Seminar Simulation und Bildanalyse mit Java

Einführung in Markov-Chain-Monte-Carlo

Monika Trapp

19.05.2003

Motivation und Einleitung

- zeitliche Abläufe weisen Abhängigkeit von der Vergangenheit auf
  - Verkaufszahlen eines Produkts
  - Luftverschmutzung
  - morgige Börsenkurse
- volle Abhängigkeit: $X_n = f(X_0, X_1, \ldots, X_{n-1})$
  - realistisches Abbild der Wirklichkeit
  - große Schwierigkeiten bei der w'keitstheoretischen Behandlung
- Kompromiß: $X_n = f(X_{n-1})$
  - Markov-Prozeß / Markov-Kette $(X_n)_{n\in\mathbb{N}}$

Markov-Prozesse

Definition 1 (Stochastischer Prozess)
Sei $(\Omega, \mathcal{F}, \mathcal{P})$ W'keitsraum, $\mathcal{I} \neq \emptyset$, $(E, \mathcal{E})$ meßbarer Raum.
Eine Familie $(X_n)_{n \in \mathcal{I}}$ von Zufallsvariablen mit Werten in E heißt stochastischer Prozeß mit Parameterbereich $\mathcal{I}$ und Zustandsraum E.

Definition 2 (Markov-Kette)
Sei $(X_n)_{n \in \mathbb{N}}$ ein stochastischer Prozeß mit abzählbarem Zustandsraum E. Dann heißt (X_n) Markov-Kette, wenn gilt:

$$\mathcal{P}\left(X_{n+1} = e_j | X_0 = e_{i0}, \ldots, X_n = e_i\right) = \mathcal{P}\left(X_{n+1} = e_j | X_n = e_i\right) =: p_{ij}$$

$\forall n, i_0, \ldots, i, j \in \mathbb{N}, e_{i_0}, \ldots, e_i, e_j \in E$ *mit* $\mathcal{P}\left(X_0 = e_{i0}, \ldots, X_n = e_i\right) > 0$*,*
so daß $p_{ij} \geq 0$ *und* $\sum_{j \in \mathcal{I}} p_{ij} = 1 \; \forall i, j \in \mathbb{N}$.

Bemerkung 1
Im folgenden sei $E = \{e_1, \ldots, e_k\}$ stets ein endlicher Zustandsraum.

Bayes'sche Modellierung

- Hauptproblem in der stat. Modellbildung:
  Auswertung hochdimensionaler Integrale
  - Bayes'sche Statistik $\Rightarrow$ a-posteriori-Verteilung erfordert Integration über viele Parameter
  - i.a. nicht analytisch lösbar
  - zu aufwendig für praktische Anwendung
- Entwicklung von Verfahren, die näherungsweise Integralberechnung zulassen
- MCMC-Methoden:
  - verbinden Monte-Carlo-Verfahren zur numerischen Berechnung von Integralen mit der Theorie der Markov-Ketten
  - Monte-Carlo-Integration: Zufallszahlen mit gewünschter Ausgangsverteilung, um Erwartungswerte zu schätzen
  - bei mehrdimensionalen Verteilungen oft keine elementaren Generatoren $\Rightarrow$ an Stelle von unabhängigen Realisierungen der Ausgangsverteilung irreduzible Markov-Ketten

Eigenschaften von Markov-Ketten

Definition 3
Sei $(X_n)_{n\in\mathbb{N}}$ Markov-Kette.

- *Dann heißt $(X_n)_{n\in\mathbb{N}}$ homogen / Kette mit stationären Übergangsw'keiten p_{ij}, falls $\forall i,j \in \mathcal{I}\ p_{i,j} := \mathcal{P}\left(X_{n+1} = e_j | X_n = e_i\right)$ unabh. von n ist*
- $p_{ij}^{(m)} := \mathcal{P}\left(X_{n+m} = e_j | X_n = e_i\right)$ *heißt m-Schritt Übergangsw'keit*
- $P_{(m)} = \begin{pmatrix} p_{11}^{(m)} & p_{12}^{(m)} & \cdots & p_{1k}^{(m)} \\ p_{21}^{(m)} & p_{22}^{(m)} & \cdots & p_{2k}^{(m)} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k1}^{(m)} & p_{k2}^{(m)} & \cdots & p_{kk}^{(m)} \end{pmatrix}$ *heißt m-Schritt-Übergangsmatrix, für $P_{(1)}$ schreibt man auch kurz P*
- *die W'keitsfunktion $\alpha = (\alpha_i)_{i\in\mathcal{I}}$, $\alpha_i = \mathcal{P}\left(X_0 = e_i\right)$, heißt Anfangs-/Startverteilung*
- *Die Anfangsverteilung α heißt stationär, falls gilt $\alpha^T = \alpha^T P$.*

MCMC

- umgekehrt zur Frage in der Markovtheorie:
  - existiert stationäre Verteilung zu gegebener Markovkette
  - welcher Gestalt

- hier: erzeuge eine Markov-Kette, deren stationäre Verteilung die vorgegebene Funktion $\Pi(.)$ ist

- eine homogene Markov-Kette ist eindeutig durch die Verteilung des Startwertes X_0 und die Übergangsw'keiten P festgelegt

- genügt sie bestimmten Bedingungen, so nähert sie sich mit der Zeit einem stationären Zustand, der unabh. von der Verteilung des stationären Wertes ist

$\Rightarrow$ Suche nach einer Markov-Kette mit bestimmter stationärer Verteilungsfunktion ist äquivalent zur Bestimmung ihrer Übergangsw'keiten

Eigenschaften

Definition 4 (Eigenschaften von Markov-Ketten)
Eine Markov-Kette heißt

- *irreduzibel, wenn* $\forall i, j \in \{1, \ldots, k\} \exists n \in \mathbb{N} : P_{ij}^n > 0$
- *aperiodisch, wenn* $\forall i \in \{1, \ldots, k\}$*: der ggT von* $\{\ n : P_{ii}^n > 0\ \} = 1$
- *ergodisch, wenn* $\exists \Pi = (\Pi_1, \ldots, \Pi_k)$ *mit*
  - $\sum_{j=1}^{k} \Pi_j = 1, \quad \forall j \in \{1, \ldots, k\} : \quad \Pi_j \geq 0$
  - $\lim_{n \to \infty} P_{ij}^n = \Pi_j$
  - $\sum_{i=1}^{k} \Pi_i P_{ij} = \Pi_j$

  Π *ist dann stationäre Verteilung der Markov-Kette*
- Π *heißt reversibel* $\Leftrightarrow \Pi_i P_{ij} = \Pi_j P_{ji}$

Satz 1 (Aperiodizität)
Sei $\{X_n\}$ *eine aperiodische Markov-Kette. Dann*

$$\exists N < \infty \ \forall i \in \{1, \ldots, k\}, n \geq N : P_{ii}^n > 0$$

Proof.*[Literatur]* $\square$

Korollar 1
Sei $\{X_n\}$ *eine irreduzible, aperiodische Markov-Kette. Dann*

$$\exists M < \infty \ \forall i, j \in \{1, \ldots, k\}, n \geq M : P_{ij}^n > 0$$

Proof.*[Literatur]* $\square$

Satz 2 (Existenz einer stationären Verteilung)
Für jede irreduzible, aperiodische Markov-Kette existiert eine stationäre Verteilung.

Proof.*[Literatur]* $\square$

Satz 3 (Reversibilität)
Sei $\{X_n\}$ eine Markov-Kette mit endlichem Zustandsraum und Übergangsmatrix P. Wenn die Anfangsverteilung α reversibel ist, so ist sie auch stationär.

Proof.[Literatur] □

Definition 5 (Totalvariation)

- *Seien $\nu^{(1)}$ und $\nu^{(2)}$ W'keitsverteilungen auf E. Die Totalvariations-Distanz zwischen $\nu^{(1)}$ und $\nu^{(2)}$ ist definiert als $d_{TV}(\nu^{(1)}, \nu^{(2)}) = 1/2 * \sum_{i=1}^{k} \left|\nu_i^{(1)} - \nu_i^{(2)}\right|$*
- *Seien $\nu^{(1)}, \nu^{(2)}, \ldots, \nu$ W'keitsverteilungen auf E. Dann konvergiert $\nu^{(n)}$ in Totalvariation gegen ν, wenn gilt $\lim_{n\to\infty} d_{TV}(\nu^{(n)}, \nu) = 0$. Man schreibt $\nu^{(n)} \to^{TV} \nu$.*

Satz 4 (Konvergenzsatz für Markov-Ketten)
Sei $\{X_n\}$ eine irreduzible, aperiodische Markov-Kette mit beliebiger Anfangsverteilung $\alpha^{(0)}$. Dann gilt für jede zu $\{X_n\}$ stationäre Verteilung Π : $\alpha^{(n)} \to^{TV} \Pi$

Proof.[Literatur] □

Satz 5 (Eindeutigkeit)

Jede irreduzible, aperiodische Markov-Kette hat genau eine stationäre Verteilung

Beweis. 1. *Sei* $(X_0, X_1, \ldots)$ *eine irreduzible, aperiodische Markov-Kette mit Übergangsmatrix* P. *Nach Satz 2 gibt es mindestens eine stationäre Verteilung für* P.

2. *Seien* Π, Π' *zwei stationäre Verteilungen für* P, $\alpha^{(0)} = \Pi'$ *die Anfangsverteilung. Dann gilt, da* Π' *stationär,* $\alpha^{(n)} = \Pi' \; \forall n$.
3. *Weiter gilt aber nach Satz 4, daß* $\alpha^{(n)} \to^{TV} \Pi$, *d.h.* $0 = \lim_{n\to\infty} d_{TV}(\alpha^{(n)}, \Pi) = \lim_{n\to\infty} d_{TV}(\Pi', \Pi)$.
4. *Da aber* $d_{TV}(\Pi', \Pi)$ *nicht von n abhängt, ist dies 0 und damit* $\Pi = \Pi'$.

$\Rightarrow$ Existenz und Konvergenz sind geklärt

Motivierendes Beispiel: Hard-Core-Modell

Beispiel 1

- $G = (V, E)$ *Graph, Ecken* $E = \{e_1, \ldots, e_k\}$, *Kanten* $V = \{v_1, \ldots, v_l\}$
- *zufällige Zuordnung von 0en / 1en, so daß keine 2 benachbarten Ecken den Wert 1 haben* $\Rightarrow$ *zulässige Konfigurationen aus* $\{0,1\}^{|E|}$
- *Sei* $Z = \{\xi \in \{0,1\}^{|E|}$: *keine 2 benachbarten Ecken haben Wert 1*$\}$
- *wir wählen jedes* $\xi_z \in Z$ *mit gleicher W'keit*

$$\Rightarrow \mu_G(\xi) = \begin{cases} \frac{1}{|Z|} & \text{falls } \xi \in Z, \\ 0 & \text{sonst} \end{cases}$$

- *Sei* $n(\xi) := \{$*Anzahl von 1en in* $\xi\}$
- *Sei* X *zufällig gewählte Konfiguration mit Verteilung* μ_G
- *erwartete Anzahl von 1en in* $X =$

$$\boldsymbol{E}[n(X)] = \sum_{\xi \in \{0,1\}^{|E|}} n(\xi)\mu_G(\xi) = \frac{1}{|Z|} \sum_{\xi \in \{0,1\}^{|E|}} n(\xi) 1_Z = ?$$

Lösung: Markov-Chain-Monte-Carlo

- konstruiere irreduzible, aperiodische Markov-Kette $\{X_n\}$ mit gewünschter stationärer Verteilung Π
- mit jeder beliebigen Anfangsverteilung wird die Verteilung der Kette gg. Π konvergieren
- beliebig nahe Annäherung an Π mit wachsendem n
- Konstruktion einer Markov-Kette mit gewünschter Eigenschaft:
  - setze $E = Z$
  - wähle Ecke e beliebig (mit W'keit $\frac{1}{|E|}$)
  - werfe faire Münze und setze bei Kopf und alle benachbarten Ecken $= 0$ $X_{n+1}(e) = 1$, sonst $X_{n+1}(e) = 0$
  - $\forall \hat{e} \in E \setminus \{e\}: \quad X_{n+1}(\hat{e}) = X_n(\hat{e})$

- die resultierende Markov-Kette ist irreduzibel, aperiodisch und μ_G ist die zugehörige stationäre Verteilungsfunktion

Beweis. *1. offensichtlich ist die Kette irreduzibel und aperiodisch. Somit ist mit Satz 3 nur zu zeigen, daß μ_G reversibel ist.*

2. Seien $\xi, \xi' \in E$. Dann gilt μ_G reversibel $\Leftrightarrow \mu_G(\xi)P_{\xi\xi'} = \mu_G(\xi')P_{\xi'\xi}$.

3. Sei $d := d(\xi, \xi') = \{$Nummer von Ecken mit $e_\xi \neq e_{\xi'}\}$
1.Fall: $d = 0 \Rightarrow \xi = \xi'$
2.Fall: $d = 1 \Rightarrow \exists 1 e \in E : \quad e_\xi \neq e_{\xi'}$
$\Rightarrow \forall \ \ \hat{e}$ Nachbar von e_ξ, $\ \tilde{e}$ Nachbar von $e_{\xi'} : \quad \hat{e} = \tilde{e} = 0$
$\Rightarrow \mu_G(\xi)P_{\xi\xi'} = \frac{1}{|E|}\frac{1}{2k} = \mu_G(\xi')P_{\xi'\xi}$
3. Fall: $d \geq 2 \Rightarrow \mu_G(\xi)P_{\xi\xi'} = 0 = \mu_G(\xi')P_{\xi'\xi}$

4. mit Satz 4 und 5 konvergiert die Verteilung der Markov-Kette daher gegen die gewünschte Verteilung

- eine Zufallsvariable mit der gewünschten Verteilung wäre viel aufwendiger auszuwerten
- Beispiel
  - Generierung von $X_i \sim \mu_G$ mit $i \in \{1, \ldots, N\}$
  - mit den SGGZ konvergiert das arithmetische Mittel dann fast-sicher gegen den Erwartungswert ($\sum_{i=1}^{N} \frac{n(X_i)}{N} \longrightarrow_{N\to\infty} \boldsymbol{E}[n(X)]$ f.-s.)
  - $X := \phi(U)$ mit $U \sim \boldsymbol{U}[0,1]$ und $\phi : [0,1] \to E$
  - $$\phi(x) = \begin{cases} e_1 & \text{für } x \in [0, \Pi_{e_1}), \\ \vdots & \\ e_i & \text{für } x \in [\sum_{j=1}^{i-1} \Pi_{e_j}, \sum_{j=1}^{i} \Pi_{e_j}), \\ \vdots & \\ e_k & \text{für } x \in [\sum_{j=1}^{k-1} \Pi_{e_j}, 1] \end{cases}$$

    $\Rightarrow X \sim \Pi$
- ergibt gewünschte Zufallsvariable, aber hoher numerischer Aufwand

Algorithmen

- 2 Haupttechniken in MCMC:
  - vereinfachen höherdimensionaler Probleme durch sukzessive Simulierung aus verschiedenen niedriger-dimensionalen Komponenten von $X = (X^{(1)}, \ldots, X^{(d)}) \Rightarrow$ Gibbs-Sampler
  - nutze zusätzliche Verwerfungsmethode, um zufällige Markov-Kette so zu korrigieren, daß sich stationäre Verteilung Π ergibt $\Rightarrow$ Metropolis-Hastings-Algorithmus

- wir werden sehen: Gibbs-Sampler ist Spezialfall des M-H-Algorithmus

Metropolis-Hastings-Algorithmus

- multivariate Verteilungsfunktion $\Pi(x) = \Pi(x^{(1)}, \ldots, x^{(d)})$
- $q(x, y)$ Übergangsw'keit (von x nach y), die leicht simuliert werden kann
- Idee: suche einen "Kandidaten" y^* und akzeptiere / verwerfe nach bestimmten Kriterien

- Algorithmus
  1. gegeben sei der n-te Wert $X_n = x$, generiere y^* aus der Fkt $q(x, .)$
  2. berechne Akzeptanzw'keit $\alpha(x, y^*)$ mit

$$\alpha(x, y) = \begin{cases} min\{\frac{\Pi(y)q(y,x)}{\Pi(x)q(x,y)}, 1\} & \text{falls } \Pi(x)q(x, y) > 0, \\ 1 & \text{sonst} \end{cases}$$

  3. akzeptiere y^* mit W'keit $\alpha(x, y^*)$, setze dann $X_{n+1} = y^*$, sonst $X_{n+1} = x$
  4. wiederhole

- Beispiel: Sei E wiederum endlicher Zustandsraum, $Q = (q_{ij})$ beliebige irreduzible Übergangsmatrix
- zu jedem Zeitpunkt $(n+1) \in \mathbb{N}$ mit $X_n = e_i$
  1. generiere Z Zufallsvariable mit $P(Z = e_j) = q_{ij}$
  2. akzeptiere/verwerfe e_j mit W'keit α_{ij}
- die Matrix $P = (q_{ij}\alpha_{ij})$ ist dann die Übergangsmatrix
- die resultierende Markov-Kette ist reversibel und ergodisch
- Π ist die stationäre Verteilung der Kette, denn
  1. nach Theorem 2 ist nur noch z.z $\Pi_i P_{ij} = \Pi_j P_{ji} \Leftrightarrow \Pi_i q_{ij}\alpha_{ij} = \Pi_j q_{ji}\alpha_{ji}$
  2. dies ist aber sofort erfüllt, denn wenn $\alpha_{ij} = \frac{\Pi_j q_{ji}}{\Pi_i q_{ij}}$ ist, dann muss $\alpha_{ji} = 1$ gelten und umgekehrt

Gibbs-Sampler

- multivariate Verteilungsfunktion $\Pi(x) = \Pi(x^{(1)}, \ldots, x^{(d)})$
- Zufallsvariable $X = (X^{(1)}, \ldots, X^{(d)}) \sim \Pi$
- bed. Verteilung der i-ten Komponente $X^{(i)}$: $P(.|X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(d)})$
- Idee: simuliere $X^{(i)}$ aus $P(.|X^{(1)}, \ldots, X^{(i-1)}, X^{(i+1)}, \ldots, X^{(d)})$
- Algorithmus
  1. Initialisierung: $X_0 = (X_0^{(1)}, \ldots, X_0^{(d)})$
  2. simuliere
     - $X_1^{(1)}$ aus $P(.|X_0^{(2)}, \ldots, X_0^{(d)})$
     - $X_1^{(2)}$ aus $P(.|X_1^{(1)}, X_0^{(3)}, \ldots, X_0^{(d)})$
     - usw.
     - $X_1^{(d)}$ aus $P(.|X_1^{(1)}, \ldots, X_1^{(d-1)})$
  3. wiederhole 2.
- bedingte Erwartungen liefern Markov-Kette $\Rightarrow$ falls Konvergenzkriterien erfüllt, dann $X_1, X_2, \ldots \longrightarrow_D \Pi(x^{(1)}, \ldots, x^{(d)})$

- Beispiel: Hard-Core-Modell
- Zufallsvariable $X = (X^{(1)}, \ldots, X^{(d)})$ mit W'keitsfunktion $p(x)$ und Werten aus E
- für beliebiges $Z \subsetneq E$ wollen wir eine Zufallsvariable X generieren mit

$$\forall x \in Z : \quad \Pi_x = \frac{p(x)}{P(X \in Z)}$$

- Algorithmus:
  zu jedem Zeitpunkt $(n+1) \in \mathbb{N}$ mit $X_n = x = (x_1, \ldots, x_i, \ldots, x_d) \in Z$
  1. wähle (zuf. oder det.) eine Koordinate i
  2. $\forall j \neq i : \quad X_{n+1}(j) := x_j$

3. generiere eine Zufallsvariable Y mit W'keit

$$P(Y=y) = P(X(i)=y|X(j)=x_j, j \neq i)$$
$$= \frac{p((x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_k))}{\sum\limits_{x \in Z, X(j)=x_j, j\neq i} p((x_1,\dots,x_{i-1},x_i,x_{i+1},\dots,x_k))} =: p_y$$

4. für $Y=y$ setze $w=(x_1,\dots,x_{i-1},y,x_{i+1},\dots,x_k)$ und

$$X_{n+1} := \begin{cases} w & \text{mit W'keit } \alpha(x,w) = min\{\frac{\Pi(w)q(w,x)}{\Pi(x)q(x,w)},1\}, \\ x & \text{mit W'keit } 1-\alpha(x,w) \end{cases}$$

   wobei $q(x,w) = \frac{1}{|Z|} * p_y$, d.h. hier $\alpha(x,w) \equiv 1$

5. durchlaufe alle anderen i, setze dann $n \mapsto n+1$

MCMC für q-Colouring

Beispiel 2

Sei $G = (V, E), q \geq 4, q \in \mathbb{N}$

Den Ecken E sollen q verschiedene Farben zugewiesen werden, so daß keine 2 benachbarten Pixel die selben Farben besitzen (Definition der Menge Z*).*

- *ein zufälliges* q*-Colouring wird gleichverteilt auf Z generiert*
- *Gibbs-Sampler für zufällige* q*-Colouring:*
  zu jedem Zeitpunkt $(n + 1) \in \mathbb{N}$ *mit* $X_n = x = (x_1, \ldots, x_i, \ldots, x_d) \in Z$
  1. *wähle* $e \in E$ *(gleichverteilt)*
  2. *generiere eine Farbe für* $X_{n+1}(e)$ *(gleichverteilt auf der Menge der für e zulässigen Farben)*
  3. $\forall \hat{e} \in E \setminus \{e\} : \quad X_{n+1}(\hat{e}) = X_n(\hat{e})$
- *Diese Kette ist irreduzibel, aperiodisch und hat die stationäre Verteilung* μ_G
- *Alternativ kann* e *auch systematisch ermittelt werden, periodisch wird zum Zeitpunkt* $i, k + i, 2k + i, \ldots$ *die* k*-te Ecke gewählt*

Weitere Beispiele

Beispiel 3 (Travelling Salesman)
Ein Handlungsreisender soll N Städte genau einmal besuchen und anschließend heimkehren. Gegeben sind die Abstände zwischen je zwei Städten (die nicht unbedingt symmetrisch sein müssen). Gesucht ist die kürzeste Route.

Zu diesem Problem gibt es zahlreiche äquivalente Formulierungen.

Beispiel 4
Ein Lötkolben soll automatisch eine Reihe von Lötstellen auf einen Chip setzen. Die Ausfallrate der Chips wächst mit der Weglänge, die der Lötkolben durchläuft. Deshalb sollte die totale Weglänge minimiert werden.

Wir bleiben der Einfachheit halber beim symmetrischen Travelling Salesman Problem.

- Sei N die Anzahl der Städte und somit oBdA $C = \{1, \ldots, N\}$ die Menge der Städte.
- Der Abstand zwischen Stadt i und j sei $d(i, j) \geq 0$.
- Eine Tour sei eine Abbildung $\rho : C \rightarrow C$, so daß $\forall_{k=1,\ldots,N-1} \rho^k(i) \neq i$ und $\forall_i \ \rho^N(i) = i$, d.h. eine zyklische Permutation von C.
- Die Menge E aller Touren hat $(N-1)!$ Elemente, und die Kosten einer Tour entsprechen der totalen Länge

$$H(\rho) = \sum_{i \in C} d(i, \rho(i))$$

- Eine Tour ρ_1 wird als Nachfolger einer Tour ρ_0 mit W'keit 1 akzeptiert, falls $H(\rho_1) \leq H(\rho_0)$ gilt oder mit W'keit $exp(H(\rho_0) - H(\rho_1))$, falls $H(\rho_1) > H(\rho_0)$ gilt.

Dieses Vorgehen führt wiederum auf den Metropolis-Hastings-Algorithmus

Literaturliste

1. Pierre Brémaud(2001): Markov Chains: Gibbs Fields, Monte Carlo Simulation and Queues, Springer Texts in Applied Mathematics
2. Petros Dellaportas und Gareth Roberts(2002): An Introduction to MCMC
3. Elenonore Gräf(2002): Endliche Markov-Ketten und Monte-Carlo-Simulation, Universität Ulm
4. Olle Häggström(2002): Finite Markov Chains and Algorithmic Applications, Cambridge University Press
5. Jesper Moller(1999): Notes on Markov Chain Monte Carlo Methods, Dina Research Summer School
6. Gerhard Winkler(2000): Stochastische Prozesse in der statistischen Modellierung, Institut für Biomathematik und Biometrie Oberschleißheim