Konvergenz von Markovketten

Dieser Vortrag basiert im Wesentlichen auf §6.1 und §6.2 des Buchs

R. Durrett: Random Graph Dynamics, Cambridge University Press, 2007.

Definition 1. Eine Markovkette in diskreter Zeit mit endlichem Zustandsraum Ω, $|\Omega| = n$ und Übergangsmatrix K ist eine Folge von Zufallsvariablen $\{X_t\}_{t\in\mathbb{N}}$ mit Werten in Ω und der folgenden Eigenschaft

$$P(X_{t+1} = x_{t+1}|X_t = x_t, \ldots, X_0 = x_0) = P(X_{t+1} = x_{t+1}|X_t = x_t) = K(x_t, x_{t+1}),$$

Das heißt die bedingte Verteilung zu einem Zeitpunkt hängt nur vom vorherigen Zeitpunkt ab, jedoch nicht von der gesamten Vergangenheit. Desweiteren soll obige Gleichung bedeuten, dass die Einschritt-Übergangswahrscheinlichkeit nicht von der Zeit abhängt. Aufgrund dieser Zeitunabhängigkeit gilt für die n-Schritt Übergangswahrscheinlichkeit

$$P(X_n = y|X_0 = x) = K^n(x, y).$$

Definition 2. Eine Markovkette mit Übergangsmatrix K heißt irreduzibel falls für alle $x, y \in \Omega$ eine Zeit $n \in \mathbb{N}$ existiert, sodass $K^n(x, y) > 0$, das heißt x geht nach hinreichend langer Zeit mit positiver Wahrscheinlichkeit in den Zustand y über.

Stellt man die Markovkette als Graphen mit $V(G) = \Omega$ und $E(G) = \{(x, y) \in \Omega^2 : K(x, y) > 0\}$ dar, dann heißt Irreduzibilität, dass G stark zusammenhängend ist. Die Übergangsmatrizen K von Markovketten haben die Eigenschaft, dass ihre Zeilensumme gleich eins ist und dass ihre Einträge nichtnegativ sind. Solche Matrizen nennt man *stochastisch.*
Damit ist $(1, \ldots, 1)^T$ ein Eigenvektor zum Eigenwert $\lambda = 1$ und man kann sich mittels diagonaldominanz überlegen, dass $(K - \lambda)^{-1}$ für alle $\lambda \in \mathbb{C}$ mit $|\lambda| > 1$ existiert. Es gibt also keinen Eigenwert mit Betrag größer als eins. Aus dem Satz von Perron-Frobenius folgt für irreduzible Markovketten, dass $\lambda = 1$ der betragsmäßig größte Eigenwert ist mit *algebraischer* Vielfachheit eins.

Satz 3 (Perron-Frobenius). Ist K die Übergangsmatrix einer irreduziblen Markovkette, dann ist $\lambda = 1$ ein algebraisch einfacher Linkseigenwert von K mit Linkseigenvektor π, sodass $\pi(x) > 0$ für alle $x \in \Omega$.

Erfüllt die Verteilung $\{\pi(x)\}_{x\in\Omega}$ für eine Markovkette mit Übergangsmatrix K die *Detailed-Blance-Bedingung*

$$\pi(x)K(x, y) = \pi(y)K(y, x), \quad x, y \in \Omega,$$

so heißt die Markovkette *reversibel.* Die Bedeutung dieser Bedingung erkennt man nach Einführen des gewichteten inneren Produktes $< f, g >_\pi := \sum_{x\in\Omega} \pi(x)f(x)g(x)$, denn es gilt

$$< Kf, g >_\pi = \sum_{x,y\in\Omega} \pi(x)K(x, y)f(x)g(y) = \sum_{x,y\in\Omega} \pi(y)K(y, x)f(x)g(y) = < f, Kg >_\pi .$$

Das heißt bezüglich dieses inneren Produktes ist K symmetrisch. Man beachte, dass für irreduzible Markovketten nach dam Satz von Perron-Frobenius $\pi(x) > 0$ für alle $x \in \Omega$

folgt. In diesem Fall ist $< .,. >_\pi$ ein Skalarprodukt. Der entsprechende Operator im $\mathbb{R}^n$ mit dem üblichen Skalarprodukt ist

$$D^{1/2}KD^{-1/2}, \quad D = diag(\pi(x))_{x\in\Omega}.$$

Aus der Detailed-Blance-Bedingung folgt außerdem

$$\sum_{y\in\Omega} \pi(y)K(y,x) = \sum_{y\in\Omega} \pi(x)K(x,y) = \pi(x), \quad x \in \Omega,$$

denn die Zeilensumme in K ist gleich eins. Somit ist π aufgefasst als Zeilenvektor ein Linkseigenvektor von K zum Eigenwert 1 und es scheint naheliegend π als *stationäre Verteilung* dieser Markovkette zu bezeichnen.
Die symmetrische Matrix $D^{1/2}KD^{-1/2}$ besitzt die reellen Eigenwerte $1 = \lambda_0 \geq \lambda_1 \geq \ldots \geq \lambda_{n-1} \geq -1$. Ist die Markovkette irreduzibel, so folgt aus dem Satz von Perron-Frobenius sogar $\lambda_1 < 1$.
Die Verteilung einer irreduziblen Markovkette konvergiert für $t \to \infty$ gegen die stationäre Verteilung und die Konvergenzgeschwindigkeit gegen die stationäre Verteilung π wird durch den betragsmäßig zweitgrößten Eigenwert $\lambda_{max} := \{\lambda_1, |\lambda_{n-1}|\}$ bestimmt.

Satz 4. Eine irreduzible reversible Markovkette mit Übergangsmatrix K erfüllt

$$\max_x \sum_y |K^t(x,y) - \pi(y)| \leq \max_{x,y} \left| \frac{K^t(x,y)}{\pi(y)} - 1 \right| \leq \frac{\lambda_{\max}^t}{\min_{x\in\Omega} \pi(x)}.$$

Erläuterung: Der zweitgrößte Eigenwert dominiert, denn oben betrachtet man

$$\|xK^t - \pi\| \sim \|xD^{-1/2}D^{1/2}K^tD^{-1/2} - \pi D^{-1/2}\| = \|D^{1/2}K^tD^{-1/2}D^{-1/2}x^T - D^{-1/2}\pi^T\|.$$

Also liegt der Eigenvektor $x^T = \pi^T$ zum größten Eigenwert $\lambda_0 = 1$ im Kern obigen Matrix.
Für einen elementaren Beweis dieses Satzes verweisen wir auf Durret.
Wir wollen nun die Konvergenz und ihre Geschwindigkeit im zeitkontinuierlichen Fall untersuchen. Hierfür müssen wir zunächst eine Übergangsmatrix für Markovketten mit kontinuierlicher Zeit einführen.
Hierfür betrachten wir einen Poisson-Prozess mit der Rate $\lambda = 1$, um die Verteilung der Sprünge bis zur Zeit t zu beschreiben. Die erwartete Anzahl der Sprünge zur Zeit t ist dann genau t selbst und wir erhalten somit für die Übergangsmatrix

$$H_t(x,y) = e^{-t} \sum_{m=0}^{\infty} \frac{t^m}{m!} K^m(x,y) = e^{-(I-K)t}, \quad t \in \mathbb{R}_+.$$

Für den Analytiker steht hier einfach die von $I - K$ erzeugte Halbgruppe $\{e^{-(I-K)t}\}_{t\geq 0}$. Bezüglich des oben definierten Skalarproduktes $< .,. >_\pi$ ist der Erzeuger $I - K$ symmetrisch, das heißt selbstadjungiert, und nichtnegativ, damit besteht die Halbgruppe sogar aus Kontraktionen. Beachte, dass nach dem Spektralabbildungssatz für Matrizen

$$\begin{aligned} \lambda \in \sigma(K) \quad &\Longleftrightarrow \quad e^{-t+\lambda t} \in \sigma(e^{-(I-K)t}), \quad t \in \mathbb{R}_+ \\ \lambda \in \sigma(D^{1/2}KD^{-1/2}) \quad &\Longleftrightarrow \quad e^{-t+\lambda t} \in \sigma(e^{-(I-D^{1/2}KD^{-1/2})t}), \quad t \in \mathbb{R}_+ \end{aligned}$$

gilt, die Jordanketten übertragen sich ebenfalls entsprechend. Im Gegensatz zur Behauptung des Autors kann auch H_t komplexe Eigenwerte haben, da K solche besitzen kann. Der zweitgrößte Eigenwert von H_t ist jedoch $e^{-t+\lambda_1 t}$. Wie in der zeitdiskreten Version bestimmt der zweitgrößte Eigenwert der betrachteten Matrix, hier H_t, die Konvergenz.

Satz 5. Ist $\beta = 1 - \lambda_1$ und $\lambda_{\max} = \lambda_1$, dann gilt die Abschätzung

$$\max_{x,y\in\Omega} \left| \frac{H_t(x,y)}{\pi(y)} - 1 \right| \le \frac{e^{-\beta t}}{\min_{x\in\Omega} \pi(x)}.$$

Beweisidee:
Man betrachtet die Ableitung $\frac{\partial}{\partial t} H_t(x,y)$ der Operator-Exponentialfunktion H_t, und schätzt diese nach oben ab. Man erhält hieraus eine obere Schranke für H_t und dann eine Abschätzung für die oben angegebene Differenz. Nutze auch Variationscharakterisierung der Spektrallücke. Nette Randbemerkung im Beweis: Zeitumkehr der Markovkette entspricht Adjunktion von H_t bezüglich $< .,. >_\pi$.

Die Größe $\beta = 1 - \lambda_{max}$ heißt *Spektrallücke.*
Wir definieren die Leitfähigkeit einer reversiblen irreduziblen Markovkette durch

$$h = \min_{\pi(S)\le 1/2} \frac{\sum_{x\in S, y\in S^c} \pi(x)K(x,y)}{\pi(S)}, \quad \pi(S) := \sum_{s\in S} \pi(s), \quad S \subseteq \Omega.$$

Da sich der Eigenwert λ_{max} theoretisch nur schlecht bestimmen lässt, ist es gebräuchlich diesen mit Hilfe der Leitfähigkeit h abzuschätzen. In Anlehnung an die Eigenwertabschätzung des Laplace-Beltrami-Operators auf Riemannschen Mannigfaltigkeiten wurde die folgende Ungleichung nach Cheeger benannt.

Satz 6. Die Spektrallücke erfüllt

$$\frac{h^2}{2} \le 1 - \lambda_1 \le 2h.$$

Beweisidee: Nutze die Variationscharakterisierung von $1-\lambda_1$ (vgl. Ergänzung am Ende) und konstruiere geschickt Funktionen f für diese Charakterisierung.

Wir wollen das Ganze nun auf Graphen anwenden. Sei $G = (V, E)$ ein Graph und $d(x)$ der Grad der Ecke x in G. Für $\Omega = V(G)$ definieren wir die Übergangsmatrix K durch

$$K(x,y) := \begin{cases} 1/2, & \text{falls } x = y, \\ \frac{1}{2d(x)}, & \text{falls } x \sim y, \\ 0, & \text{sonst.} \end{cases} \tag{1}$$

Eine Markovkette mit $K(x,x) \ge 1/2$ für alle $x \in \Omega$ heißt *faul.* Insbesondere gilt $K = (I + p)/2$, für eine weitere Übergangsmatrix p, also ist $\sigma(D^{1/2}KD^{-1/2}) \subset [0,1]$ und $\lambda_{max} = \lambda_1$. Durch $\pi(x) := d(x)/D$ mit $D = \sum_{y\in V(G)} d(y)$ wird eine stationäre Verteilung definiert, welche

$$\pi(x)K(x,y) = \frac{1}{2D} = \pi(y)K(y,x)$$

also die Detailed-Balance-Bedingung erfüllt.

Definition 7. Die Mischungszeit einer Markovkette ist für ein hinreichend kleines $\varepsilon > 0$ und mit der Totalvariation

$$d_{TV}(K^t(x,\cdot),\pi) := \sum_y |K^t(x,y) - \pi(y)|$$

definiert als

$$T_{mix} := \max_x \min_t \{d_{TV}(K^t(x,\cdot),\pi) \le \varepsilon\}.$$

Die konkrete Wahl von ε spielt für uns keine große Rolle, es ändert nur eine Konstante in der Zeit. Da wir im Folgenden die Mischungszeit in $\mathcal{O}(\log n)$ abschätzen ist der Wert dieser Konstante egal.

Wir wollen nun die Mischungszeit von Zufallsgraphen aus $ER(n,(c\log n)/n)$ mit $c > 1$ untersuchen. In Theorem 2.8.1 wurde gezeigt, dass die Graphen aus dieser Klasse mit hoher Wahrscheinlichkeit für große n zusammenhängend sind. Damit ist das Setting sinnvoll und wir können die obigen Resultate anwenden.

Satz 8. Graphen aus $ER(n,(c\log n)/n)$, $c > 1$ mit der in (1) definierten faulen Markovkette K haben eine Mischungszeit T_{mix} der Ordnung $\mathcal{O}(\log n)$.

Beweisidee:
Bestimme Schranken für den Eckengrad d und folgere hieraus Schranken für die Leitfähigkeit h. Nun gibt die Cheeger-Ungleichung eine untere Schranke für die Spektrallücke $\beta = 1-\lambda_1$, woraus man eine Abschätzung für die Totalvariation und damit die Mischzeit erhält.

Ergänzung: Dirichlet-Formen, Rayleigh-Ritz-Prinzip

Zu $f,g \in \mathbb{R}^n$ betrachten wir die Dirichlet-Form

$$\mathcal{E}(f,g) = \frac{1}{2}\sum_{x,y\in\Omega}(f(x)-f(y))(g(x)-g(y))\pi(x)K(x,y).$$

Mit Hilfe des gewichteten Skalarprodukt $< .,. >_\pi$ lässt sich die Bilinearform durch einen Operator darstellen:

$$\mathcal{E}(f,f) =< f,(I-K)f >_\pi .$$

Der entsprechende symmetrische Operator bezüglich des ursprünglichen Skalarproduktes ist

$$D^{1/2}(I-K)D^{-1/2} = I - D^{1/2}KD^{-1/2}.$$

Nun kann man definieren, dass $\lambda \in \mathbb{C}$ ein Eigenwert der Form $\mathcal{E}$ ist, falls ein $f \in \mathbb{R}^n \setminus \{0\}$ existiert, sodass

$$\mathcal{E}(f,f) = \lambda \|f\|^2$$

erfüllt ist. Die Eigenwerte der Form entsprechen also denen des darstellenden Operators $I - D^{1/2}KD^{-1/2}$. Diese sind jedoch gerade die Werte $\{1-\lambda_i\}_{i=0}^{n-1}$. Das Variationsprinzip von Rayleigh-Ritz besagt für unser Problem nun, dass sich der zweitkleinste Eigenwert $1-\lambda_1$ bestimmen lässt durch

$$1-\lambda_1 = \inf\left\{\frac{\mathcal{E}(f,f)}{\|f\|^2} : f \in \operatorname{span}\{\pi\}^\perp,\ f \neq 0\right\}.$$

Beachte, dass π der Eigenvektor zum Eigenwert 0, also dem kleinsten Eigenwert von $I - D^{1/2}KD^{-1/2}$ ist.