Endliche Markov-Ketten - eine Übersicht

Diese Übersicht über endliche Markov-Ketten basiert auf dem Buch „Monte Carlo-Algorithmen“ von Müller-Gronbach et. al. und dient als Sammlung von Definitionen und Sätzen, mit welchen man auch im Blockseminar zur Wahrscheinlichkeitstheorie konfrontiert wird bzw. werden könnte. Wer eine ausführlichere Darstellung bzw. Einführung wünscht, die sich dennoch schnell erarbeiten lässt, dem sei die oben genannte Quelle nahegelegt.

Im Folgenden sei Z eine beliebige endliche Menge, stillschweigend ausgestattet mit ihrer Potenzmenge als σ-Algebra. Die Elemente aus $\mathbb{R}^Z$ bzw. $\mathbb{R}^{Z\times Z}$ verstehen wir als Zeilenvektoren bzw. als quadratische Matrizen, wobei eine beliebige Anordnung der Indexmenge Z zugrunde gelegt sei.

Definition 1. Wir definieren:

1) Ein Vektor $\mu = (\mu_z)_{z\in Z} \in \mathbb{R}^Z$ heißt *Wahrscheinlichkeitsvektor*, falls gilt:
   i) $\forall z \in Z : \mu_z \geq 0$
   ii) $\sum_{z\in Z} \mu_z = 1$
2) Eine Matrix $Q = (q_{z,z'})_{z,z'\in Z} \in \mathbb{R}^{Z\times Z}$ heißt *stochastische Matrix*, falls ihre Zeilen Wahrscheinlichkeitsvektoren sind, d.h. falls gilt:
   i) $\forall z, z' \in Z : q_{z,z'} \geq 0$
   ii) $\forall z \in Z : \sum_{z'\in Z} q_{z,z'} = 1$

Definition 2. Ein stochastischer Prozess $(X_n)_{n\in\mathbb{N}_0}$ mit Zustandsraum Z heißt *endliche Markov-Kette*, falls er die *Markov-Eigenschaft* erfüllt, dh. falls gilt:

$$\mathbb{P}(X_{n+1} = z_{n+1}|(X_0, \dots, X_n) = (z_0, \dots, z_n)) = \mathbb{P}(X_{n+1} = z_{n+1}|X_n = z_n)$$

für alle $n \in \mathbb{N}$ und alle $z_0, \dots, z_{n+1}$ mit $\mathbb{P}((X_0, \dots, X_n) = (z_0, \dots, z_n)) > 0$. Die endliche Markov-Kette heißt *homogen*, falls für alle $z, z' \in Z$ ein $q_{z,z'} \in \mathbb{R}$ existiert, sodass für alle $n \in \mathbb{N}$ mit $\mathbb{P}(X_n = z) > 0$ gilt:

$$\mathbb{P}(X_{n+1} = z'|X_n = z) = q_{z,z'}$$

Anmerkung. Im Folgenden betrachten wir nur noch homogene endliche Markov-Ketten mit Zustandsraum Z und nennen diese schlichtweg „Markov-Kette“.

Definition 3. Sei $(X_n)_{n\in\mathbb{N}_0}$ eine Markov-Kette, so heißt eine stochastische Matrix $Q \in \mathbb{R}^{Z\times Z}$ *Übergangsmatrix* dieser Kette, falls für alle $z, z' \in Z$ gilt: Ist $n \in \mathbb{N}_0$ beliebig mit $\mathbb{P}(X_n = z) > 0$, so folgt $\mathbb{P}(X_{n+1} = z'|X_n = z) = q_{z,z'}$.

Anmerkung. Im Kontext einer Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$ bezeichnen wir mit Q stets ihre Übergangsmatrix. Ferner setzen wir $\forall n \in \mathbb{N}_0 : \mu^{(n)} := \mathbb{P}^{X_n}$, also die Verteilung der Kette zum Zeitpunkt n. Die Verteilung $\mu^{(0)}$ bezeichnet man auch als Startverteilung der Markov-Kette. Die so definierten Verteilungen $(\mu^{(n)})_{n\in\mathbb{N}_0}$ identifizieren wir auch stets mit entsprechenden Wahrscheinlichkeitsvektoren aus $\mathbb{R}^Z$.

Lemma 4. *Eine Markov-Kette $(X_n)_{n\in\mathbb{N}_0}$ besitzt die folgenden Eigenschaften:*

i) Für alle $n, l \in \mathbb{N}_0$ gilt: $\mu^{(n+l)} = \mu^{(n)} \cdot Q^l$

ii) Für alle $n \in N$ und $z_0, \dots, z_n \in Z$ gilt:
$P((X_0, \dots, X_n) = (z_0, \dots, z_n)) = \mu^{(0)}_{z_0} \cdot q_{z_0,z_1} \cdots q_{z_{n-1},z_n}.$

Anmerkung. Eine Markov-Kette (bzw. die Verteilung einer Markov-Kette) ist also eindeutig bestimmt durch ihre Anfangsverteilung $\mu^{(0)}$ sowie ihre Übergangsmatrix Q. Ferner nennt man für alle $l \in N$ die Matrix Q^l auch l-Schritt-Übergangsmatrix. Wir schreiben dann $Q^l = (q^{(l)}_{z,z'})_{z,z'\in Z}$.

Definition 5. Eine stochastische Matrix $Q \in \mathbb{R}^{Z\times Z}$ heißt irreduzibel, falls

$$\forall z, z' \in Z : \ \exists n > 0 : (q^{(n)}_{z,z'}) > 0$$

und aperiodisch, falls

$$\forall z \in Z : \mathrm{ggT}\{n > 0 | q^{(n)}_{z,z} > 0\} = 1.$$

Definition 6. Eine Markov-Kette heißt irreduzibel bzw. aperiodisch, falls ihre Übergangsmatrix diese Eigenschaft erfüllt.

Anmerkung. Markov-Ketten, welche sowohl irreduzibel, als auch aperiodisch sind, sind in der Theorie und in den Anwendungen extrem wichtig. Irreduzibilität bedeutet, dass die Kette von jedem Zustand $z \in Z$ ausgehend jeden Zustand $z' \in Z$ mit positiver Wahrscheinlichkeit nach einer positiven Anzahl von Schritten erreicht. Aperiodizität bedeutet zum einen, dass die Markov-Kette ausgehend von einem Zustand $z \in Z$ nach einer positiven Anzahl von Schritten mit positiver Wahrscheinlichkeit in diesen Zustand zurückkehren kann, und dass jene möglichen Rückkehrzeitpunkte nicht periodisch sind, das heißt nicht in einer Menge der Form $\{d, 2d, \dots\}$ für ein $d \geq 2$ liegen.

Bemerkung 7. Viele der folgenden Aussagen werden wir für stochastische Matrizen Q formulieren. Die Aussagen gelten dann also für alle Markov-Ketten mit Übergangsmatrix Q und beliebiger Anfangsverteilung $\mu^{(0)}$.

Lemma 8. *Sei Q eine irreduzible stochastische Matrix mit einem positiven Diagonalelement. Dann ist Q aperiodisch.*

Anmerkung. Die Diagonalelemente einer Übergangsmatrix einer Markov-Kette nennt man auch Verweilwahrscheinlichkeiten, da diese gerade die Wahrscheinlichkeit angeben, dass die Kette im nächsten Schritt im aktuellen Zustand stehen bleibt. Hat man also eine irreduzible Markov-Kette mit mindestens einer positiven Verweilwahrscheinlichkeit, so ist sie auch aperiodisch. Im Folgenden nennen wir einen Vektor oder eine Matrix positiv, falls alle Einträge positiv sind.

Lemma 9. *Für jede stochastische Matrix $Q \in \mathbb{R}^{Z \times Z}$ sind folgende Aussagen äquivalent:*

i) *Q ist irreduzibel und aperiodisch*

ii) $\exists\, n > 0 : Q^n$ *ist positiv*

iii) $\exists\, N > 0 : \ \forall n \geq N : Q^n$ *ist positiv*

Definition 10. Sei $Q \in \mathbb{R}^{Z \times Z}$ eine stochastische Matrix. Ein Wahrscheinlichkeitsvektor $\mu \in \mathbb{R}^Z$ mit $\mu \cdot Q = \mu$ heißt *stationäre Verteilung* oder *invariantes Maß* der Matrix Q (bzw. einer Markov-Kette mit Übergangsmatrix Q).

Anmerkung. Die Verteilungen $\mathbb{P}^{X_n}$ einer Markov-Kette zu den Zeiten $n \in \mathbb{N}_0$ stimmen genau dann überein, wenn die Anfangsverteilung der Kette eine stationäre Verteilung der Übergangsmatrix der Kette ist. Die Gleichverteilung auf Z ist offenbar genau dann eine stationäre Verteilung von Q, wenn auch die Spaltensummen von Q gleich eins sind. Solche Übergangsmatrizen werden auch als doppelt-stochastisch bezeichnet. Symmetrische stochastische Matrizen sind z.B. stets doppelt-stochastisch.

Definition 11. Eine stochastische Matrix $Q = (q_{z,z'}) \in \mathbb{R}^{Z \times Z}$ (bzw. eine Markov-Kette mit jener Übergangsmatrix) heißt *reversibel*, falls ein Wahrscheinlichkeitsvektor $\mu \in \mathbb{R}^Z$ existiert, der die *detailed-balance-Gleichungen* erfüllt, das heißt

$$\forall\, z, z' \in Z : \mu_z \cdot q_{z,z'} = \mu_{z'} \cdot q_{z',z}.$$

In diesem Fall sagt man auch, Q (bzw. eine Markov-Kette mit Übergangsmatrix Q) sei reversibel bzgl. μ.

Lemma 12. *Sei Q reversibel bezüglich μ. Dann ist μ eine invariante Verteilung von Q.*

Anmerkung. Ist Q symmetrisch, dann auch reversibel bzgl. der Gleichverteilung auf Z.

Satz 13. *Zu jeder irreduziblen und aperiodischen Matrix Q existiert eine positive stationäre Verteilung.*

Satz 14. *Sei $Q \in \mathbb{R}^{Z \times Z}$ eine irreduzible und aperiodische stochastische Matrix und sei $\mu \in \mathbb{R}^Z$ eine stationäre Verteilung von Q. Dann existieren Konstanten $c > 0$ und $\alpha \in (0,1)$, sodass*

$$\sup_{z \in Z} |(\mu^{(0)} \cdot Q^n)_z - \mu_z| \leq c \cdot \alpha^n$$

für jeden Wahrscheinlichkeitsvektor $\mu^{(0)} \in \mathbb{R}^Z$ und alle $n \in \mathbb{N}$ gilt.

Definition 15. Seien μ und ν beliebige Wahrscheinlichkeitsverteilungen auf Z, dann definiert man den Totalvariationsabstand zwischen μ und ν durch

$$\|\mu - \nu\|_{TV} := \sup_{A \subseteq Z} |\mu(A) - \nu(A)|$$

Anmerkung. Der Totalvariationabstand bildet eine Metrik auf der Menge aller Wahrscheinlichkeitsverteilungen auf Z, d.h. man kann von *Konvergenz in Totalvariation* sprechen.

Corollar 16. 1. Ergodensatz: *Jede irreduzible und aperiodische Matrix Q besitzt eine eindeutig bestimmte stationäre Verteilung μ. Diese ist positiv, und die Folge der Verteilungen $\mu^{(n)}$ jeder Markov-Kette mit Übergangsmatrix Q konvergiert für jede Anfangsverteilung $\mu^{(0)}$ exponentiell schnell gegen μ in Totalvariation. Genauer existieren Konstanten $c > 0$ sowie $\alpha \in (0,1)$, sodass für alle $n \in \mathbb{N}_0$ gilt:*

$$\|\mu^{(n)} - \mu\|_{TV} = \sup_{A \subseteq Z} |\mu^{(n)}(A) - \mu(A)| \leq c \cdot \alpha^n$$

Definition 17. Das Gleichgewichtsmaß $\tilde{\mu}$ einer Markov-Kette $(X_n)_{n \in \mathbb{N}_0}$ gibt die Langzeitverteilung dieser an und ist definiert durch

$$\tilde{\mu}(z) = \lim_{n \to \infty} \mathbb{P}(X_n = z | X_0 = z') \ \forall z, z' \in Z,$$

falls dies wohldefiniert ist, d.h. der Grenzwert auf der rechten Seite nicht von z' abhängt. Falls alle Voraussetzungen für den 1. Ergodensatz erfüllt sind, so ist das Gleichgewichtsmaß genau die stationäre Verteilung der Markov-Kette, es gilt also $\tilde{\mu} = \mu$.

Satz 18. 2. Ergodensatz: *Sei $(X_n)_{n \in \mathbb{N}_0}$ eine irreduzible und aperiodische Markov-Kette mit stationärer Verteilung $\mu \in \mathbb{R}^Z$. Für jede Abbildung $f : Z \longrightarrow \mathbb{R}$ gilt fast sicher*

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} f(X_k) = \sum_{z \in Z} f(z) \mu_z$$

Anmerkung. Man beachte in der Aussage des 2. Ergodensatzes, dass rechts der μ-Erwartungswert von f steht (also $\mathbb{E}_\mu f$) und denke an das starke Gesetz der großen Zahlen.

Satz 19. *Sei die stochastische Matrix $Q \in \mathbb{R}^{Z\times Z}$ irreduzibel, aperiodisch und reversibel bzgl. ihrer stationären Verteilung μ. Dann gilt:*

i) Q besitzt nur reelle Eigenwerte.

ii) Q besitzt den Eigenwert 1. Jener ist einfach und dominiert betraglich alle anderen Eigenwerte strikt.

Definition 20. Sei $X = (X_n)_{n\in\mathbb{N}_0}$ eine irreduzible, aperiodische und bzgl. μ reversible Markov-Kette mit Übergangsmatrix $Q \in \mathbb{R}^{Z\times Z}$. Dann gilt für die (nicht notwendigerweise verschiedenen) Eigenwerte $(\lambda_z)_{z\in Z}$ nach geeineter Nummerierung:

$$1 = \lambda_1 > \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_{|Z|} > -1$$

Wir setzen nun $\Delta(X) := 1 - \lambda_2$ und nennen diesen Wert Spektrallücke von X. Ferner setzen wir $\Delta^*(X) := 1 - max(\lambda_2, |\lambda_{|Z|}|)$ und nennen diesen Wert absolute Spektrallücke von X.

Satz 21. *Sei $X = (X_n)_{n\in\mathbb{N}_0}$ eine irreduzible, aperiodische und bzgl. μ reversible Markov-Kette. Dann gilt $\Delta^*(X) \in (0,1)$ und es existiert ein $c > 0$ sodass für alle $n \in \mathbb{N}_0$ gilt :*

$$\|\mathbb{P}^{X_n} - \mu\|_{TV} \leq c \cdot (1 - \Delta^*(X))^n$$

Satz 22. *Sei $X = (X_n)_{n\in\mathbb{N}_0}$ eine irreduzible, aperiodische und bzgl. μ reversible Markov-Kette sowie $f : Z \longrightarrow \mathbb{R}$ eine Funktion. Dann gilt ein zentraler Grenzwertsatz im folgenden Sinne:*

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}_\mu f\right) \xrightarrow{d} N(0, \mathbb{V}_\mu^a f) \quad \textit{für } n \to \infty$$

wobei für die asymptotische Varianz $\mathbb{V}_\mu^a f \in (0,\infty)$ gilt:

$$\mathbb{V}_\mu^a f \leq \left(\frac{2}{\Delta(X)}\right) \mathbb{V}_\mu f$$

Hierbei ist $\Delta(X)$ die Spektrallücke der Kette X.