Nieasymptotyczne oszacowania błędu dla regeneracyjnych algorytmów MCMC[1]

Wojciech Niemiro

Uniwersytet Mikołaja Kopernika, Toruń i Uniwersytet Warszawski

Statystyka Matematyczna
Wisła, grudzień 2009

[1] Współautorzy: Krzysztof Łatuszyński, University of Warwick, U.K. i Błażej Miasojedow, Uniwersytet Warszawski,
Praca finansowana przez Grant No. N N201387234.

Plan

1. Wstęp
   - Obliczanie całek metodą MCMC
   - Oszacowania dokładności
2. Regeneracja
   - Mały zbiór
   - Regeneracja
   - Estymator sekwencyjno-regeneracyjny
3. Główny wynik
   - Błąd średniokwadratowy
   - Warunek dryfu
   - Jawne oszacowania przy warunku dryfu

Plan

Plan

1. Wstęp
   - Obliczanie całek metodą MCMC
   - Oszacowania dokładności
2. Regeneracja
   - Mały zbiór
   - Regeneracja
   - Estymator sekwencyjno-regeneracyjny
3. Główny wynik
   - Błąd średniokwadratowy
   - Warunek dryfu
   - Jawne oszacowania przy warunku dryfu

Obliczanie całek metodą MCMC

Obliczyć całkę

$$\theta = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x,$$

gdzie

- $\mathcal{X}$ – przestrzeń wielowymiarowa,
- π gęstość rozkładu prawdopodobieństwa na $\mathcal{X}$, zwykle znana tylko z dokładnością do stałej normującej.

Metoda MCMC: generujemy (symulujemy) **łańcuch Markowa**

$$X_0, X_1, \ldots, X_t, \ldots \qquad \pi_t(A) = \mathbb{P}(X_t \in A) \to \pi(A), \quad (t \to \infty).$$

Estymator MCMC:

$$\hat{\theta}_T = \frac{1}{n}\sum_{i=0}^{T-1} f(X_i) \to \theta \qquad (T \to \infty).$$

Obliczanie całek metodą MCMC

Obliczyć całkę

$$\theta = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x,$$

gdzie

- $\mathcal{X}$ – przestrzeń wielowymiarowa,
- π gęstość rozkładu prawdopodobieństwa na $\mathcal{X}$, zwykle znana tylko z dokładnością do stałej normującej.

Metoda MCMC: generujemy (symulujemy) **łańcuch Markowa**

$$X_0, X_1, \ldots, X_t, \ldots \qquad \pi_t(A) = \mathbb{P}(X_t \in A) \to \pi(A), \quad (t \to \infty).$$

Estymator MCMC:

$$\hat{\theta}_T = \frac{1}{n}\sum_{i=0}^{T-1} f(X_i) \to \theta \qquad (T \to \infty).$$

Obliczanie całek metodą MCMC

Obliczyć całkę

$$\theta = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x,$$

gdzie

- $\mathcal{X}$ – przestrzeń wielowymiarowa,
- π gęstość rozkładu prawdopodobieństwa na $\mathcal{X}$, zwykle znana tylko z dokładnością do stałej normującej.

Metoda MCMC: generujemy (symulujemy) **łańcuch Markowa**

$$X_0, X_1, \ldots, X_t, \ldots \qquad \pi_t(A) = \mathbb{P}(X_t \in A) \to \pi(A), \quad (t \to \infty).$$

Estymator MCMC:

$$\hat{\theta}_T = \frac{1}{n} \sum_{i=0}^{T-1} f(X_i) \to \theta \qquad (T \to \infty).$$

Obliczanie całek metodą MCMC

Obliczyć całkę

$$\theta = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x,$$

gdzie

- $\mathcal{X}$ – przestrzeń wielowymiarowa,
- π gęstość rozkładu prawdopodobieństwa na $\mathcal{X}$, zwykle znana tylko z dokładnością do stałej normującej.

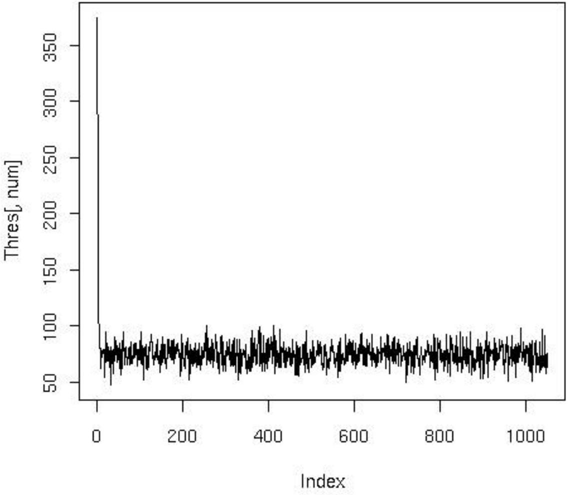
Metoda MCMC: generujemy (symulujemy) **łańcuch Markowa**

$$X_0, X_1, \ldots, X_t, \ldots \qquad \pi_t(A) = \mathbb{P}(X_t \in A) \to \pi(A), \quad (t \to \infty).$$

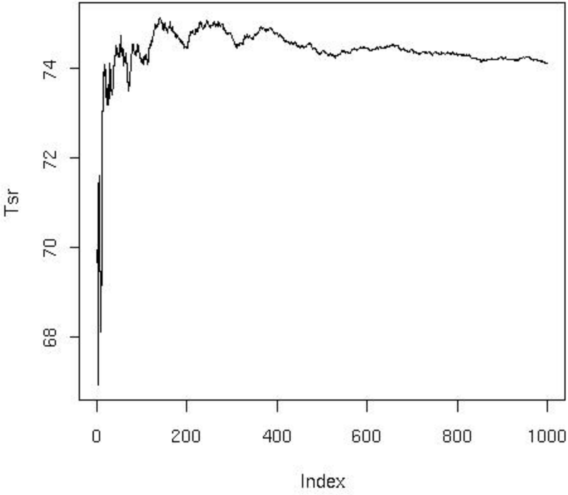
Estymator MCMC:

$$\hat{\theta}_T = \frac{1}{n} \sum_{i=0}^{T-1} f(X_i) \to \theta \qquad (T \to \infty).$$

Zbieżność do rozkładu stacjonarnego:



Zbieżność średnich do estymatora bayesowskiego:



Oszacowania dokładności

Błąd średniokwadratowy:

$$\text{MSE} \;=\; \mathbb{E}\,(\hat{\theta}_T - \theta)^2 \leq ?$$

Poziom ufności:

$$\mathbb{P}(|\hat{\theta}_T - \theta| > \varepsilon) \leq ?$$

Oszacowania dokładności

Błąd średniokwadratowy:

$$\text{MSE} = \mathbb{E}\,(\hat{\theta}_T - \theta)^2 \leq ?$$

Poziom ufności:

$$\mathbb{P}(|\hat{\theta}_T - \theta| > \varepsilon) \leq ?$$

Mały zbiór

Łańcuch Markowa X_n na przestrzeni $\mathcal{X}$ z jądrem przejścia:

$$P(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x).$$

ZAŁOŻENIA

- ***Rozkład stacjonarny.*** *Istnieje rozkład π na $\mathcal{X}$ taki, że $\pi P = \pi$ i P jest π-nieprzywiedlny.*
- ***Mały zbiór.*** *Istnieją: $J \subseteq \mathcal{X}$ z $\pi(J) > 0$, miara probabilistyczna ν i $\beta > 0$ takie, że*

$$P(x, \cdot) \geq \beta \mathbb{I}(x \in J) \nu(\cdot).$$

Mały zbiór

Łańcuch Markowa X_n na przestrzeni $\mathcal{X}$ z jądrem przejścia:

$$P(x, A) = \mathbb{P}(X_n \in A | X_{n-1} = x).$$

ZAŁOŻENIA

- ***Rozkład stacjonarny.*** *Istnieje rozkład π na $\mathcal{X}$ taki, że $\pi P = \pi$ i P jest π-nieprzywiedlny.*
- ***Mały zbiór.*** *Istnieją: $J \subseteq \mathcal{X}$ z $\pi(J) > 0$, miara probabilistyczna ν i $\beta > 0$ takie, że*

$$P(x, \cdot) \geq \beta \mathbb{I}(x \in J) \nu(\cdot).$$

Regeneracja

„Resztowe" jądro przejścia:

$$Q(x,\cdot) := \frac{P(x,\cdot) - \beta\mathbb{I}(x \in J)\nu(\cdot)}{1 - \beta\mathbb{I}(x \in J)}.$$

- Jeśli $X_{n-1} \notin J$ to losuj $X_n \sim P(X_{n-1},\cdot)$, **nie ma regeneracji**;
- Jeśli $X_{n-1} \in J$ to
  - z prawdopodobieństwem $1-\beta$ losuj $X_n \sim Q(X_{n-1},\cdot)$, **nie ma regeneracji**;
  - z prawdopodobieństwem β losuj $X_n \sim \nu(\cdot)$, **Regeneracja!**

Remark: Faktycznie losowanie $Q(X_{n-1},\cdot)$ **nie** jest konieczne (Mykland et al. 1995).

Czasy odnowienia

Momenty regeneracji (odnowienia) dzielą łańcuch na **niezależne** „bloki” losowej długości.

$$\underbrace{X_0, \dots, X_{T_1 - 1}}_{T_1}, \quad \underbrace{X_{T_1}, \dots, X_{T_2 - 1}}_{T_2 - T_1}, \quad \underbrace{X_{T_2}, \dots, X_{T_3 - 1}}_{T_3 - T_2}, \quad \dots$$

$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \text{R} & \text{R} & \text{R} \end{array}$$

Dodatkowo zakładamy, że $X_0 \sim (\pi(\cdot|J)$, wtedy 0 jest momentem odnowienia. i wszystkie bloki mają **ten sam rozkład**.

Estymator sekwencyjno-regeneracyjny

$0 = T_0, T_1, \ldots, T_r, \ldots$ – losowe momenty regeneracji.

$$R(n) := \min\{r : T_r \geq n\}.$$

Estymator:

$$\hat{\theta}_{T_{R(n)}} = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i).$$

$$\begin{array}{ccccc} 0, \ldots, T_1 - 1, & T_1, \ldots\ldots, & T_{R(n)-1}, \ldots, & n, \ldots, T_R(n) - 1, & T_{R(n)}, \ldots \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \mathrm{R} & \mathrm{R} & \mathrm{R} & n & \mathrm{R} \end{array}$$

Błąd średniokwadratowy

Całka i jej estymator MCMC:

$$\theta = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x, \qquad \hat{\theta}_{T_{R(n)}} = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i)$$

TWIERDZENIE

Zakładamy: **Rozkład stacjonarny** *i* **Mały zbiór**. *Wtedy*

$$(i) \quad MSE = \mathbb{E}\,(\hat{\theta}_{T_{R(n)}} - \theta)^2 \leq \frac{\sigma^2_{\mathrm{as}}(f)}{n}\left(1 + \frac{n_0}{n}\right),$$

$$(ii) \quad \mathbb{E}\,T_{R(n)} \leq n + n_0,$$

gdzie

$$\sigma^2_{\mathrm{as}}(f) = \frac{1}{\mathbb{E}\,T_1}\mathbb{E}\left[\sum_{i=0}^{T_1-1}(f(X_i) - \theta)\right]^2, \qquad n_0 = \frac{\mathbb{E}\,T_1^2}{\mathbb{E}\,T_1} - 1.$$

Błąd średniokwadratowy

Całka i jej estymator MCMC:

$$\theta = \pi(f) = \int_{\mathcal{X}} f(x)\pi(x)\mathrm{d}x, \qquad \hat{\theta}_{T_{R(n)}} = \frac{1}{T_{R(n)}} \sum_{i=0}^{T_{R(n)}-1} f(X_i)$$

TWIERDZENIE

Zakładamy: **Rozkład stacjonarny** *i* **Mały zbiór**. *Wtedy*

$$(i) \quad MSE \;=\; \mathbb{E}\,(\hat{\theta}_{T_{R(n)}} - \theta)^2 \leq \frac{\sigma^2_{\mathrm{as}}(f)}{n}\left(1 + \frac{n_0}{n}\right),$$

$$(ii) \quad \mathbb{E}\, T_{R(n)} \leq n + n_0,$$

gdzie

$$\sigma^2_{\mathrm{as}}(f) = \frac{1}{\mathbb{E}\, T_1}\mathbb{E}\left[\sum_{i=0}^{T_1-1}(f(X_i) - \theta)\right]^2, \qquad n_0 = \frac{\mathbb{E}\, T_1^2}{\mathbb{E}\, T_1} - 1.$$

Błąd średniokwadratowy

Uwaga

- *Nieasymptotyczne oszacowanie z „asymptotycznie poprawnym wyrazem wiodącym”:*

$$MSE = \mathbb{E}\,(\hat{\theta}_{T_{R(n)}} - \theta)^2 \leq \frac{\sigma^2_{\mathrm{as}}(f)}{n}\left(1 + \frac{n_0}{n}\right),$$

bo

$$\sqrt{n}(\hat{\theta}_{T_{R(n)}} - \theta) \to \mathcal{N}(0, \sigma^2_{\mathrm{as}}(f)) \qquad (n \to \infty).$$

- *“Niemal deterministyczna” długość symulacji:*

$$\mathbb{E}\,T_{R(n)} \leq n + n_0.$$

Błąd średniokwadratowy

Uwaga

- *Nieasymptotyczne oszacowanie z „asymptotycznie poprawnym wyrazem wiodącym”:*

$$MSE = \mathbb{E}\,(\hat{\theta}_{T_{R(n)}} - \theta)^2 \leq \frac{\sigma^2_{\mathrm{as}}(f)}{n}\left(1 + \frac{n_0}{n}\right),$$

bo

$$\sqrt{n}(\hat{\theta}_{T_{R(n)}} - \theta) \to \mathcal{N}(0, \sigma^2_{\mathrm{as}}(f)) \qquad (n \to \infty).$$

- *“Niemal deterministyczna” długość symulacji:*

$$\mathbb{E}\, T_{R(n)} \leq n + n_0.$$

Błąd średniokwadratowy

Uwaga

- *Nieasymptotyczne oszacowanie z „asymptotycznie poprawnym wyrazem wiodącym”:*

$$MSE = \mathbb{E}\,(\hat{\theta}_{T_{R(n)}} - \theta)^2 \le \frac{\sigma^2_{\mathrm{as}}(f)}{n}\left(1 + \frac{n_0}{n}\right),$$

bo

$$\sqrt{n}(\hat{\theta}_{T_{R(n)}} - \theta) \to \mathcal{N}(0, \sigma^2_{\mathrm{as}}(f)) \qquad (n \to \infty).$$

- *“Niemal deterministyczna” długość symulacji:*

$$\mathbb{E}\, T_{R(n)} \le n + n_0.$$

Warunek dryfu do małego zbioru

Dryf do małego zbioru J

ZAŁOŻENIE

- ***Dryf.*** *Istnieje funkcja* $V : \mathcal{X} \to [1, \infty[$, *stałe* $\lambda < 1$ *i* $K < \infty$ *takie, że*

$$PV^2(x) := \int_{\mathcal{X}} P(x, dy) V^2(y) \leq \begin{cases} \lambda^2 V^2(x) & \text{dla } x \notin J, \\ K^2 & \text{dla } x \in J. \end{cases}$$

Uwaga

Notacja V^2, λ^2, K^2 *– upraszcza wzory.*

Jawne oszacowania przy warunku dryfu

Oszacowania na stałe $\sigma^2_{\mathrm{as}}(f), n_0$.

TWIERDZENIE

Przy warunku **Dryfu***,*

$$\sigma^2_{\mathrm{as}}(f) \leq \|\bar{f}\|^2_V \frac{K^2(2+\beta) - 2K(2\lambda+\beta) + 2\lambda^2 + 2\lambda\beta - \lambda^2\beta}{(1-\lambda)^2\beta},$$

gdzie $\|\bar{f}\|_V := \sup_x |\bar{f}(x)|/V(x) < \infty$ *i* $\bar{f} := f - \theta$*,*

$$n_0 \leq \frac{2}{(1-\lambda)\beta}\left[K\frac{1-\lambda(1-\beta)}{1-\lambda} - \beta\left(1 + \frac{\lambda^2}{1-\lambda}\right) - \lambda\right].$$

Problem

Czy można otrzymać praktycznie użyteczne i dające się obliczyć oszacowania na $\sigma^2_{\mathrm{as}}(f)$?