9 Metropolis-Hastings-Algorithmus

9.1 Einleitung

in Arbeit: Die Einleitung fehlt noch.

9.2 Der Metropolis-Hastings-Algorithmus

9.1 Definition Eine ergodische Markov-Kette heißt *(zeit-)reversibel* (im Englischen mitunter auch *in detailed balance*), wenn für die stationäre Verteilung $\mathbf{w}$ und alle Zustände i und j gilt:

$$\mathbf{w}_i P_{ij} = \mathbf{w}_j P_{ji} . \qquad \diamond$$

Zur Begründung dieses Namens mache man sich klar, dass $\mathbf{w}_i P_{ij}$ die Wahrscheinlichkeit ist, dass man bei einer Markov-Kette im stationären Zustand bei einem Schritt gerade einen Übergang von Zustand i nach Zustand j beobachtet, und $\mathbf{w}_j P_{ji}$ die Wahrscheinlichkeit, dass man bei einem Schritt umgekehrt einen Übergang von Zustand j nach Zustand i.

Als Beispiele können die folgenden Ketten dienen:

9.2 Definition Es sei $G = (V, E)$ ein zusammenhängender ungerichteter Graph ohne Schlingen und $0 < \beta \leqslant 1$ eine reelle Zahl. Mit $d(i)$ bezeichnen wir den Grad von Knoten i und es sei $d = \max_{i \in V} d(i)$. Die Übergangswahrscheinlichkeiten P_{ij} der Markov-Kette $M_{G,\beta}$ sind dann wie folgt definiert:

$$P_{ij} = \begin{cases} \beta/d & \text{falls } i \neq j \text{ und } (i,j) \in E \\ 0 & \text{falls } i \neq j \text{ und } (i,j) \notin E \\ 1 - d(i)\beta/d & \text{falls } i = j \end{cases} \qquad \diamond$$

9.3 Man kann sich überlegen, dass für die so definierten Markov-Ketten die Gleichverteilung die stationäre Verteilung ist. Da der Graph zusammenhängend ist, ist die Kette irreduzibel. Für $\beta < 1$ ist sie außerdem aperiodisch und reversibel.

9.4 Eine kleine Verallgemeinerung liefert schon eine ganz einfache Variante des sogenannten *Metropolis-Hastings-Algorithmus*. Sei dazu $\mathbf{p}$ eine Wahrscheinlichkeitsverteilung auf V, die nirgends 0 ist. Dann betrachte man folgende Festlegungen:

$$P_{ij} = \begin{cases} \min(1, \frac{\mathbf{p}_j}{\mathbf{p}_i}) \cdot \beta/d & \text{falls } i \neq j \text{ und } (i,j) \in E \\ 0 & \text{falls } i \neq j \text{ und } (i,j) \notin E \\ 1 - \sum_{i \neq k} P_{ik} & \text{falls } i = j \end{cases}$$

Auch diese Markov-Kette ist reversibel, was man zum Beispiel so sieht. Sei $i \neq j$ und $(i,j) \in E$ (die anderen Fälle sind trivial). Dann gilt:

$$\frac{\mathbf{p}_i P_{ij}}{\mathbf{p}_j P_{ji}} = \frac{\mathbf{p}_i \min(1, \frac{\mathbf{p}_j}{\mathbf{p}_i}) \cdot \beta/d}{\mathbf{p}_j \min(1, \frac{\mathbf{p}_i}{\mathbf{p}_j}) \cdot \beta/d} = \begin{cases} \frac{\mathbf{p}_i}{\mathbf{p}_j} / \frac{\mathbf{p}_i}{\mathbf{p}_j} & \text{falls } \mathbf{p}_i \leqslant \mathbf{p}_j \\ \frac{\mathbf{p}_i}{\mathbf{p}_j} \cdot \frac{\mathbf{p}_j}{\mathbf{p}_i} & \text{falls } \mathbf{p}_i > \mathbf{p}_j \end{cases} = 1$$

Wegen des folgenden Lemmas kennt man auch sofort die stationäre Verteilung dieser Markov-Kette: Es ist $\mathbf{p}$.

9.5 Lemma. Ist M eine ergodische Markov-Kette und $\mathbf{q}$ eine Verteilung mit der Eigenschaft, dass für alle Zustände i und j gilt, dass $\mathbf{q}_i P_{ij} = \mathbf{q}_j P_{ji}$ ist, dann ist $\mathbf{q}$ die (eindeutig bestimmte) stationäre Verteilung von M.

9.6 Beweis. Für alle i ist

$$(\mathbf{q}P)_i = \sum_j \mathbf{q}_j P_{ji} = \sum_j \mathbf{q}_i P_{ij} = \mathbf{q}_i \sum_j P_{ij} = \mathbf{q}_i \ .$$

■

9.7 Für den klassischen Algorithmus von Metropolis u. a. (1953) ist eine zeilenstochastische Matrix Q gegeben, die symmetrisch ist (diese Einschränkung wurde von Hastings (1970) beseitigt). Ziel ist aber eine reversible Markov-Kette mit Übergangsmatrix P, deren stationäre Verteilung $\mathbf{w}$ nur indirekt gegeben ist. Festgelegt ist nämlich nur eine Funktion $H\colon S \to \mathbb{R}$ mit der Eigenschaft, dass $e^{-H(i)}$ proportional zur Wahrscheinlichkeit von Zustand i in $\mathbf{w}$ ist. Der Proportionalitätsfaktor ist natürlich $Z = \sum_{i \in S} e^{-H(i)}$.

Q heißt üblicherweise *proposal matrix*, H *energy function* und Z *partition function*.

Das Schöne am Metropolis-Algorithmus ist, dass man Z gar nicht kennen muss. (In manchen Anwendungen in der Physik ist S *sehr* groß.)

in Arbeit: Der Rest fehlt noch.

9.3 Simulated Annealing

Dieser Abschnitt fehlt noch.

Literatur

Hastings (1970). „Monte Carlo Sampling Methods Using Markov Chains and Their Applications". In: *Biometrika* 57.1, S. 97–109. doi: 10.1093/biomet/57.1.97 (siehe S. 80).

Metropolis, Nicholas u. a. (1953). „Equations of State Calculations by Fast Computing Machines". In: *Journal of Chemical Physics* 21.6, S. 1087–1092 (siehe S. 80).

