# Stochastik für die Informatik, Vorlesung 24

## Inhalt

- Invariante Verteilung, deren Bedeutung und Berechnung
- strukturelle Eigenschaften von Markov-Ketten
- Konvergenz gegen die invariante Verteilung

## Lernziele

- Invariante Verteilungen berechnen können
- Die Bedeutung von invarianten Verteilungen kennen

# Erinnerung: Markov-Ketten und stochastische Matrizen

Übergangswahrscheinlichkeiten $p_{a,b} = \mathbb{P}(X_n = b \,|\, X_{n-1} = a)$

Übergangsmatrix $P = (p_{a,b})_{a,b \in S}$

$n-$Schritt-Übergangswahrscheinlichkeiten:
$\mathbb{P}(X_n = b \,| X_0 = a) = p^{(n)}_{a,b}$, wobei $p^{(n)}_{a,b}$ den Eintrag an Zeile $a$ und Spalte $b$ der $n-$ten Matrix-Potenz $P^n$ von $P$ bezeichnet (Matrizenmultiplikation).

Startverteilung $\nu_a = \mathbb{P}(X_0 = a)$

Verteilung nach $n$ Schritten $\mathbb{P}(X_n = b) = \sum_{a \in S} \nu_a p^{(n)}_{a,b}$.

# Invariante Verteilung

(Def. 11.5) Sei $X_0, X_1, \ldots$ eine homogene Markov-Kette mit Zustandsraum $S$. Ein (Spalten-)vektor $(\pi_a)_{a \in S}$ heißt invariante Verteilung für die Markov-Kette $X = (X_n)_{n \geq 0}$, falls gelten

(1) $\pi_a = \sum_{b \in S} \pi_b p_{b,a}$, für alle $a \in S$,

(2) $\pi_a \geq 0$ für alle $a \in S$, und $\sum_{a \in S} \pi_a = 1$.

- Andere Bezeichnungen: Stationäre Verteilung, Gleichgewichtsverteilung
- (2) besagt dass $\pi$ eine Wahrscheinlichkeitsverteilung auf dem Zustandsraum $S$ ist.
- Andere Schreibweise für (1): $\pi^T = \pi^T P$ (Vektor-Matrix-Multiplikation)

# Invariante Verteilung

(Satz 11.4) Falls $\pi$ eine invariante Verteilung für $(X_n)_{n\geq 0}$ ist, und $\nu(a) = \mathbb{P}(X_0 = a) = \pi_a$ für alle $a \in S$ gilt, so gilt für alle $n \in \mathbb{N}$,

$$\mu_n(a) = \mathbb{P}(X_n = a) = \pi_a.$$

(Satz 11.5) Eine homogene Markov-Kette auf einem endlichen Zustandsraum $S$ besitzt immer mindestens eine invariante Verteilung. Invariante Verteilungen sind Lösungen des Gleichungssystems

$$(P - I)^T \pi = 0, \quad \sum_{a\in S} \pi_a = 1.$$

Hier bezeichnet $I$ die Einheitsmatrix der Größe $|S|$,

$$I = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \ddots & \ddots & & \vdots \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}$$

- (Beispiel 11.8, 11.9)

# Beispiel

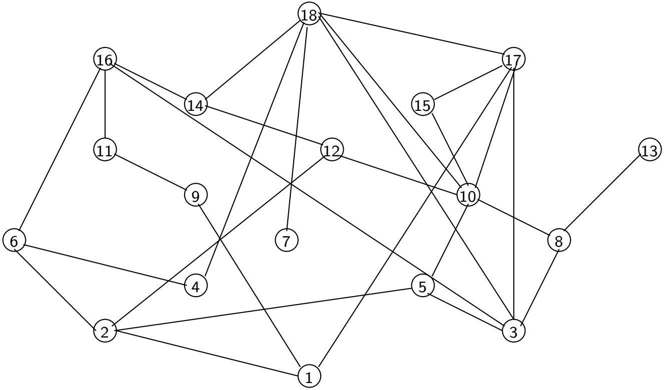



Abbildung: Netzwerk aus Knoten und Kanten

# Übergangsmatrix aus dem Beispiel

$$P = \begin{bmatrix}
0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 \\
1/4 & 0 & 0 & 0 & 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/5 & 0 & 0 & 1/5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/5 & 1/5 & 1/5 \\
0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 \\
0 & 1/3 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1/3 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 & 0 \\
1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1/6 & 0 & 0 & 1/6 & 0 & 0 & 0 & 1/6 & 0 & 0 & 1/6 & 0 & 1/6 & 1/6 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 \\
0 & 1/3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/3 & 0 & 0 & 0 & 1/3 & 0 & 1/3 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 0 \\
0 & 0 & 1/4 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 & 1/4 & 0 & 0 & 1/4 & 0 & 0 & 0 & 0 \\
1/5 & 0 & 1/5 & 0 & 0 & 0 & 0 & 0 & 0 & 1/5 & 0 & 0 & 0 & 0 & 1/5 & 0 & 0 & 1/5 \\
0 & 0 & 1/6 & 1/6 & 0 & 0 & 1/6 & 0 & 0 & 1/6 & 0 & 0 & 0 & 1/6 & 0 & 0 & 1/6 & 0
\end{bmatrix}$$

# Im Beispiel:

Lösung des Gleichungssystems

$$(P - I)^T \pi = 0,$$

$$\pi_1 + \ldots + \pi_{18} = 1$$

ergibt:

$$\pi \approx (0.0526, 0.0702, 0.0877, 0.0351, 0.0526, 0.0526, 0.0175, 0.0526, 0.0351, 0.0877, 0.0351, 0.0526, 0.0175, 0.0526, 0.0351, 0.0702, 0.0877, 0.1053)$$

- $\pi_i$ ist hier proportional zur Anzahl Kanten von Knoten $i$
- Welche Informationen über das Verhalten der Markov-Kette erhalten wir aus der invarianten Verteilung?

# Strukturelle Eigenschaften

(Def. 11.6) Sei $(X_n)_{n\in\mathbb{N}}$ eine Markov-Kette mit Zustandsraum $S$. Eine Folge von Zuständen $(a_0, a_1, ...a_k)$ heißt (guter) Pfad, falls $p_{a_i,a_{i+1}} > 0$ ist für alle $i = 0, ..., k-1$.

Das bedeutet, dass die Kette von $a_0$ nach $a_1$ und weiter nach $a_2, a_3, ...$ bis $a_k$ tatsächlich springen kann.

Eine Markov-Kette heißt irreduzibel, falls für jede Wahl von zwei Zuständen $a, b \in S$ ein Pfad existiert, welcher $a$ und $b$ verbindet, d.h. ein Pfad mit $a_0 = a$ und $a_k = b$. Die Länge $k$ ist dabei egal.

(Def. 11.7) Eine irreduzible Markov-Kette heißt aperiodisch, falls für alle $a \in S$ gilt

$$\mathrm{ggT}\{k \in \mathbb{N} : \exists \text{Pfad der Länge } k \text{ von } a \text{ nach } a\} = 1.$$

(Beispiele)

# Strukturelle Eigenschaften

(Def. 11.6) Sei $(X_n)_{n\in\mathbb{N}}$ eine Markov-Kette mit Zustandsraum $S$. Eine Folge von Zuständen $(a_0, a_1, ...a_k)$ heißt (guter) Pfad, falls $p_{a_i,a_{i+1}} > 0$ ist für alle $i = 0, ..., k-1$.

Das bedeutet, dass die Kette von $a_0$ nach $a_1$ und weiter nach $a_2, a_3, ...$ bis $a_k$ tatsächlich springen kann.

Eine Markov-Kette heißt irreduzibel, falls für jede Wahl von zwei Zuständen $a, b \in S$ ein Pfad existiert, welcher $a$ und $b$ verbindet, d.h. ein Pfad mit $a_0 = a$ und $a_k = b$. Die Länge $k$ ist dabei egal.

(Def. 11.7) Eine irreduzible Markov-Kette heißt aperiodisch, falls für alle $a \in S$ gilt

$$\mathrm{ggT}\{k \in \mathbb{N} : \exists \text{Pfad der Länge } k \text{ von } a \text{ nach } a\} = 1.$$

(Beispiele)

# Existenz und Eindeutigkeit von invarianten Verteilungen

(Satz 11.5) Eine homogene Markov-Kette auf einem endlichen Zustandsraum $S$ besitzt immer mindestens eine invariante Verteilung.

(Satz 11.6) Falls die Markov-Kette im obigen Fall irreduzibel ist, so ist die invariante Verteilung eindeutig.

- (Bem.) unendlicher Zustandsraum
- Beispiel

# Konvergenz gegen die invariante Verteilung

(Theorem 11.7) Sei $(X_i)_{i\in\mathbb{N}_0}$ eine irreduzible und aperiodische homogene Markov-Kette mit Zustandsraum $S$. Sei $(\pi_a)_{a\in S}$ eine invariante Verteilung. Dann gilt für alle $a \in S$, unabhängig von der Startverteilung,

$$\lim_{n\to\infty} \mathbb{P}(X_n = a) = \pi_a.$$

- Egal wo man startet: Für große $n$ gilt $\mathbb{P}(X_n = a) \approx \pi_a$.
- Falls die Kette in einer (beliebigen) Startverteilung $\nu$ startet, befindet sie sich nach hinreichend langer Zeit in der stationären Verteilung
- Information über Struktur der Kette. Bsp. Netzwerk: Knoten mit vielen Kanten werden oft besucht, für solche Knoten $a$ ist $\pi_a$ größer als für solche die selten besucht werden.

# Konvergenz gegen die invariante Verteilung

(Theorem 11.7) Sei $(X_i)_{i\in\mathbb{N}_0}$ eine irreduzible und aperiodische homogene Markov-Kette mit Zustandsraum $S$. Sei $(\pi_a)_{a\in S}$ eine invariante Verteilung. Dann gilt für alle $a \in S$, unabhängig von der Startverteilung,

$$\lim_{n\to\infty} \mathbb{P}(X_n = a) = \pi_a.$$

- Egal wo man startet: Für große $n$ gilt $\mathbb{P}(X_n = a) \approx \pi_a$.
- Falls die Kette in einer (beliebigen) Startverteilung $\nu$ startet, befindet sie sich nach hinreichend langer Zeit in der stationären Verteilung
- Information über Struktur der Kette. Bsp. Netzwerk: Knoten mit vielen Kanten werden oft besucht, für solche Knoten $a$ ist $\pi_a$ größer als für solche die selten besucht werden.

# Invariante Verteilung

(Satz 11.8) Sei $(X_n)_{n\in\mathbb{N}_0}$ eine irreduzible und aperiodische homogene Markov-Kette mit Zustandsraum $S$ und invarianter Verteilung $\pi$. Sei

$$T_i := \inf\{n \geq 1 : X_n = i\}$$

der erste Zeitpunkt (nach dem Start), zu dem sich die Kette im Zustand $i$ befindet. Dann gilt für die invariante Verteilung $\pi$,

$$\pi_i = \frac{1}{\mathbb{E}[T_i \mid X_0 = i]}.$$

Das heit, die erwartete erste Rückkehrzeit zu $i$ bei Start in $i$ ist gegeben durch das Reziproke der invarianten Verteilung.

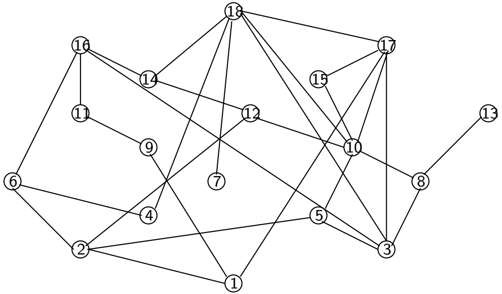
- Beweis siehe z.B. Kapitel 7 von *Mitzenmacher, Upfal: Probability and Computing*

# Invariante Verteilung: Informationen über die Struktur der Kette

- Falls die Kette in einer (beliebigen) Startverteilung $\nu$ startet, befindet sie sich nach hinreichend langer Zeit in der stationären Verteilung
- Information über Struktur der Kette. Bsp. Netzwerk: Knoten mit vielen Kanten werden oft besucht, für solche Knoten $a$ ist $\pi_a$ größer als für solche die selten besucht werden.
- Je größer $\pi_i$, desto kürzer die Zeit bis die Kette zu $i$ zurückkehrt

# Beispiel 11.11: Page Rank

Zufallsgraph vom letzten Mal:




Invariante Verteilung:

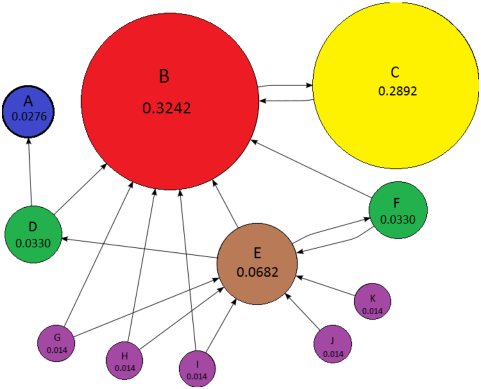
$$\pi^T \approx (0.0517, 0.0690, 0.0862, 0.0345, 0.0517, 0.0517, 0.0172, 0.0517, 0.0345, \\ 0.1034, 0.0345, 0.0517, 0.0172, 0.0517, 0.0345, 0.0690, 0.0862, 0.1034)$$

- Die Knoten mit großen Werten von $\pi$ werden häufiger besucht als die anderen (weil sie mehr Nachbarn haben): Sie sind wichtiger da sie mehr Verknüpfungen besitzen
- Grundprinzip des PageRank Algorithmus

# Beispiel 11.12: Page Rank

- Bewertung/Gewichtung von Dokumenten mit Verknüpfungen, z.B. Webseiten mit Links
- $k$ Webseiten, $S = \{1, ..., k\}$
- $c_i$ =Anzahl Links von Seite $i \in S$ ausgehend (gerichteter Graph)
- $$q_{ij} := \frac{1_{\{i \text{ verlinkt auf } j\}}}{c_i} = \frac{1_{\{i \to j\}}}{c_i} = \begin{cases} \frac{1}{c_i} & \text{falls } i \to j \\ 0 & \text{sonst.} \end{cases}$$
- $Q = (q_{ij})_{i,j \in S}$ ist die Übergangsmatrix für eine Markov-Kette auf dem (gerichteten) Graphen, der durch die Webseiten und die Links beschrieben wird, d.h. $q_{ij}$ ist die Wahrscheinlichkeit mit der man von $i$ nach $j$ springt.
- Grundidee: Finde invariante Verteilung von $Q$. Je größer deren $i-$te Komponente, desto wichtiger ist die Seite $i$.

# Beispiel 11.12: Page Rank






- Praktische Berechnung: Z.B. Jacobi-Iteration
- Zufallssurfermodell, Erweiterungen

# Beispiel 11.12: Page Rank

- Problem: Seiten, auf die keine andere Seite verlinkt
- Erweiterung: Sei $d \in (0,1)$ gegeben
- $$p_{ij} := dq_{ij} + \frac{1-d}{k}$$
  d.h. mit Wahrscheinlichkeit $d$ folgt man den Links, mit Wahrscheinlichkeit $1-d$ springt man zu einer zufällig ausgewählten Seite.
- Def. PageRank der Seite $i$ : $PR_i := \tilde{\pi}_i$, wobei $\tilde{\pi}$ die invariante Verteilung von $P = (p_{ij})_{i,j=1,\dots,k}$ ist
- $PR$ ist also die (normierte) Lösung des Gleichungssystems
  $$PR_i = \frac{1-d}{k} + d\sum_{j=1}^{k} \frac{1_{\{j \to i\}}}{c_j} PR_j$$
- Referenz: Sergei Brin, Lawrence Page: The Anatomy of a Large-Scale Hypertextual Web Search Engine. In: Computer Networks and ISDN Systems, Band 30, 1998, S. 107-117

# Beispiel 11.13 Warteschlangen

- Zeitabschnitte von fester Länge
- Parameter $k \in \mathbb{N}, \mu \in [0,1], \lambda \in [0,1]$, mit $\lambda + \mu \leq 1$.
- In jedem Zeitabschnitt geschieht eines der folgenden Ereignisse:
  - Falls weniger als $k$ Kunden da sind, so kommt ein neuer Kunde mit Wahrscheinlichkeit $\lambda$ an
  - Falls mindestens ein Kunde da ist, so verlässt ein Kunde die Schlange mit Wahrscheinlichkeit $\mu$
  - Andernfalls geschieht nichts
- $X_n$:=Anzahl Kunden in der Schlange im $n-$ten Zeitabschnitt
- $(X_n)_{n \in \mathbb{N}_0}$ ist Markov-Kette mit Zustandsraum $S = \{0, ..., k\}$

# Beispiel 11.13 Warteschlangen

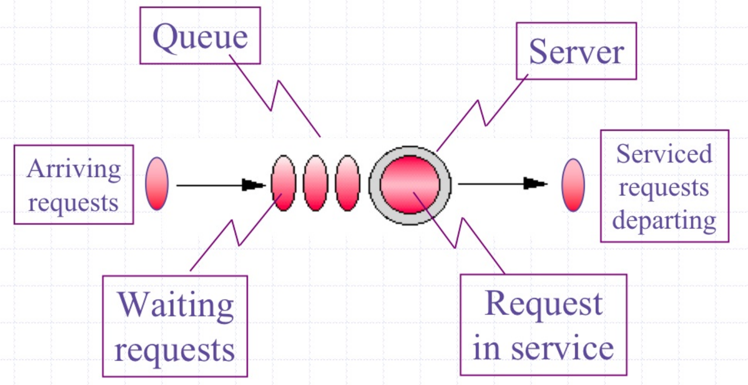



Bild: http://www.perfdynamics.com/

# Beispiel 7.12 Warteschlangen

Mögliche Übergänge:

$$
\begin{aligned}
\color{blue}{p_{i,i+1}} &= \lambda \text{ falls } i < k \\
\color{blue}{p_{i,i-1}} &= \mu \text{ falls } i > 0.
\end{aligned}
$$

d.h.

$$
P = \begin{bmatrix}
1-\lambda & \lambda & 0 & \cdots & & & \cdots & 0 \\
\mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & & 0 \\
0 & \mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & 0 \\
\vdots & & \ddots & \ddots & \ddots & & & \vdots \\
0 & \cdots & & & 0 & \mu & 1-\mu-\lambda & \lambda \\
0 & \cdots & & & & 0 & \mu & 1-\mu
\end{bmatrix}
$$

Frage: Wie viele Kunden sind auf lange Sicht in der Schlange?
Stabilität für $k \to \infty$?

# Beispiel 7.12 Warteschlangen

Mögliche Übergänge:

$$\begin{aligned} p_{i,i+1} =& \lambda \text{ falls } i < k \\ p_{i,i-1} =& \mu \text{ falls } i > 0. \end{aligned}$$

d.h.

$$P = \begin{bmatrix} 1-\lambda & \lambda & 0 & \cdots & & & \cdots & 0 \\ \mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & & 0 \\ 0 & \mu & 1-\mu-\lambda & \lambda & 0 & \cdots & & 0 \\ \vdots & & \ddots & \ddots & \ddots & & & \vdots \\ 0 & \cdots & & & 0 & \mu & 1-\mu-\lambda & \lambda \\ 0 & \cdots & & & & 0 & \mu & 1-\mu \end{bmatrix}$$

Frage: Wie viele Kunden sind auf lange Sicht in der Schlange?
Stabilität für $k \to \infty$?

# Beispiel 11.13 Warteschlangen

(Satz 11.9) Die invariante Verteilung der beschriebenen Warteschlange ist gegeben durch

$$\pi_i = \frac{1}{\sum_{j=0}^{k} (\lambda/\mu)^j} \cdot \left(\frac{\lambda}{\mu}\right)^i, \quad i = 0, \ldots, k$$

Was passiert für $k \to \infty$?

- Falls $\lambda/\mu < 1$ ist, gilt $\sum_{j=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^j = \frac{\mu}{\mu-\lambda} < \infty$, die Warteschlange hat für $k \to \infty$ eine invariante Verteilung, ist stabil
- Andernfalls ist $\sum_{j=0}^{\infty} \left(\frac{\lambda}{\mu}\right)^j = \infty$, die Warteschlange ist instabil