Oszacowania błędów estymatorów stosowanych w markowowskich metodach Monte Carlo

Błażej Miasojedow

autoreferat pracy doktorskiej

W wielu modelach statystyki bayesowskiej kluczowym problemem jest obliczanie całek względem rozkładów *a posteriori*, które są skomplikowane i możliwe jest jedynie wyznaczenie ich gęstości z dokładnością do stałej normującej. W tej sytuacji najczęściej stosowanym i bardzo skutecznym narzędziem są markowowskie metody Monte Carlo (*Markov Chain Monte Carlo*, MCMC). Jest to rodzina algorytmów, które pozwalają wyznaczyć przybliżone wartości całki

$$\theta := \pi f = \int_{\mathcal{X}} f(x)\pi(dx), \tag{1}$$

gdzie π jest rozkładem znanym z dokładnością do stałej normującej, $\mathcal{X}$ jest podzbiorem przestrzeni wielowymiarowej, a f jest funkcją o wartościach rzeczywistych. Procedura Monte Carlo do aproksymacji całki (1) polega na symulowaniu ergodycznego łańcucha Markowa $(X_n)_{n\geqslant 0}$ o jądrze przejścia $P(x, A)$ i rozkładzie stacjonarnym π, czyli takiego, że X_n zbiega według rozkładu do π. Następnie aproksymuje się nieznaną wielkość θ za pomocą estymatora

$$\hat{\theta}_n := \frac{1}{n}\sum_{i=0}^{n-1} f(X_i). \tag{2}$$

Pierwszym algorytmem pozwalającym na generowanie takich łańcuchów Markowa był algorytm Metropolisa, zaproponowany w [26]. Od tego czasu pojawiło się wiele róznych tego typu procedur oraz metod estymacji θ opartych na algorytmach MCMC [28, 25, 6]. Badanie jakości takich algorytmów jest kluczowe dla oceny jakości i przydatności wyników obliczeń w modelu bayesowskim.

Dostępna jest bogata literatura dotycząca zarówno algorytmów MCMC jak i łańcuchów Markowa w zastosowaniu do MCMC. Wiele prac dotyczy badania tempa zbieżności do rozkładu stacjonarnego [3, 7, 13, 33, 34], równie wiele prac jest poświęconych badaniu asymptotycznych własności estymatorów [14, 15, 27]. Wyniki dotyczące tempa zbieżności do rozkładu stacjonarnego pozwalają tylko kontrolować obciążenie estymatorów i nie przekładają się ani na oszacowania błędu średniokwadratowego ani na konstrukcje przedziałów ufności. Natomiast wyniki asymptotyczne mogą okazać się w praktyce bezużyteczne, a nawet wprowadzające w błąd [31].

Podstawowym celem pracy jest uzyskanie nieasymptotycznych oszacowań błędów estymatorów (1) nieznanej wielkości θ. Tego typu wyniki istnieją tylko przy dość ograniczających założeniach: dla dyskretnej przestrzeni stanów $\mathcal{X}$ i ograniczonej funkcji f [1, 9, 23, 24, 28], w przypadku łańcuchów jednostajnie ergodycznych na ogólnej przestrzeni stanów i ograniczonej funkcji f [10, 16]. Założenie o jednostajnej ergodyczności w praktyce oznacza, że przestrzeń stanów $\mathcal{X}$ jest zwarta, co jest rzadko spotykane w zastosowaniach statystycznych.

Niewiele jest znanych wyników przy bardziej realistycznych założeniach: dla odwracalnych łańcuchów Markowa oszacowania błędu średniokwadratowego $MSE = \mathbb{E}(\hat{\theta}_n - \theta)^2$ za pomocą luki spektralnej w $L^2(\pi)$ można znaleźć w [32], dla geometrycznie ergodycznych łańcuchów na ogólnej przestrzeni stanów MSE jest oszacowany w [21] lub [19].

W rozdziale 3 pracy doktorskiej są przedstawione oszacowania MSE dla geometrycznie ergodycznych łańcuchów Markowa. Rozdział ten oparty jest na pracy [22], w której zawarte są oszacowania MSE dla sekwencyjnego estymatora o losowej długości symulacji. Jeżeli spełniony warunek małego zbioru

$$P(x, A) \geqslant \beta \mathbb{I}(x \in C)\nu(A), \quad \forall x \in \mathcal{X}, \forall A \in \mathcal{B}(\mathcal{X}), \tag{3}$$

gdzie $\beta > 0$ oraz ν jest miarą probabilistyczną, to można korzystać z techniki regeneracji. Technika ta była przedstawiona równolegle w pracach [29] oraz [2], pozwala ona na rozbicie sumy $S_n := \sum_{i=0}^{n-1} f(X_i)$ na mniejsze bloki losowej długości, które są niezależne. Za pomocą tej techniki wykazane jest następujące twierdzenie.

Twierdzenie 4. *Jeżeli zachodzi (3), to*

$$\sqrt{\mathbb{E}_\xi\,(\hat{\theta}_n - \theta)^2} \leqslant \frac{\sigma_{\mathrm{as}}(f)}{\sqrt{n}}\left(1 + \frac{C_0}{n}\right) + \frac{C_1(f)}{n} + \frac{C_2(f)}{n},$$

gdzie ξ oznacza rozkład początkowy.

W tym twierdzeniu $\sigma^2_{\mathrm{as}}(f)$ jest asymptotyczną wariancją występującą w Centralnym Twierdzeniu Granicznym (CTG), co powoduje, że wynik ten jest asymptotycznie optymalny, tzn.

$$\lim_{n\to\infty} n\mathbb{E}_\xi\,(\hat{\theta}_n - \theta)^2 = \sigma^2_{\mathrm{as}}(f)$$

Dodatkowo w tym rozdziale przedstawione są jawne oszacowania stałych występujących w Twierdzeniu 4, przy założeniu warunku dryfu do małego zbioru. Załóżmy, że istnieje funkcja $V \geqslant 1$, stałe $\lambda < 1, K < \infty$ oraz zbiór C spełniający (3) takie, że:

$$PV(x) \leqslant \begin{cases} \lambda V(x) & \text{dla } x \notin C, \\ K & \text{dla } x \in C. \end{cases}$$

Wtedy można oszacować $\sigma_{\mathrm{as}}(f), C_0, C_1(f), C_2(f)$ w terminach wielkości β, λ, K oraz $\sup_{x\in\mathcal{X}} \frac{|f(x)|}{V(x)}$.

Inne podejście do oszacowań dokładności estymatorów MCMC jest związane z nierównościami wykładniczymi. Pozwalają one na budowę dokładnych (nieasymptotycznych) przedziałów ufności. W rozdziale 4 pracy doktorskiej przedstawione są uogólnienia nierówności Hoeffdinga [12] dla łańcuchów Markowa. Podstawowy wynik jest otrzymany przy następującym założeniu

$$\|P - \Pi\|_{L^2(\pi)} = \rho < 1, \tag{5}$$

gdzie $\|\cdot\|_{L^2(\pi)}$ oznacza normę operatorową w przestrzeni $L^2(\pi)$ funkcji całkowalnych w kwadracie względem rozkładu π z iloczynem skalrnym

$$\langle f, g \rangle := \int_{\mathcal{X}} f(x)g(x)\pi(dx).$$

Twierdzenie 6 (Nierówności Hoeffdinga)**.** *Załóżmy, że $0 \leqslant f \leqslant 1$. Przy założeniu (5) dla każdego $0 < \varepsilon < 1 - \theta$ zachodzą następujące nierówności:*

$$\begin{aligned} \mathbb{P}_\pi\left(S_n \geqslant n(\theta + \varepsilon)\right) &\leqslant \left[\frac{\theta + \bar{\theta}\rho}{1 - 2(\bar{\theta} - \varepsilon)/(1 + \sqrt{\Delta})}\right]^{n(\theta+\varepsilon)} \left[\frac{\bar{\theta} + \theta\rho}{1 - 2(\theta + \varepsilon)/(1 + \sqrt{\Delta})}\right]^{n(\bar{\theta}-\varepsilon)} \\ &\leqslant \exp\left\{-2\frac{1-\rho}{1+\rho}\varepsilon^2 n\right\}, \end{aligned}$$

gdzie $\bar{\theta} = 1 - \theta$ oraz $\Delta = 1 + \frac{4\rho(\theta+\varepsilon)(\bar{\theta}-\varepsilon)}{\theta\bar{\theta}(1-\rho)^2}$.

Wynik ten jest uogólnieniem rezultatu z pracy [23], która zawiera analogiczny rezultat dla odwracalnych łańcuchów na skończonej przestrzeni stanów. Założenie (5) w przypadku odwracalnym jest równoważne geometrycznej ergodyczności. Dla nieodwracalnych łańcuchów Markowa (5) wyznacza węższą klasę łańcuchów Markowa. Klasa ta opisana jest przez własności jądra przejścia $P(x, A)$ oraz jądra przejścia $P^*(x, a)$ sprzężonego do $P(x, A)$ zdefinowanego przez równanie

$$\pi(dx)P(x, dy) = \pi(dy)P^*(y, dx).$$

W przypadku niezależnych zmiennych losowych własność posiadania ogonów wykładniczych jest wystarczająca do uzyskania wykładniczych nierówności dla wielkich odchyleń. Niestety dla łańcuchów Markowa ta własność nie jest wystarczająca do zagwarantowania istnienia funkcji tworzącej momenty dla sumy S_n. W rozdziale 4 pracy doktorskiej przedstawione są również nierówności podwykładnicze dla prawdopodobieństw błędu estymatora przy założeniu, że istnieją stałe $a > 0$ i $b > 0$ oraz $M_0 > 0$ takie, że dla każdego $M > M_0$ zachodzi

$$\mathbb{P}_\pi(|f(X)| > \frac{M}{2}) \leqslant ae^{-bM}. \tag{7}$$

Wówczas dla dostatecznie dużych n oraz dostatecznie małych ε zachodzi następujący wniosek z Twierdzenia 6:

Wniosek 8. *Niech X_n będzie łańcuchem Markowa z rozkładem stacjonarnym π i operatorem przejścia P spełniającym $\|P - \Pi\|_{L^2(\pi)} \leqslant \rho$. Niech f będzie funkcją $f : \mathcal{X} \to \mathbb{R}$ oraz niech $\theta := \pi f$ oraz dla każdego $M > M_0$ zachodzi $\mathbb{P}_\pi(|f(x)| > \frac{M}{2}) \leqslant ae^{-bM}$. Wówczas dla dowolnego rozkładu początkowego ξ absolutnie ciągłego względem π, dla dowolnego $1 \leqslant p, q \leqslant \infty$ takich, że $\frac{1}{p} + \frac{1}{q} = 1$ zachodzi:*

$$\mathbb{P}_\xi\left(S_n \geqslant n(\theta + \varepsilon)\right) \leqslant \left\|\frac{d\xi}{d\pi}\right\|_p 2\exp\left\{-\frac{2}{q}\frac{1-\rho}{1+\rho}\varepsilon^2\frac{n}{{M^*}^2}\right\},$$

gdzie M^* jest rozwiązaniem

$$M^3 - M^2\frac{\log(an)}{b} - 2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2 = 0.$$

Ponadto

$$\max\left\{\left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}, \frac{\log(an)}{b}\right\} \leqslant M^* \leqslant 1.465571\max\left\{\left(2\frac{1-\rho}{b(1+\rho)}n\varepsilon^2\right)^{\frac{1}{3}}, \frac{\log(an)}{b}\right\}.$$

Dla odwracalnych łańcuchów Markowa użytecznym narzędziem jest pojęcie luki spektralnej w przestrzeni $L^2(\pi)$. Pozwala ono na: łatwiejsze i w wielu przypadkach dokładniejsze oszacowanie asymptotycznej wariancji [8, 30], na osłabienie założeń CTG [30] oraz na oszacowania MSE [32]. W rozdziale 5 pracy doktorskiej pokazane jest, że wiele z tych własności przenosi się na nieodwracalne łańcuchy Markowa posiadające lukę spektralną w $L^2(\pi)$. W szczególności wówczas zachodzi CTG dla wszystkich funkcji całkowalnych w kwadracie względem π [18], dokładnie tak samo jak dla łańcuchów odwracalnych. W tym rozdziale przedstawiony jest nowy dowód tego faktu za pomocą równania Poissona, pozwalający na oszacowanie asymptotycznej wariancji przez

$$\sigma^2_{\mathrm{as}}(f) \leqslant K\left[\pi(f^2) - (\pi f)^2\right],$$

przy czym stała K zależy tylko od luki spektralnej. Jest to również uogólnieniem znanego faktu dla łańcuchów odwracalnych.

Twierdzenie 9. *Jeśli łańcuch Markowa $(X_n)_{n\geqslant 0}$ z jądrem przejścia $P(x, dy)$ i rozkładem stacjonarnym π ma lukę spektralną $1 - \rho$ w $L^2(\pi)$, to dla każdej funkcji f takiej, że $\pi(f^2) < \infty$ zachodzi CTG:*

$$\frac{\bar{S}_n}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2_{\mathrm{as}}(f)), \quad \textit{gdy } n \to \infty,$$

gdzie $\bar{S}_n := \sum_{i=0}^{n-1}\bar{f}(X_i)$ oraz $\bar{f} = f - \pi f$. Ponadto

$$\sigma^2_{\mathrm{as}}(f) = \pi(\bar{f}^2) + 2\mathbb{E}_\pi\sum_{n=1}^{\infty}\bar{f}(X_0)\bar{f}(X_n)$$

Niestety w przypadku nieodwracalnym istnienie luki spektralnej nie jest równoważne geometrycznej ergodyczności, co jest prawdą w przypadku odwracalnym [18]. W rozdziale 5 pracy doktorskiej przedstawiony jest prosty przykład geometrycznie ergodycznego łańcucha bez luki spektralnej. Dodatkowo ten łańcuch jest nowym prostszym przykładem geometrycznie ergodycznego łańcucha Markowa, dla którego nie jest spełnione CTG dla pewnej funkcji z $L^2(\pi)$.

W rozdziale 5 wprowadzone są dwa nowe warunki dryfu. Jeden z nich gwarantuje zachodzenie (5), zaś drugi pociąga za sobą słabszy warunek istnienia luki spektralnej w $L^2(\pi)$.

Literatura

[1] D. Aldous, *On the Markov Chain Simulation Method for Uniform Combinatorial Distributions and Simulated Annealing,* Probability in the Engineering and Informational Sciences 1, 33-46, 1987.

[2] K.B. Athreya and P. Ney, *A new approach to the limit theory of recurrent Markov chains*, Trans. Amer. Math. Soc. 245, 493–501, 1978.

[3] P.H. Baxendale, *Renewal Theory and Computable Convergence Rates for Geometrically Ergodic Markov Chains,* Ann. Appl. Prob. 15, 700–738, 2005.

[4] W. Bednorz, R. Latała and K. Łatuszyński, *A Regeneration Proof of the Central Limit Theorem for Uniformly Ergodic Markov Chains,* Elect. Comm. in Probab. 13, 85–98, 2008.

[5] R.C. Bradley, Jr. *Information regularity and the central limit question.* Rocky Mountain J. Math., 13(1), 77–97, 1983.

[6] G. Casella, C. P. Robert, *Monte Carlo Statistical Methods.* Springer-Verlag, New York, 1999.

[7] R. Douc, E. Moulines, J. S. Rosenthal, *Quantitative bounds on convergence of time-inhomogeneous Markov Chains.* Ann. Appl. Prob. 14, 1643-1665, 2003.

[8] C. J. Geyer, *Practical Markov Chain Monte Carlo.* Stat. Sci. 7 (4), 473-511, 1992.

[9] D. Gillman, *A Chernoff Bound for Random Walks on Expander Graphs.* SIAM J. Comput. 27 (4), 1203-1220, 1998.

[10] P. W. Glynn, D. Ormoneit, *Hoeffding's inequality for uniformly ergodic Markov chains.* Statist. Probab. Lett., 56(2):143–146, 2002.

[11] O. Häggström, *On the Central Limit Theorem for Geometrically Ergodic Markov Chains*, Probability Theory and Related Fields, 132, 74–82, 2005.

[12] W. Hoeffding, *Probability inequalities for sums for bounded random variables.* J. Amer. Statist. Assoc. 58 13-30, 1963.

[13] G.L. Jones, J.P. Hobert, *Sufficient Burn-in for Gibbs Samplers for a Hierarchical Random Effects Model.* The Annals of Statistics 32 (2), 784-817, 2004.

[14] G. L. Jones, M. Haran, B. S. Caffo, R. Neath, *Fixed-Width Output Analysis for Markov Chain Monte Carlo,* Journal of the American Statatistical Association, 101, 1537-1547, 2006.

[15] C. Kipnis, S.R.S. Varadhan, *Central Limit Theorem for Additive Functionals of Reversible Markov Processes and Applications to Simple Exclusions,* Commun. Math. Phys. 104, 1–19, 1986.

[16] I. Kontoyiannis, L. Lastras-Montano, S. P. Meyn, *Relative Entropy and Exponential Deviation Bounds for General Markov Chains,* 2005 IEEE International Symposium on Information Theory, 2005.

[17] I. Kontoyiannis, S.P. Meyn *Large deviation asymptotics and the spectral theory of multiplicatively regular Markov processes,* Electron. J. Probab., 10(3):61-123, 2005.

[18] I. Kontoyiannis, S.P. Meyn *Geometric ergodicity and the spectral gap of non-reversible Markov chains.*, preprint, 2009.

[19] K. Łatuszyński, *Regeneration and Fixed-Width Analysis of Markov Chain Monte Carlo Algorithms.* PhD Dissertation, 2008. Available at arXiv:0907.4716v1

[20] K. Łatuszyński, W. Niemiro, $(\varepsilon - \alpha)$-*MCMC approximation under drift condition*, in: Proceedings of the 6th International Workshop on Rare Event Simulation (RESIM 2006), 2006.

[21] K. Łatuszyński, W. Niemiro, *Rigorous confidence bounds for MCMC under a geometric drift condition*, 2010, ukaże się w J. Complex. arXiv:0908.2098

[22] K. Łatuszyński, B. Miasojedow, W. Niemiro, *Nonasymptotic bounds on the estimation error for regenerative MCMC algorithms*, dostępna w arXiv:0907.4915v1 .

[23] C. A. León, F. Perron, *Optimal Hoeffding bounds for discrete reversible Markov chains.* Ann. Appl. Probab. Volume 14, Number 2, 958-970, 2004.

[24] P. Lézaud, *Chernoff-type bound for finite Markov chains.* Ann. Appl. Probab. 8 849-867, 1998.

[25] J. S. Liu, *Monte Carlo Strategies in Scientific Computing.* Springer, 2001.

[26] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *Equations of state calculations by fast computing machines.* J. Chem. Phys. 21, 1087-1091, 1953.

[27] S.P. Meyn, R.L. Tweedie: *Markov Chains and Stochastic Stability.* Springer-Verlag, 1993.

[28] W. Niemiro, P. Pokarowski, *Fixed Precision MCMC Estimation by Median of Products of Averages.* Journal of Applied Probability, 46(2), 309-329, 2009.

[29] E. Nummelin, *A splitting technique for Harris recurrent Markov chains,* Z. Wahr. Verw. Geb. 43, 309–318, 1978.

[30] G.O. Roberts, J.S. Rosenthal, *Geometric ergodicity and hybrid Markov chains,* Elec. Comm. Prob. 2 (2), 1997.

[31] G.O. Roberts, J.S. Rosenthal, *General state space Markov chains and MCMC algorithms,* Probability Surveys 1, 20–71, 2004.

[32] D. Rudolf, *Explicit error bounds for lazy reversible Markov chain Monte Carlo,* J. of Complexity. 25, 11–24, 2008.

[33] G. O. Roberts, R. L.Tweedie, *Bounds on Regeneration Times and Convergence Rates for Markov Chains,* Stochastic Process. Appl. 91, 337-338, 1999.

[34] J. S. Rosenthal, *Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo,* Journal of the American Statistical Association, 90, 558-566, 1995.