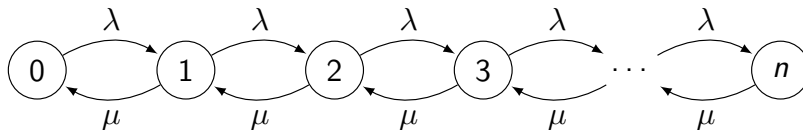
- Mit $\sum_{i=0}^{\infty} \pi_i = 1$ ergibt sich $\pi_0 = \frac{1}{1 + \sum_{k \geq 1} \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}}$.

4/110

Beispiel 1: Beschränkter Warteraum

M/M/1-Warteschlange mit nur n Plätzen

Neue Jobs werden abgewiesen, wenn bereits n Jobs im System sind. Es ergibt sich der folgende Birth-and-Death Prozess:



Wir erhalten:

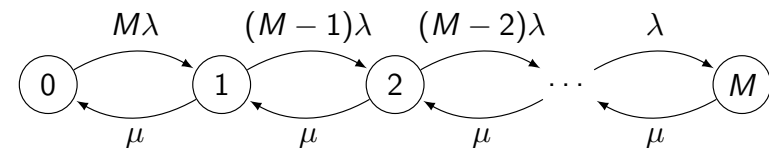
$$\pi_k = \rho^k \cdot \pi_0 \text{ für } 1 \leq k \leq n$$

$$\pi_0 = \frac{1}{\sum_{i=0}^n \rho^i} = \begin{cases} \frac{1}{n+1} & \text{für } \rho = 1 \\ \frac{1-\rho}{1-\rho^{n+1}} & \text{sonst} \end{cases}$$

4/111

Beispiel 2: Beschränkte Benutzerzahl

Anfragesystem mit M Terminals und einem Server



Wir erhalten:

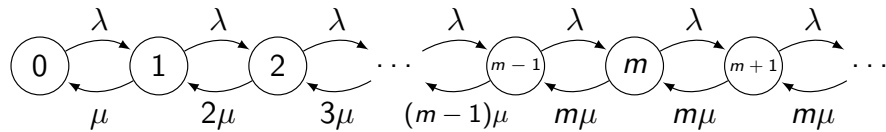
$$\pi_k = \pi_0 \cdot \prod_{i=0}^{k-1} \frac{\lambda(M-i)}{\mu} \text{ für } 1 \leq k \leq M$$

$$\pi_0 = \frac{1}{\sum_{k=0}^M \left(\frac{\lambda}{\mu}\right)^k \cdot M^k}$$

wobei $M^k := M(M-1)(M-2) \cdot \dots \cdot (M-k+1)$.

4/112

Das M/M/m-System (1)



Wir erhalten mit $\rho := \frac{\lambda}{m\mu} < 1$:

$$\pi_k = \begin{cases} \pi_0 \cdot \frac{\lambda^k}{\mu^k \cdot k!} = \pi_0 \cdot \frac{(\rho m)^k}{k!} & \text{für } 1 \leq k \leq m \\ \pi_0 \cdot \frac{\lambda^k}{\mu^k \cdot m! \cdot m^{k-m}} = \pi_0 \cdot \frac{\rho^k m^m}{m!} & \text{für } k \geq m \end{cases}$$

$$\pi_0 = \frac{1}{1 + \sum_{k=1}^{m-1} \frac{(\rho m)^k}{k!} + \sum_{k=m}^{\infty} \frac{\rho^k m^m}{m!}} = \frac{1}{\sum_{k=0}^{m-1} \frac{(\rho m)^k}{k!} + \frac{(\rho m)^m}{m!(1-\rho)}}$$

4/113

Das M/M/m-System (2)

Die Wahrscheinlichkeit P_Q , dass ein ankommender Job in der Warteschlange des M/M/m-Systems warten muss, ist also:

$$\begin{aligned} P_Q &= \sum_{k=m}^{\infty} \pi_k = \sum_{k=m}^{\infty} \frac{\pi_0 \rho^k m^m}{m!} \\ &= \frac{\pi_0 (\rho m)^m}{m!} \sum_{k=m}^{\infty} \rho^{k-m} = \frac{\pi_0 (\rho m)^m}{m!(1-\rho)} \end{aligned}$$

$$\Rightarrow P_Q = \frac{(\rho m)^m / (m!(1-\rho))}{\sum_{k=0}^{m-1} \frac{(\rho m)^k}{k!} + \frac{(\rho m)^m}{m!(1-\rho)}} \quad (\text{für } \rho = \frac{\lambda}{m\mu} < 1)$$

Diese Formel wird nach A.K. Erlang (1878-1929) die **Erlang C-Formel** genannt.

4/114

Das M/M/m-System (3)

Nun können wir aus P_Q weitere Größen ableiten:

- Für N_Q (erwartete Anzahl von Jobs in der Warteschlange) erhalten wir:

$$\begin{aligned} N_Q &= \sum_{k=1}^{\infty} k \cdot \pi_{m+k} = \sum_{k=1}^{\infty} k \cdot \pi_0 \cdot \frac{\rho^{m+k} m^m}{m!} \\ &= \pi_0 \frac{\rho^m m^m}{m!} \sum_{k=1}^{\infty} k \rho^k = \pi_0 \frac{\rho^{m+1} m^m}{m!(1-\rho)^2} \\ &= \frac{P_Q m!(1-\rho)}{\rho^m m^m} \cdot \frac{\rho^{m+1} m^m}{m!(1-\rho)^2} = P_Q \cdot \frac{\rho}{1-\rho} \end{aligned}$$

4/115

Das M/M/m-System (4)

- Für W (mittlere Wartezeit in der Queue) erhalten wir mit der Formel von Little:

$$W = \frac{N_Q}{\lambda} = P_Q \cdot \frac{\rho}{\lambda(1-\rho)} = \frac{\rho P_Q}{\lambda(1-\rho)}$$

- Die mittlere Antwortzeit T ist dann:

$$T = W + \frac{1}{\mu} = \frac{\rho P_Q}{\lambda(1-\rho)} + \frac{1}{\mu} = \frac{P_Q}{m\mu - \lambda} + \frac{1}{\mu}$$

- Erneute Anwendung der Formel von Little liefert die mittlere Anzahl N von Jobs im System:

$$N = \lambda T = \frac{\lambda P_Q}{m\mu - \lambda} + \frac{\lambda}{\mu} = \frac{\rho P_Q}{1-\rho} + m\rho$$

4/116

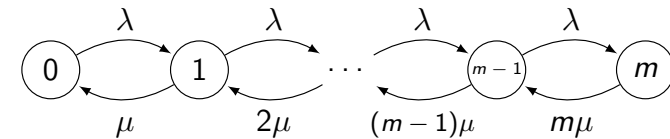
Das M/M/m/m-System (1)

- ▶ System mit m Servern und maximal m Jobs im System.
- ▶ Jobs, die ankommen, wenn alle m Server belegt sind, werden abgewiesen.
- ▶ Klassisches Modell für Analyse von Leitungsvermittlung im Telefonnetz:
 - ▶ Ankunftsrate von Telefongesprächen λ .
 - ▶ Gesprächsdauern exponentialverteilt mit Parameter μ .
 - ▶ Kapazität für m gleichzeitige Telefongespräche.
 - ▶ Anzahl Benutzer ist viel grösser als m .

4/117

Das M/M/m/m-System (2)

Modellierung als Birth-and-Death Prozess:



Wir erhalten:

$$\pi_k = \pi_0 \cdot \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} \quad \text{für } 1 \leq k \leq m$$

Mit $\sum_{k=0}^m \pi_k = 1$ ergibt sich:

$$\pi_0 = \frac{1}{\sum_{k=0}^m \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}}$$

4/118

Das M/M/m/m-System (3)

Die Blockierungswahrscheinlichkeit (W'keit, dass ein neu ankommender Job abgewiesen wird), ist damit:

$$\pi_m = \frac{\left(\frac{\lambda}{\mu}\right)^m \frac{1}{m!}}{\sum_{k=0}^m \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!}}$$

Diese Formel ist als **Erlang B-Formel** bekannt.

Bemerkung: Die Erlang B-Formel gilt auch für M/G/m/m-Systeme (d.h. wenn die Bearbeitungszeiten (Gesprächsdauern) Erwartungswert $1/\mu$ haben, aber ansonsten beliebig verteilt sind).

4/119

Warteschlangen-Netzwerke

- ▶ Warteschlangen-Netzwerke sind Graphen, bei denen die Knoten Warteschlangen-Systeme darstellen (z.B. M/M/1), und gerichtete Kanten die Jobs von einem Knoten zum nächsten führen.
- ▶ Man unterscheidet zwischen **offenen** und **geschlossenen** Warteschlangen-Netzwerken:
 - ▶ Offene Netzwerke erlauben, dass Jobs von aussen zum Netzwerk dazustossen oder das Netzwerk verlassen.
 - ▶ Bei geschlossenen Netzwerken sind die Jobs im Netzwerk gefangen; die Anzahl der Jobs im Netzwerk bleibt deshalb konstant.

4/120

Burke's Theorem

- ▶ Gegeben ein M/M/m ($m = 1, \dots, \infty$) System mit Ankunftsrate λ . Wir nehmen an, dass das System im stationären Zustand gestartet wird. Dann ist der Ausgangsprozess (der Prozess, der das System verlässt) auch ein Poisson-Prozess mit Rate λ .
- ▶ Dank Burke's Theorem kann man direkt Warteschlangen-Netzwerke analysieren.
- ▶ Allerdings muss man vereinfachend annehmen, dass die Servicezeit eines Jobs beim betreten jedes weiteren Warteschlangen-Systems wieder unabhängig ist.
- ▶ Wenn man diese vereinfachende Annahme nicht trifft, kann bisher schon ein einfaches Tandem-System (zwei M/M/1 Systeme in Serie) nicht analysiert werden.

4/121

Jackson's Theorem für offene Netze

- ▶ Jobs kommen bei Knoten j als Poisson-Prozess mit Rate r_j von aussen an.
- ▶ Jobs verlassen Knoten i mit Wahrscheinlichkeit p_{ij} Richtung Knoten j , oder verlassen das Netzwerk mit Wahrscheinlichkeit $p_{i,exit}$, wobei $p_{i,exit} + \sum_{\forall j} p_{ij} = 1$.
- ▶ Dann ist der gesamte Ankunftsprozess bei Knoten j gegeben durch:

$$\lambda_j = r_j + \sum_{\forall i} \lambda_i p_{ij}$$

- ▶ Die Lösung dieses linearen Gleichungssystems ergibt direkt die gesamten Ankunftsrate λ_j .
- ▶ Geschlossene Netze sind etwas komplexer...

4/122

Simulation

- ▶ Kompliziertere und realistischere Warteschlangen-Systeme und -Netzwerke werden in der Regel simuliert. Eine vereinfachte Analyse (z.B. M/M/1) kann aber schon einen ersten Eindruck vermitteln.
- ▶ "And now for something completely different..."

4/123