Algorytmy MCMC (Markowowskie Monte Carlo) dla skokowych procesów Markowa

Wojciech Niemiro[1]

Uniwersytet Warszawski i UMK Toruń

XXX lat IMSM, Warszawa, kwiecień 2017

[1] Wspólne prace z Błażejem Miasojedowem, Johnem Noble, Krzysztofem Opalskim.

Plan

Plan

Plan

Skokowe procesy Markowa (czas ciągły, przestrzeń dyskretna)

- $X = \{X(t), t^{\min} \leqslant t \leqslant t^{\max}\}$ jest procesem Markowa na dyskretnej przestrzeni stanów $\mathcal{S}$.
- Intensywność przejścia ze stanu $s \in \mathcal{S}$ do $s' \neq s$:

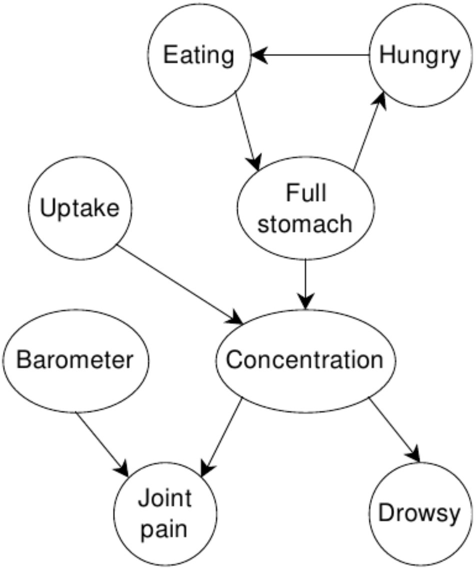
$$\lim_{h \to 0} \frac{1}{h} \mathbb{P}\left(X(t+h) = s' | X(t) = s\right) = Q_t(s, s').$$

- Intensywność wyjścia ze stanu: $Q_t(s) = \sum_{s' \neq s} Q_t(s, s')$.
- Proces jest jednorodny jeśli $Q_t = Q$.

Modele reakcji (bio)chemicznych

- „Równania reakcji”:
  - reakcja 1: $A + 2B \to C$,
  - $\ldots$
  - reakcja k: $\ldots$.
- Stan procesu: $s = (n_A, n_B, n_C, n_D) \in \{0, 1, \ldots\}^4$.
- Intensywność zachodzenia reakcji $Q^{(1)}((n_A, n_B, n_C, n_D), (n_A - 1, n_B - 2, n_C + 1, n_D))$ zależy od stanu.
- Proces jednorodny: $Q = Q^{(1)} + \cdots Q^{(k)}$.

Przykład sieci bayesowskiej z czasem ciągłym (CTBN)



Sieci bayesowskie z czasem ciągłym (CTBN)

Graf skierowany $(\mathcal{N}, \mathcal{E})$. Dla $i \in W$, niech $\mathrm{pa(i)} = \{\mathrm{j : i \to i}\}$. Stan procesu to konfiguracja $s = (s_i)_{i \in W}$, gdzie s_i jest elementem skończonego „alfabetu stanów” wierzchołka i.

CTBN jest jednorodnym procesem Markowa z czasem ciągłym na przestrzeni konfiguracji, z macierzą intensywności

$$Q(s, s') = \begin{cases} Q^{(i)}(s_i, s'_i | s_{\mathrm{pa(i)}}) & \text{jeśli } s_{-i} = s'_{-i} \text{ i } s_i \neq s'_i \\ & \qquad\qquad \text{dla pewnego } i; \\ 0 & \text{jeśli } s_{-i} \neq s'_{-i} \text{ dla wszystkich } i, \end{cases}$$

dla $s \neq s'$.

Ukryte modele Markowa

- Proces X jest nieobserwowany.
- Obserwujemy zmienną losową Y o rozkładzie prawdopodobieństwa $p(Y|X)$.
- Zadanie: obliczyć rozkład *a posteriori* $p(X|Y)$. Powiedzmy, że znamy $p(X)$ i $p(Y|X)$.
- Szczególny przypadek: obserwujemy stan procesu w momentach $t^{\min} \leqslant t_1^{\mathrm{obs}} < \cdots < t_k^{\mathrm{obs}} \leqslant t^{\max}$ z błędem losowym. Formalnie, ciąg zmiennych losowych $Y = \{Y_1, \ldots, Y_k\}$ taki, że $p(Y|X) = \prod_i p(Y_i|X(t_i^{\mathrm{obs}}))$.
- Szczególny przypadek: obserwujemy pewne wierzchołki CTBN, inne są „ukryte”.

Ukryte modele Markowa

- Proces X jest nieobserwowany.
- Obserwujemy zmienną losową Y o rozkładzie prawdopodobieństwa $p(Y|X)$.
- Zadanie: obliczyć rozkład *a posteriori* $p(X|Y)$. Powiedzmy, że znamy $p(X)$ i $p(Y|X)$.

- Szczególny przypadek: obserwujemy stan procesu w momentach $t^{\text{min}} \leqslant t_1^{\text{obs}} < \cdots < t_k^{\text{obs}} \leqslant t^{\text{max}}$ z błędem losowym. Formalnie, ciąg zmiennych losowych $Y = \{Y_1, \ldots, Y_k\}$ taki, że $p(Y|X) = \prod_i p(Y_i|X(t_i^{\text{obs}}))$.
- Szczególny przypadek: obserwujemy pewne wierzchołki CTBN, inne są „ukryte”.

Algorytmy MCMC dla skokowych procesów Markowa

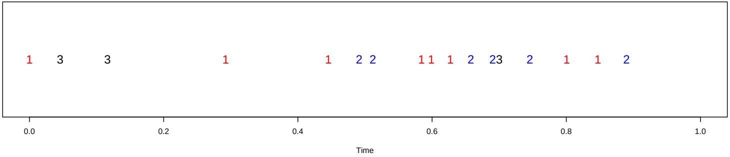
Rozkład *a posteriori* $\pi(X) = p(X|Y)$ jest rozkładem docelowym, Y jest ustalone.

Generujemy łańcuch Markowa $X_0, X_1, \ldots, X_m, \ldots$ taki, że $\pi(X)$ jest rozkładem stacjonarnym i $X_m \to \pi$.

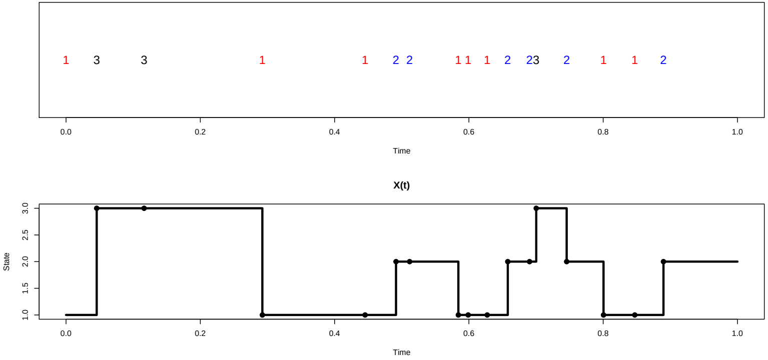
Rozkład empiryczny $\hat{\pi}_m = \dfrac{1}{m} \sum_{j=1}^{m} \delta_{X_j}$ jest aproksymacją π.

Rozkład π jest miarą probabilistyczną na przestrzeni trajektorii skokowego procesu Markowa, $X_m = (X_m(t) : t^{\min} \leqslant t \leqslant t^{\max})$.

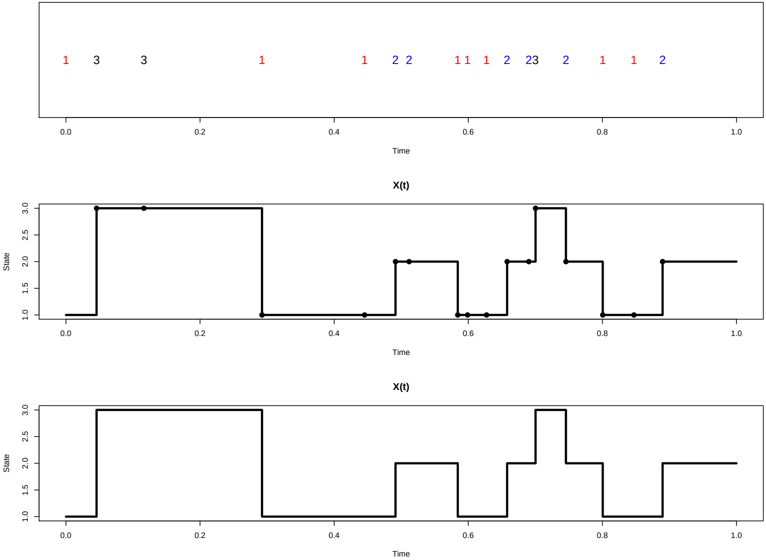
„Nadmiarowa” reprezentacja trajektorii



„Nadmiarowa” reprezentacja trajektorii



„Nadmiarowa” reprezentacja trajektorii



Reprezentacja trajektorii

„Nadmiarowa reprezentacja” procesu $X = (X(t) : t^{\min} \leqslant t \leqslant t^{\max})$:

- $T_1, T_2, \ldots, T_N$ – losowe momenty skoków (być może „wirtualnych”).
- $S_0, S_1, \ldots, S_N$ – „szkielet”: $S_i = X(T_i)$.
- $X(t) = S_{i-1}$ dla $T_{i-1} \leqslant t < T_i$.
- T_i jest skokiem wirtualnym jeśli $S_i = S_{i-1}$.

$$X = \begin{pmatrix} t^{\min} & T_1 & \cdots & T_i & \cdots & T_N & t^{\max} \\ S_0 & S_1 & \cdots & S_i & \cdots & S_N & \end{pmatrix},$$

gdzie $N = \max\{n : T_n < 1\}$.

„Uniformizacja”

Niech $\lambda \geqslant \max_s Q(s)$, gdzie $Q(s) = \sum_{s'} Q(s, s')$.

- $T_1, T_2, \ldots, T_N$ – jednorodny proces Poissona z intensywnością λ.
- $S_0, S_1, \ldots, S_N$ – jednorodny łańcuch Markowa z prawdopodobieństwami przejścia

$$P(s, s') = \begin{cases} \dfrac{Q(s, s')}{\lambda} & \text{jeśli } s \neq s' \text{ (Rzeczywisty skok)} \\[2ex] 1 - \dfrac{Q(s)}{\lambda} & \text{jeśli } s = s' \text{ (Wirtualny skok)} \end{cases}.$$

Algorytm Metropolisa-Hastingsa

B. Miasojedow, WN, J. Noble, K. Opalski (2014, 2016).

Niech X oznacza nadmiarową reprezentację trajektorii, π jest rozkładem *a posteriori*.

Krok algorytmu MCMC: $X_m = X \to X' = X_{m+1}$.

Generujemy „propozycję" $X^* \sim q(X, \cdot)$ i akceptujemy ruch $X \to X'$ z prawdopodobieństwem

$$\alpha(X, X^*) = 1 \wedge \frac{\pi(X^*)q(X^*, X)}{\pi(X)q(X, X^*)}.$$

- Jeśli akceptujemy, to $X' = X^*$.
- Jeśli nie, to $X' = X$.

Prawdopodobieństwo akceptacji jest tak dobrane, że π jest rozkładem stacjonarnym.

Algorytm Metropolisa-Hastingsa

4 typy „ruchów”:

ChangeTime: $T_i \to T_i'$

ChangeState: $S_i \to S_i'$

RemoveJump/AddJump: $T = (T_1, \ldots, T_{i-1}, T_i, T_{i+1}, \ldots, T_N) \leftrightarrow T' = (T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_N)$.

+ odpowiednie prawdopodobieństwa akceptacji.

Algorytm Rao i Teha

V. Rao, Y.W. Teh (2012, 2013).

$(T, S) = (J, V, S) = (\text{Rzeczywiste skoki}, \text{Wirtualne skoki}, \text{Szkielet})$.

Próbnik Gibbsa:

$$(T, S) = (J, V, S) \to (J, V', S) = (T', S) \to (T', S').$$

Geometryczna ergodyczność algorytmu Rao-Teha

TWIERDZENIE (B. Miasojedow i WN)

Załóżmy, że macierz intensywności Q *jest nieprzywiedlna i* $\lambda > \max_s Q(s)$. *Wtedy łańcuch* X_m *generowany przez algorytm Rao i Teha jest geometrycznie ergodyczny. Istnieje stała* $\gamma < 1$ *i funkcja* M *takie, że dla każdego* X_0,

$$\|p(X_m|X_0) - \pi(X)\|_{TV} \leqslant \gamma^m M(X_0).$$

Geometryczna ergodyczność algorytmu Metropolisa-Hastingsa

TWIERDZENIE (B. Miasojedow, WN, J. Noble, K. Opalski)

Załóżmy dodatkowo, że $\inf_{s \neq s'} Q(s, s') \geqslant \delta > 0$*. Wtedy łańcuch* X_m *generowany przez algorytm Metropolisa-Hastingsa (skonstruowany przez nas) jest geometrycznie ergodyczny.*

Przykład

Niech $S = \{0, 1, 2, 3, 4\}$, $Q(s, s + 1 \bmod 5) > 0$ i $Q(s, s - 1 \bmod 5) > 0$, pozostałe intensywności $= 0$. Łańcuch generowany przez algorytm Metropolisa-Hastingsa *nie jest* geometrycznie ergodyczny.

Geometryczna ergodyczność algorytmu Metropolisa-Hastingsa

TWIERDZENIE (B. Miasojedow, WN, J. Noble, K. Opalski)

Załóżmy dodatkowo, że $\inf_{s \neq s'} Q(s, s') \geqslant \delta > 0$*. Wtedy łańcuch* X_m *generowany przez algorytm Metropolisa-Hastingsa (skonstruowany przez nas) jest geometrycznie ergodyczny.*

Przykład

Niech $\mathcal{S} = \{0, 1, 2, 3, 4\}$, $Q(s, s + 1 \bmod 5) > 0$ i $Q(s, s - 1 \bmod 5) > 0$, pozostałe intensywności $= 0$. Łańcuch generowany przez algorytm Metropolisa-Hastingsa *nie jest* geometrycznie ergodyczny.

- V. Rao, Y. W. Teh. MCMC for continuous-time discrete-state systems. In *Advances in Neural Information Processing Systems*, 701–709, 2012.
- V. Rao, Y. W. Teh. Fast MCMC sampling for Markov jump processes and extensions. *Journal of Machine Learning Research*, 14, 3207–3232, 2013.
- B. Miasojedow, W. Niemiro, J. Noble, K. Opalski. Metropolis-type algorithms for Continuous Time Bayesian Networks. `arXiv:1403.4035v1`.
- B. Miasojedow, W. Niemiro. Geometric ergodicity of Rao and Teh's algorithm for homogeneous Markov jump processes. *Statistics and Probability Letters* 113, 1–6, 2016.