Vorlesung 10 Markov Prozesse cont'd

Erinnerung
- Ein Markov-Prozess X auf einem Graphen G heißt <u>aperiodisch</u>, falls:
$$ggT(\{ k \in \mathbb{N} \mid (P^k)_{uu} > 0 \}) = 1$$
für alle $u \in V$ (P ist die Übergangsmatrix von X).
- X heißt irreduzibel, falls für alle $u, v \in V$ ein $t \in \mathbb{N}$ existiert mit
$$(P^t)_{uv} > 0.$$

Theorem 10.1

Sei X ein aperiodischer und irreduzibler Markov-Prozess auf G. Sei P die Übergangsmatrix von X.

1. X hat eine eindeutige stationäre Verteilung $\pi : V \to \mathbb{R}$. (d.h. $P\pi = \pi$, $\pi(u) \geq 0$ für alle u, $\sum_{u \in V} \pi(u) = 1$).
2. $\lim_{k \to \infty} P^k = \pi e^T$, $e = (1, \ldots, 1)^T \in \mathbb{R}^n$, $n = |V|$.
3. Für alle Whk.-Verteilungen $f : V \to \mathbb{R}$ gilt: $\lim_{k \to \infty} (P^k \cdot f) = \pi$

Wir brauchen für den Beweis zwei Hilfsresultate.

Proposition 10.2 (Perron-Frobenius)
Sei $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ eine Matrix mit $a_{ij} > 0$ für alle i,j und $A^T e = e$, $e = (1, \ldots, 1)^T \in \mathbb{R}^n$. Dann ist 1 ein <u>einfacher</u> Eigenwert von A und alle anderen Eigenwerte $\lambda \in \mathbb{C}$ von A erfüllen $|\lambda| < 1$.

Beweis

Da $A^T e = e$, hat A^T den Eigenwert 1.

Daher hat A den Eigenwert 1.

Sei $k \geq 1$ fest und $M := A^k$, $M = (m_{ij})$.

Dann gilt:
- $m_{ij} > 0$ für alle i,j
- $\max\limits_{1 \leq i,j \leq n} m_{ij} < 1$, weil $M^T e = \left[\sum_{i=1}^{n} m_{ij} \right]_{j=1}^{n}$

Angenommen $\lambda \in \mathbb{C}$ ist kein einfacher Eigenwert von A. Dann hat M ein Jordan-Block der Form:

$$\begin{bmatrix} \lambda & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda \end{bmatrix}^k = \begin{bmatrix} \lambda^k & k\lambda^{k-1} & & 0 \\ & \ddots & \ddots & \\ & & \ddots & k\lambda^{k-1} \\ 0 & & & \lambda^k \end{bmatrix}$$

Wenn A diag.bar: $A = SDS^{-1}$, $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$

I.A. haben wir Jordan-Zerlegung $A = SDS^{-1}$ $D = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_n \end{pmatrix}$

$B_i = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix}$

$\Rightarrow$ Da die Einträge von M beschränkt sind, muss gelten:
- $\lambda = 1$ ist einfacher Eigenwert
- alle anderen Eigenwerte müssen $|\lambda| \leq 1$ erfüllen.

Sei nun $\lambda \in \mathbb{C}$, $\lambda \neq 1$, ein Eigenwert von A^T. Sei $x = (x_1, \dots, x_n)^T \in \mathbb{C}^n$ ein Eigenvektor von λ: $A^T x = \lambda x$. Dann gilt:

$$\lambda x_i = a_{1i} x_1 + \dots + a_{ni} x_n. \quad \text{für } 1 \leq i \leq n.$$

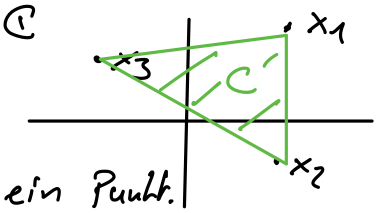
Es gilt:

$$a_{1i} + \dots + a_{ni} = 1, \quad \text{weil } A^T e = e. \text{ und } a_{ij} > 0$$

D.h. λx_i ist eine Konvex-Kombination von $x_1, \dots, x_n$.

Sei C die konvexe Hülle von $x_1, \dots, x_n$ $\mathbb{C}$

D.h. $\lambda x_i \in C$ für $1 \leq i \leq n$.



Es gibt $i \neq j$ mit $x_i \neq x_j$. $\Rightarrow$ C ist nicht nur ein Punkt.

Da $a_{ij} > 0$ für alle i,j, muss λx_i im relativen Inneren von C liegen. Daraus folgt:

$$|\lambda x_i| < \max_{1 \le j \le n} |x_j| \quad \text{für alle } i.$$

$\Rightarrow |\lambda| < 1$. $\square$

Jetzt: Sei $S \subset \mathbb{N}$ eine Teilmenge, s.d. für alle $r, s \in S$: $r+s \in S$. Dann nennen wir S eine Semigruppe.

<u>Lemma</u> 10.3

Sei $S \subset \mathbb{N}$ eine Semigruppe, s.d. $ggT(S) = 1$. Dann ist $\mathbb{N} \setminus S$ endlich.

<u>Beweis</u>

siehe Skript.

<u>Beweis von Theorem 10.1</u>

Wir müssen zeigen, dass:

1. X hat eine eindeutige stationäre Verteilung $\pi: V \to \mathbb{R}$.
   (d.h. $P\pi = \pi$, $\pi(u) \geq 0$ für alle u, $\sum_{u \in V} \pi(u) = 1$).
2. $\lim_{k \to \infty} P^k = \pi e^T$, $e = (1, \dots, 1)^T \in \mathbb{R}^n$, $n = |V|$.
3. Für alle Whk.-Verteilungen $f: V \to \mathbb{R}$ gilt: $\lim_{k \to \infty} (P^k f) = \pi$

Behauptung: Unter der Annahme des Theorems gilt: es existiert $M \in \mathbb{N}$ mit $(P^m)_{uv} > 0$ für alle $u, v \in V$ und $m \geq M$.

Wir zeigen erst wie 1.+2.+3. aus der Behauptung folgen.

Wir wissen, dass $P^T e = e$. Daher: $(P^T)^m e = e$ für alle $m \in \mathbb{N}$.

D.h. für $m \geq M$ gilt nach Prop. 10.2:

- 1 ist einfacher Eigenwert von P^m
- alle anderen Eigenwerte von P^m erfüllen $|\lambda| < 1$.

Daraus folgt, da (λ EW von $P \Rightarrow \lambda^m$ EW von P^m)

a) 1 ist einfacher Eigenwert von P

b) alle anderen Eigenwerte von P erfüllen $|\lambda| < 1$.

Aus a) folgt: P hat einen <u>eindeutigen</u> Rechtseigenvektor $\pi \in \mathbb{R}^n$ mit

$$P\pi = \pi.$$

Sei nun $\lambda \neq 1$ ein Eigenwert von P. Dann gilt nach b) $\lim_{k \to \infty} \lambda^k = 0$.

D.h. P^k für $k \to \infty$ konvergiert gegen eine Rang-1-Matrix

$$P^* := \lim_{k \to \infty} P^k = xy^T \quad \text{mit } x, y \in \mathbb{R}^n.$$

Es gilt $P^k \pi = \pi$ für alle k, $(P^T)^k e = e$ für alle k

$$\Rightarrow \quad P^* \pi = \pi \quad \text{und} \quad P^* e = e$$

$$\Rightarrow \quad \langle y, \pi \rangle x = \pi \quad \text{und} \quad \langle x, e \rangle y = e.$$

$$\Rightarrow \quad \langle y, \pi \rangle \neq 0 \quad \text{und} \quad \langle x, e \rangle \neq 0.$$

Wir können annehmen, dass $\langle x, e \rangle = 1$. $\Rightarrow y = e \Rightarrow \langle e, \pi \rangle x = \pi$.

Daher: $\langle e, \pi \rangle \neq 0$ und wir können annehmen, dass

$$\langle e, \pi \rangle = 1.$$

$$\Rightarrow \quad \lim_{k \to \infty} P^k = \pi e^T \quad \text{(Das zeigt 2.)}$$

Daraus folgt auch $\pi(u) > 0$ für alle $u \in V$. (Das zeigt 1.)

Sei nun $f: V \to \mathbb{R}$ eine Wkt.-Verteilung. Dann gilt:

$$\lim_{k\to\infty} P^k f = \pi e^T f = \pi \quad , \text{ da } \langle e, f\rangle = e^T f = 1.$$

(Dies zeigt 3.).

Wir beweisen die Behauptung:

Sei dazu $u \in V$ und

$$S_u := \{ k \in \mathbb{N} \mid (P^k)_{uu} > 0 \}$$

Da X aperiodisch, gilt: $\operatorname{ggT}(S_u) = 1$.

Falls $\alpha, \beta \in S_u$, so auch $\alpha + \beta \in S_u$. D.h. S_u ist eine Semigruppe.

$\to$ $\mathbb{N} \setminus S_u$ ist endlich

$\to$ es existiert M_u mit $(P^m)_{uu} > 0$ für alle $m \geq M_u$.

Wir setzen

$$M' := \max_{u \in V} M_u.$$

Sei nun $u, v \in V$, $u \neq v$. Definiere für $t \geq 1$:

$$l_{uv}^t := P(X_t = u \text{ und } X_i \neq u \text{ für } i < t \mid X_0 = v).$$

= Wkt. zum ersten Mal nach t Schritten bei u zu landen, wenn wir bei v starten.

Da X irreduzibel, finden wir $R > 0$ mit

$$0 < \sum_{t=1}^{R} l_{uv}^t$$

Außerdem gilt für $k \geq 1$:

$$(P^k)_{uv} = \sum_{t=1}^{k} l_{uv}^t \, (P^{k-t})_{uu}$$

Dann setzen wir $M_1 = M' + R$.

Dann gilt für $m \geq M$: $(P^m)_{uu} > 0$ für alle $u \in V$, da $M > M'$

$$(P^m)_{uv} = \sum_{t=1}^{m} l_{uv}^t \, (P^{m-t})_{uu}$$

$$\geq \sum_{t=1}^{R} l_{uv}^t \, (P^{m-t})_{uu}$$

$$> \sum_{t=1}^{R} l_{uv}^t \quad \varepsilon \quad > 0 .$$

wobei $\varepsilon = \min_{1 \leq t \leq R} (P^{m-t})_{uu} > 0$.

$\square$

Theorem 10.4

Sei $G = (V, E)$ ein zusammenhängender Graph mit $n = |V|$.

und $\pi : V \to \mathbb{R}$ eine Wkt.-Verteilung und $d > \max_{v \in V} \deg(v)$.

Sei $P = (p_{uv}) \in \mathbb{R}^{n \times n}$ definiert durch

$$p_{uv} = \begin{cases} \frac{1}{d} \cdot \min\left\{1, \frac{\pi(u)}{\pi(v)}\right\}, & \text{falls } u \neq v \text{ und } \{u,v\} \in E \\ 0, & \text{falls } u \neq v \text{ und } \{u,v\} \notin E. \\ 1 - \sum_{i \neq v} p_{iv}, & \text{falls } u = v. \end{cases}$$

Dann ist P die Übergangsmatrix eines aperiodischen und irreduziblen Markov-Prozesses mit stat. Verteilung π.

Beweis

Übung.

Theorem 10.4 motiviert einen Algorithmus zum Ziehen von Zufallszahlen einer Verteilung $\pi : V \to \mathbb{R}$. Der Algorithmus heißt Metropolis-Hastings-Algorithmus.

Beispiel 10.5

Sei $G=(V,E)$ ein Graph. Eine Funktion $F: V \to \{1,\dots,k\}$ ein k-Coloring von G. Ein k-Coloring ist <u>zulässig</u>, falls $F(v) \neq F(u)$ für alle $u,v \in V$.

Die Anzahl k-Colorings zu berechnen ist i.A. schwer.

Trotzdem können wir von allen k-Colorings uniform samplen.

Wir definieren $\hat{G}=(\hat{V},\hat{E})$ mit:

$$\hat{V} = \{ F: V \to \{1,\dots,k\} \mid F \text{ ist zulässig} \}$$

$$\hat{E} = \{ \{F,G\} \subset \hat{V} \mid F \text{ und } G \text{ unterscheiden sich in genau einem Knoten} \}$$

Dann: $\hat{G}$ zusammenhängend, wenn $k \geq \max_{v \in V} \deg(v) + 2$.

Dann können wir mit Metropolis-Hastings von $\pi: \hat{V} \to \mathbb{R}$

$$\pi(a) = \frac{1}{|\hat{V}|}$$

ziehen ohne $|\hat{V}|$ zu kennen.