Randomisierte Algorithmen

9. Metropolis-Hastings-Algorithmus

Thomas Worsch

Fakultät für Informatik
Karlsruher Institut für Technologie

Wintersemester 2019/2020

Überblick

Überblick

Definition Zeitreversibilität

ergodische Markov-Kette ist *(zeit-)reversibel* (*in detailed balance*), wenn es Wahrscheinlichkeitsverteilung $\mathbf{w}$ auf der Zustandsmenge S gibt so, dass für alle $i, j \in S$ gilt:

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

Lemma

- wenn
  - M Markov-Kette und
  - $\mathbf{q}$ Verteilung mit
    $$\mathbf{q}_i P_{ij} = \mathbf{q}_j P_{ji}$$
    für alle Zustände i und j,
- dann ist $\mathbf{q}$ eine stationäre Verteilung von M.

Beweis

Für alle i ist

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

Mitteilung: Kriterium von Kolmogorov

Satz

Eine ergodische Markov-Kette ist genau dann(zeit-)reversibel, wenn für alle Folgen $(i_0, i_1, \ldots, i_n, i_0)$ von Zuständen gilt:

$$\begin{aligned} & P_{i_0 i_1} \cdot P_{i_1 i_2} \cdots P_{i_{n-1} i_n} \cdot P_{i_n i_0} \\ = & P_{i_0 i_n} \cdot P_{i_n i_{n-1}} \cdots P_{i_2 i_1} \cdot P_{i_1 i_0} \end{aligned}$$

ohne Beweis

Von ungerichteten Graphen zu reversible Markovketten

- $G = (V, E)$ ein zusammenhängender ungerichteter Graph ohne Schlingen
- $0 < \beta \leq 1$ eine reelle Zahl
- $d(i)$ Grad von Knoten i und $d = \max_{i \in V} d(i)$
- definiere Markov-Kette $M_{G,\beta}$:

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

Von ungerichteten Graphen zu reversible Markovketten

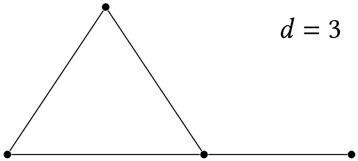
- $G = (V, E)$ ein zusammenhängender ungerichteter Graph ohne Schlingen
- $0 < \beta \leq 1$ eine reelle Zahl
- $d(i)$ Grad von Knoten i und $d = \max_{i \in V} d(i)$
- definiere Markov-Kette $M_{G,\beta}$:

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

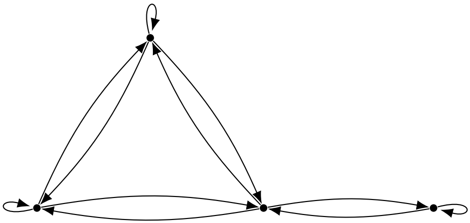
$$P_{ij} = \begin{cases} \beta/d & \text{falls } i \neq j \text{ und } (i,j) \in E \\ 0 & \text{falls } i \neq j \text{ und } (i,j) \notin E \\ 1 - \sum_{i \neq k} P_{ik} & \text{falls } i = j \end{cases}$$

Von ungerichteten Graphen zu reversible Markovketten

Beispiel



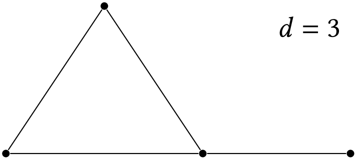
$d = 3$



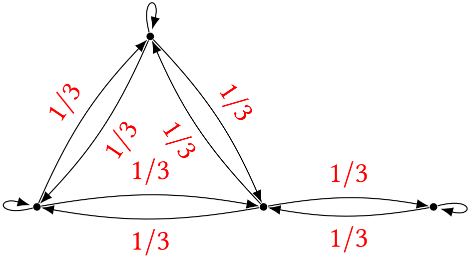
$\beta = 1$

Von ungerichteten Graphen zu reversible Markovketten

Beispiel



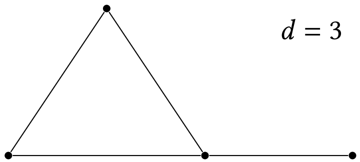
$d = 3$



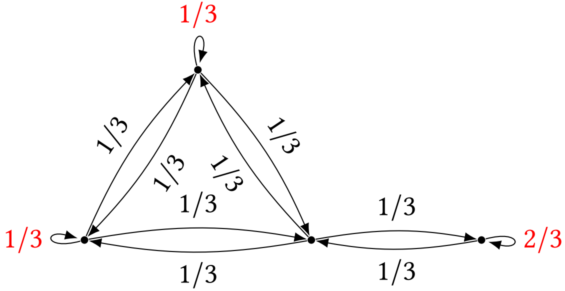
$\beta = 1$

Von ungerichteten Graphen zu reversible Markovketten

Beispiel

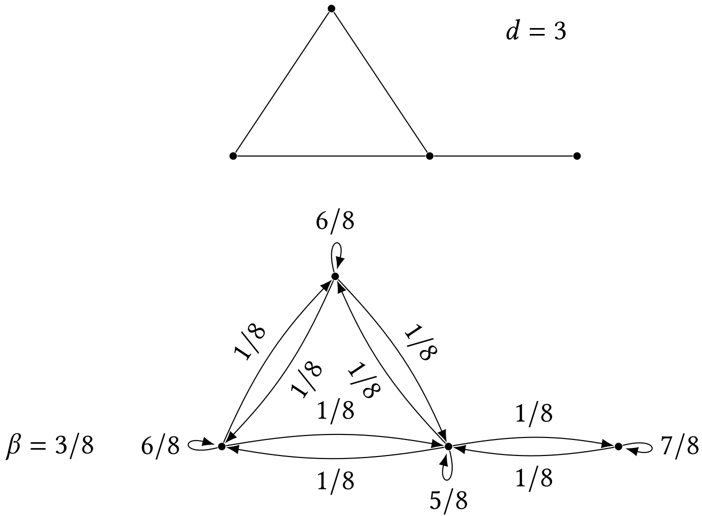


$\beta = 1$



Von ungerichteten Graphen zu reversible Markovketten

Beispiel



Beobachtung/Übung

- Kette irreduzibel, da Graph zusammenhängend
- für $\beta < 1$ Kette aperiodisch
- stationäre Verteilung $M_{G,\beta}$: Gleichverteilung
- für $\beta < 1$ Kette reversibel

Verallgemeinerung: reversible Markovketten

mit frei wählbarer Wahrscheinlichkeitsverteilung

- $G = (V, E)$ ein zusammenhängender ungerichteter Graph
  - $0 < \beta < 1$ eine reelle Zahl
  - $d(i)$ Grad von Knoten i und $d = \max_{i \in V} d(i)$
- $\mathbf{p}$ eine W.verteilung auf V, nirgends 0, sonst beliebig

Verallgemeinerung: reversible Markovketten

mit frei wählbarer Wahrscheinlichkeitsverteilung

- $G = (V, E)$ ein zusammenhängender ungerichteter Graph
  - $0 < \beta < 1$ eine reelle Zahl
  - $d(i)$ Grad von Knoten i und $d = \max_{i \in V} d(i)$
- $\mathbf{p}$ eine W.verteilung auf V, nirgends 0, sonst beliebig
- definiere Markov-Kette $M_{G,\beta,\mathbf{p}}$:

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

Verallgemeinerung: reversible Markovketten

mit frei wählbarer Wahrscheinlichkeitsverteilung

- $G = (V, E)$ ein zusammenhängender ungerichteter Graph
  - $0 < \beta < 1$ eine reelle Zahl
  - $d(i)$ Grad von Knoten i und $d = \max_{i \in V} d(i)$
- $\mathbf{p}$ eine W.verteilung auf V, nirgends 0, sonst beliebig
- definiere Markov-Kette $M_{G,\beta,\mathbf{p}}$:

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

Lemma

$M_{G,\beta,\mathbf{p}}$ ist zeitreversibel mit stationärer Verteilung $\mathbf{p}$.

Verallgemeinerung

Beweis

- Reversibilität:
  Sei $i \neq j$ und $(i, j) \in E$ (alles andere trivial):

$$\frac{\mathbf{p}_i P_{ij}}{\mathbf{p}_j P_{ji}} = \frac{\mathbf{p}_i \min(1, \frac{\mathbf{p}_j}{\mathbf{p}_i}) \cdot \beta/d}{\mathbf{p}_j \min(1, \frac{\mathbf{p}_i}{\mathbf{p}_j}) \cdot \beta/d}$$

$$= \begin{cases} \frac{\mathbf{p}_i}{\mathbf{p}_j} / \frac{\mathbf{p}_i}{\mathbf{p}_j} & \text{falls } \mathbf{p}_i \leq \mathbf{p}_j \\ \frac{\mathbf{p}_i}{\mathbf{p}_j} \cdot \frac{\mathbf{p}_j}{\mathbf{p}_i} & \text{falls } \mathbf{p}_i > \mathbf{p}_j \end{cases}$$

$$= 1$$

- stationäre Verteilung: klar

Überblick

MCMC: Markov Chain Monte Carlo

- eine Methode für *Sampling*
  - zufällige Auswahl von Elementen aus einer Grundmenge S
  - gemäß einer Wahrscheinlichkeitsverteilung $\mathbf{p}$
- *Idee von MCMC*
  - konstruiere Markovkette $\mathbf{P}$ mit Zustandsmenge S so, dass
  - $\mathbf{p}$ die stationäre Verteilung ist, und
  - mache Random Walk mit Übergangswahrscheinlichkeiten $\mathbf{P}_{ij}$
- *Probleme*
  - woher $\mathbf{P}$?
  - was, wenn $\mathbf{p}$ nicht explizit gegeben
    - sondern nur Zahlen z_i proportional zu den $\mathbf{p}_i$ und
    - S *extrem* groß und
    - Normalisierungsfaktor $Z = \sum_{i \in S} z_i$ nicht handhabbar

Originalliteratur: Metropolis et al. und Hastings

- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H.; Teller, E.:
  Equations of State Calculations by Fast Computing Machines
  Journal of Chemical Physics. 21 (6): 1087–1092. 1953

- Hastings, W.K.
  Monte Carlo Sampling Methods Using Markov Chains and Their Applications
  Biometrika. 57 (1): 97–109, 1970.

laut `https://en.wikipedia.org/wiki/Metropolis-Hastings_algorithm#History` und dort angegebenen Quellen sind die Ursprünge nicht ganz so klar

Metropolis-Hastings: die Markovkette

- gegeben
  - Zustandsmenge S
  - Zahlen z_i für $i \in S$ und damit (implizit) Wahrscheinlichkeiten $p_i = z_i/Z$ mit $Z = \sum_{i \in S} z_i$
  - *proposal matrix* $\mathbf{Q}$
    - Übergangsmatrix einer irreduziblen Markovkette
- definiere für alle $i, j \in S$:
  - *Akzeptanzwahrscheinlichkeiten* α_{ij}
    - für $i \neq j$ und $\mathbf{Q}_{ij} \neq 0$: $\alpha_{ij} = \min\left\{1, \frac{p_j Q_{ji}}{p_i Q_{ij}}\right\}$
    - anderfalls $\alpha_{ij} = 0$
  - *Übergangswahrscheinlichkeiten*:

$$\mathbf{P}_{ij} = \begin{cases} \alpha_{ij}\mathbf{Q}_{ij}, & \text{falls } i \neq j \\ 1 - \sum_{k \neq i} \mathbf{P}_{ik}, & \text{sonst} \end{cases}$$

Metropolis-Hastings: die Markovkette

- gegeben
  - Zustandsmenge S
  - Zahlen z_i für $i \in S$ und damit (implizit) Wahrscheinlichkeiten $p_i = z_i/Z$ mit $Z = \sum_{i \in S} z_i$
  - *proposal matrix* $\mathbf{Q}$
    - Übergangsmatrix einer irreduziblen Markovkette
- definiere für alle $i, j \in S$:
  - *Akzeptanzwahrscheinlichkeiten* α_{ij}
    - für $i \neq j$ und $\mathbf{Q}_{ij} \neq 0$: $\alpha_{ij} = \min\left\{1, \frac{p_j Q_{ji}}{p_i Q_{ij}}\right\} = \min\left\{1, \frac{z_j Q_{ji}}{z_i Q_{ij}}\right\}$
    - anderfalls $\alpha_{ij} = 0$
  - *Übergangswahrscheinlichkeiten*:

$$\mathbf{P}_{ij} = \begin{cases} \alpha_{ij}\mathbf{Q}_{ij}, & \text{falls } i \neq j \\ 1 - \sum_{k \neq i} \mathbf{P}_{ik}, & \text{sonst} \end{cases}$$

Metropolis-Hastings: als Algorithmus

- starte in beliebigem $i_0 \in S$
- Schritt von Zustand $i_t \in S$ zu i_{t+1} in zwei Phasen:
  - wähle gemäß Verteilung $\mathbf{q}_j = \mathbf{Q}(i_t, j)$ zufällig $j \in S$
  - berechne

$$\alpha_{ij} = \begin{cases} \min\left\{1, \frac{p_j Q_{ji}}{p_i Q_{ij}}\right\}, & \text{falls } i \neq j \text{ und } Q_{ij} \neq 0 \\ 0, & \text{sonst} \end{cases}$$

  - wähle

$$i_{t+1} = \begin{cases} j & \text{mit Wahrscheinlichkeit } \alpha_{ij}\mathbf{Q}_{ij} \\ i_t & \text{sonst} \end{cases}$$

Ising Modell

- Modell für Ferromagnetismus
- gegeben (vereinfacht)
  - Gitter L
  - an jedem Knoten i ein *Spin* $\sigma_i \in \{-1, 1\}$
  - *Konfiguration* $\sigma : L \longrightarrow \{-1, 1\}$
  - *Energie* $H(\sigma) = -\sum_{ij} J_{ij}\sigma_i\sigma_j$
    - typisch $J_{ij} = \begin{cases} 1, & \text{falls } i \text{ und } j \text{ direkte Nachbarn} \\ 0, & \text{sonst} \end{cases}$
  - (Konfigurations-)Wahrscheinlichkeit von σ ist
    - proportional zu $e^{-\beta H(\sigma)}$
    - für ein $\beta \geq 0$ «inverse Temperatur»
- gewünscht: Sampling von σ mit Wahrscheinlichkeiten
  $P_\beta(\sigma) = e^{-\beta H(\sigma)}/Z(\beta)$ wobei $Z(\beta) = \sum_\sigma e^{-\beta H(\sigma)}$

Ising Modell: Metropolis-Hastings

- starte mit zufälligem σ^0
- zu gegebenem σ^t
  nutze **Q**: flippe *einen* zufällig gewählten Spin $\rightsquigarrow \sigma'$
- Metropolis-Hastings
  - falls $H(\sigma') \leq H(\sigma)$, wähle $\sigma^{t+1} = \sigma'$.
  - falls $H(\sigma') > H(\sigma)$, wähle

$$\sigma^{t+1} = \begin{cases} \sigma' & \text{mit Wahrscheinlichkeit } e^{\beta(H(\sigma^t)-H(\sigma'))} \\ \sigma & \text{sonst} \end{cases}$$

Überblick

Optimierungsproblem

- gegeben: $f : S \longrightarrow \mathbb{R}$
  - $f^* = \max\{f(x) \mid x \in S\}$ möge existieren (z. B. S endlich)
- gesucht: ein x mit $f(x) = f^*$

Markovketten M_λ

- sei $1 \leq \lambda \in \mathbb{R}_+$ («inverse Temperatur»)
- definiere stationäre Verteilungen $\mathbf{w}_\lambda$ durch

$$\mathbf{w}_{\lambda x} = \frac{\lambda^{f(x)}}{Z(\lambda)}$$

  mit $Z(\lambda) = \sum_{x \in S} \lambda^{f(x)}$
- Markovkette M_λ
  - Metropolis-Algorithmus mit
  - stationärer Verteilung $\mathbf{w}_\lambda$
- falls $f(x) > f(y)$, Übergang von Zustand x nach y mit Wahrscheinlichkeit

$$\lambda^{-(f(x)-f(y))}$$

Simulated Annealing

- Random Walk auf *sich ändernder* Markovkette
- beginne mit $\lambda = 1$, d. h.
  - «zielloses Umherirren»

Simulated Annealing

- Random Walk auf *sich ändernder* Markovkette
- beginne mit $\lambda = 1$, d. h.
  - «zielloses Umherirren»
- erhöhe λ langsam d. h.

Simulated Annealing

- Random Walk auf *sich ändernder* Markovkette
- beginne mit $\lambda = 1$, d. h.
  - «zielloses Umherirren»
- erhöhe λ langsam d. h.
  - vermeide zunehmend neue Zustände y, für die $f(y)$ kleiner als der aktuelle Wert
- für $\lambda \longrightarrow \infty$ ergibt sich stationäre Verteilung, in nur noch Zustände x mit maximalem $f(x) = f^*$ vorkommen.

Simulated Annealing (2)

Es sei $S^* = \{x \mid f(x) = f^*\}$

$$\mathbf{w}_{\lambda x} = \frac{\lambda^{f(x)}}{Z(\lambda)} = \frac{\lambda^{f(x)}}{\sum_{x \in S} \lambda^{f(x)}}$$
$$= \frac{\lambda^{f(x)}/\lambda^{f^*}}{\sum_{x \in S} \lambda^{f(x)}/\lambda^{f^*}}$$
$$= \frac{\lambda^{f(x)}/\lambda^{f^*}}{|S^*| + \sum_{x \in S \setminus S^*} \lambda^{f(x)}/\lambda^{f^*}}$$

Also

$$\lim_{\lambda \to \infty} \mathbf{w}_{\lambda x} = \begin{cases} 1/|S^*| & \text{falls } x \in S^* \\ 0 & \text{sonst} \end{cases}$$