Seminar zur Numerik WiSe 19/20 **Ruhr-Universität Bochum, 16.01.2020**

11. Konvergenzsätze für Markov-Ketten; MCMC-Methoden

Handout *von Timo Schorlepp*

Erinnerung an wichtige Definitionen und Sätze:
$(\Omega, \mathcal{F}, P)$ sei W-Raum, Z bezeichne eine endliche Menge mit der Potenzmenge als σ-Algebra, $(X_t)_{t\in\mathbb{N}_0}$ sei zeitdiskreter stochastischer Prozess mit $X_t : \Omega \to Z$ messbar $\forall\, t$, und $(\mathcal{F}_t)_{t\in\mathbb{N}_0}$ sei die zugehörige Filtration auf Ω, d.h. $\mathcal{F}_t = \sigma(X_0, \dots, X_t) \subset \mathcal{F}$. Eine Abbildung $T : \Omega \to \mathbb{N}_0 \cup \{+\infty\} =: \mathbb{N}_\infty$ heißt Stoppzeit bzgl. (X_t), falls $\{T = t\} \in \mathcal{F}_t \forall\, t \in \mathbb{N}_\infty$ (mit $\mathcal{F}_\infty = \sigma(X_0, X_1, \dots)$).

(X_t) heißt homogene Markov-Kette, falls $P(X_{n+1} = z_{n+1}|(X_0, \dots, X_n) = (z_0, \dots, z_n)) = P(X_{n+1} = z_{n+1}|X_n = z_n)$ für alle $n \in \mathbb{N}$ und $z_0, \dots, z_{n+1} \in Z$ mit $P((X_0, \dots, X_n) = (z_0, \dots, z_n)) > 0$, sowie $Q_{z,z'} := P(X_{n+1} = z'|X_n = z) \neq Q_{z,z'}(n)$ für $P(X_n = z) > 0$. Existiert $z_0 \in Z$ mit $P(X_n = z_0) = 0$ für alle n, so setze $Q_{z_0,z'} = \delta_{z_0,z'}$. Dann gilt für die so konstruierte Matrix $Q \in \mathbb{R}^{Z\times Z}$: $Q_{z,z'} \geq 0$ für alle $z, z' \in Z$, und für die Zeilensummen gilt $\sum_{z'\in Z} Q_{z,z'} = 1$, womit Q per Definition eine stochastische Matrix ist. Ist T fast sicher endlich, so ist (X_{T+n}) ebenfalls eine Markov-Kette mit Übergangsmatrix Q. Für die Verteilungen $\mu^{(n)}$ von X_n gilt $\mu^{(n+l)} = \mu^{(n)} \cdot Q^l$. Q heißt irreduzibel, falls $\forall z, z' \in Z : \exists n > 0 : (Q^n)_{z,z'} > 0$. Q heißt aperiodisch, falls $\forall z \in Z : \mathrm{ggT}(\{n > 0|(Q^n)_{z,z} > 0\}) = 1$. Ist Q irreduzibel und $Q_{z,z} > 0$ für ein $z \in Z$, so ist Q aperiodisch. Ferner ist Q genau dann irreduzibel und aperiodisch, wenn Q^n positiv ist für ein $n > 0$, d.h. $(Q^n)_{z,z'} > 0 \forall z, z' \in Z$. Für die Stoppzeit $T_z = \inf\{n > 0|X_n = z\}$ für $z \in Z$ (*Eintrittszeit*) gilt, falls Q irreduzibel und aperiodisch ist, $ET_z < \infty$ und insbesondere ist T_z fast sicher endlich.

Ein Wahrscheinlichkeitsvektor $\mu \in \mathbb{R}^Z$ (d.h. $\mu_z \geq 0 \forall z \in Z$ und $\sum_{z\in Z} \mu_z = 1$) heißt stationäre Verteilung von Q, falls $\mu = \mu \cdot Q$. Ist Q irreduzibel mit einer stationären Verteilung μ, so ist $\tilde{Q} := (I + Q)/2$ (*lazy version* von Q) irreduzibel und aperiodisch mit stationärer Verteilung μ. Da μ stationäre Verteilung von Q ist genau dann wenn $\mu_z = \sum_{z'\in Z} \mu_{z'} Q_{z',z}$, folgt sofort, dass, falls μ die *detailed-balance*-Gleichung $\mu_z Q_{z,z'} = \mu_{z'} Q_{z',z}\ \forall z, z' \in Z$ erfüllt, μ auch stationäre Verteilung von Q ist. Abschließend wurde gezeigt, dass für irreduzible und aperiodische Q immer eine positive stationäre Verteilung existiert.

Das hard core model:
Wir betrachten einen ungerichteten Graphen (E, K) mit $|E| < \infty$, $K \neq \emptyset$ und $Z \subset \tilde{Z} = \{0, 1\}^E$, sodass für $z \in Z$ gilt, dass $z(e) = z(e') = 1 \Rightarrow \{e, e'\} \notin K$. Die gesuchte stationäre Verteilung sei die Gleichverteilung auf Z, und eine natürliche Frage ist zum Beispiel: Wie groß ist die mittlere Besetzungszahl, d.h. was ist $E_\mu(f)$ für $f : Z \to \mathbb{R}, z \mapsto 1/|E| \sum_{e\in E} z(e)$? Das Problem mit einer direkten Simulation basierend auf der Verwerfungsmethode ist, dass die Akzeptanzwahrscheinlichkeit im Allgemeinen exponentiell klein in $|E|$ ist.

Das Ising-Modell:
Es sei wieder (E, K) ein ungerichteter Graph mit $|E| < \infty$, $K \neq \emptyset$ und $Z = \{-1, 1\}^E$. Das Ziel ist es, Erwartungswerte bezüglich einer Boltzmann-Verteilung μ^β auf Z mit $\beta > 0$ zu berechnen, d.h. $\mu^\beta_z = 1/C_\beta \exp(-\beta H(z))$ für $z \in Z$ mit der Zustandssumme C_β als Normierungsfaktor und der Hamilton-Funktion $H : Z \to \mathbb{R}$ von der allgemeinen Form $H(z) = \sum_{e\in E} h_e(z(e)) + \sum_{\{e,e'\}\in K} h_{\{e,e'\}}(z(e), z(e'))$, wobei die Funktionen $h_e : \{-1, 1\} \to \mathbb{R}$ bspw. ein äußeres, ortsabhängiges Magnetfeld modellieren, und die Funktionen $h_{\{e,e'\}} : \{-1, 1\}^2 \to \mathbb{R}$ Wechselwirkungen benachbarter Spins modellieren. Ein interessante Frage ist beispielsweise, wie der Erwartungswert der absoluten Magnetisierung $f : Z \to \mathbb{R}, z \mapsto 1/|E| \left|\sum_{e\in E} z(e)\right|$ im Ising-Modell ohne äußeres Magnetfeld von der (inversen) Temperatur β abhängt. Das Problem mit naiven Monte-Carlo-Simulationen wie

$$\frac{\sum_{i=1}^n f(X_i) e^{-\beta H(X_i)}}{\sum_{i=1}^n e^{-\beta H(X_i)}}$$

für gleichverteilte Z-wertige uiv. ZV X_i ist die große Varianz für große β, wenn die Boltzmann-Verteilung von wenigen Zuständen nahe an den Minima von H dominiert ist.

In diesem Vortrag: Theoretische Grundlagen für die Markov-Chain-Monte-Carlo-Methode (MCMC) + Anwendung auf obige Beispiele.

Satz 11.1. *(Konvergenz gegen stationäre Verteilungen; Satz 6.28 in [1])*
Sei $Q \in \mathbb{R}^{Z\times Z}$ eine irreduzible und aperiodische stochastische Matrix und $\mu \in \mathbb{R}^Z$ eine stationäre Verteilung von Q. Dann existieren (feste) Konstanten $c > 0$ und $\alpha \in (0,1)$ mit

$$\max_{z\in Z} \left| (\mu^{(0)} \cdot Q^n)_z - \mu_z \right| \leq c \cdot \alpha^n \tag{11.1}$$

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

Satz 11.2. *(Ergodensatz für homogene Markov-Ketten; Satz 6.30 in [1])*
Sei $(X_n)_{n\in\mathbb{N}_0}$ eine homogene Markov-Kette mit endlichem Zustandsraum Z, irreduzibler und aperiodischer Übergangsmatrix Q und stationärer Verteilung $\mu \in \mathbb{R}^Z$. Dann gilt für jede Abbildung $f : Z \to \mathbb{R}$

$$\frac{1}{n}\sum_{i=1}^{n} f(X_i) \xrightarrow[n\to\infty]{f.s.} E_\mu(f). \tag{11.2}$$

Verbleibende Aufgabe: Wie konstruiert man nun im Allgemeinen eine irreduzible und aperiodische stochastische Matrix zu einer gegebenen Verteilung $\mu \in \mathbb{R}^Z$, sodass die zugehörige Markov-Kette mit möglichst wenig Aufwand zu simulieren ist? Starte dazu mit einer *beliebigen* irreduziblen und symmetrischen stochastischen Matrix $\tilde{Q} \in \mathbb{R}^{Z\times Z}$. Wähle dann für $z \neq z'$ sogenannte Akzeptanzwahrscheinlichkeiten $\alpha_{z,z'} \in (0,1]$, die bezüglich der stationären Verteilung μ die detailed-balance-artigen Gleichungen

$$\mu_z \cdot \alpha_{z,z'} = \mu_{z'} \cdot \alpha_{z',z} \tag{11.3}$$

erfüllen. Dann ist die Matrix Q, definiert durch $Q_{z,z'} = \tilde{Q}_{z,z'} \cdot \alpha_{z,z'}$ für $z \neq z'$ und $Q_{z,z} = 1 - \sum_{z'\neq z} Q_{z,z'}$, ebenfalls eine stochastische irreduzible Matrix mit stationärer Verteilung μ. Ist $\tilde{Q}$ aperiodisch, so auch Q.

Zwei spezielle Wahlen von Akzeptanzwahrscheinlichkeiten, die für beliebige μ die detailed-balance-Gleichung erfüllen, sind der Metropolis-Algorithmus [5] mit

$$\alpha_{z,z'} = \min\left(1, \frac{\mu_{z'}}{\mu_z}\right) \tag{11.4}$$

und der Gibbs-Sampler

$$\alpha_{z,z'} = \frac{\mu_{z'}}{\mu_z + \mu_{z'}}. \tag{11.5}$$

Literatur

[1] T. Müller-Gronbach, E. Novak, K. Ritter, “Monte Carlo-Algorithmen” *Springer-Verlag*, 2012.

[2] F. Schwabl, “Statistische Mechanik” *Springer-Verlag*, 2006.

[3] M. E. J. Newman, G. T. Barkema, “Monte Carlo Methods in Statistical Physics” *Oxford University Press*, 2002.

[4] D. A. Levin, Y. Peres, “Markov Chains and Mixing Times” *American Mathematical Soc.*, 2017.

[5] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, “Equation of State Calculations by Fast Computing Machines” *J. Chem. Phys. 21, 1087*, 1953.

[6] L. Onsager, “Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition” *Phys. Rev. 65, 117*, 1944.

[7] Ausführlichere Notizen und der im Vortrag gezeigte Code für das hard core model und das Ising-Modell finden sich unter https://github.com/TimoSchorlepp/MiscCoursework/tree/master/MCSeminar