Autoreferat pracy doktorskiej:

Regeneracja i estymacja stałoprecyzyjna dla algorytmów Monte Carlo opartych na łańcuchach Markowa

Krzysztof Łatuszyński

Jedną z najważniejszych dziedzin, w których używa się algorytmów Monte Carlo opartych na łańcuchach Markowa (MCMC) jest statystyka bayesowska, która jako skuteczne narzędzie w budowie złożonych modeli dobrze opisujących rzeczywistość odgrywa coraz większą rolę w zastosowaniach statystyki. Niestety podejście bayesowskie prowadzi do skomplikowanych rozkładów a posteriori, zazwyczaj znanych jedynie z dokładnością do stałej normującej, oraz do całkowania względem tychże rozkładów.

Dlatego, aby wykorzystać główne atuty statystyki bayesowskiej, niezbędne jest stosowanie algorytmów Monte Carlo opartych na łańcuchach Markowa, które pozwalają na przybliżone wyznaczenie całki postaci

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

gdzie π jest rozkładem a posteriori, znanym jedynie z dokładnością do stałej normującej, $\mathcal{X}$ jest podzbiorem przestrzeni wielowymiarowej, a f jest funkcją o wartościach rzeczywistych. Na przykład estymator bayesowski dla kwadratowej funkcji straty jest wartością oczekiwaną względem rozkładu a posteriori, która jest całką powyższej postaci z funkcją $f(x) = x$. Badanie jakości takich procedur jest kluczowe dla oceny jakości i przydatności estymatorów uzyskanych w modelu bayesowskim.

Procedura Monte Carlo do wyznaczania przybliżonej wartości całki w (1) polega na symulacji ergodycznego łańcucha Markowa $(X_n)_{n\geqslant 0}$ o operatorze przejścia P i rozkładzie stacjonarnym π, w którym X_n zbiega według rozkładu do π, oraz na wybraniu postaci estymatora $\hat{I}$ nieznanej wartości I. Pierwszym algorytmem generowania takiego łańcucha Markowa, w sytuacji, gdy stała normująca nie jest znana, był algorytm Metropolisa, zaproponowany w pracy [Metropolis et al. 1953], a od tego czasu pojawiło się wiele innych tego tupu procedur i wiele strategii estymacji (por. [Casella & Robert 1999], [Liu, JS 2001]).

Istniejąca literatura dotycząca łańcuchów Markowa i ich zastosowań do metod Monte Carlo w dużym stopniu koncentruje się na oszacowaniu szybkości zbieżności łańcucha Markowa do rozkładu stacjonarnego ([Baxendale 2005], [Douc et al. 2003], [Jones & Hobert 2004], [Roberts & Tweedie 1999], [Rosenthal 1995b]) oraz asymptotycznych własnościach estymatorów uzyskanych metodą MCMC

([Jones et al. 2006], [Kipnis & Varadhan 1986], [Meyn & Tweedie 1993]). Niestety wyniki dotyczące tempa zbieżności do rozkładu stacjonarnego pozwalają jedynie kontrolować obciążenie estymatorów i nie przekładają się na oszacowania błędu średniokwadratowego, czy konstrukcję przedziałów ufności. Z kolei wyniki asymptotyczne mogą okazać się w praktyce bezużyteczne, albo nawet wprowadzające w błąd ([Roberts & Rosenthal 2005]).

Głównym celem pracy jest uzyskanie dla estymatora $\hat{I}$ opartego na algorytmie MCMC wyników dotyczących estymacji stałoprecyzyjnej, w szczególności uzyskanie tzw. $(\varepsilon - \alpha)-$aproksymacji, czyli

$$P(|\hat{I} - I| \geqslant \varepsilon) \leqslant \alpha. \tag{2}$$

W analizie łańcuchów Markowa i estymatorów uzyskanych metodą MCMC zastosowane zostało podejście regeneracyjne, bazujące na łańcuchu rozszczepionym. Rozszczepianie łańcuchów Markowa na ogólnej przestrzeni stanów umożliwia podział trajektorii na niezależne lub 1-zależne kawałki i okazuje się niezwykle skuteczną techniką o szerokich zastosowaniach. Metoda ta została wprowadzona niezależnie w pracach [Athreya & Ney 1978] oraz [Nummelin 1978] i ogromnie rozwinięta w monografiach [Nummelin 1984] i [Meyn & Tweedie 1993].

Wyniki typu (2) znane są w literaturze dla dyskretnej przestrzeni stanów $\mathcal{X}$ i ograniczonej funkcji f ([Aldous 1987], [Gillman 1998], [León & Perron 2004]). Dla ciągłej przestrzeni stanów $\mathcal{X}$, łańcuchów jednostajnie ergodycznych (co w praktyce jest równoważne zwartości $\mathcal{X}$) oraz ograniczonej funkcji f znane są nierówności wykładnicze ([Glynn & Ormoneit 2002], [Kontoyiannis at al. 2005]), z których $(\varepsilon - \alpha)-$aproksymacja łatwo wynika.

Dla przestrzeni $\mathcal{X}$ ciągłej i niezwartej oraz funkcji f nieograniczonej (z taką sytuacją mamy do czynienia np. dla wspomnianego wcześniej estymatora bayesowskiego dla kwadratowej funkcji straty) brak jest tego typu nieasymptotycznych rezultatów i estymację stałoprecyzyjną realizuje się przez konstrukcję asymptotycznych przedziałów ufności dla estymatora postaci

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

Konstrukcja ta wymaga dwóch kroków. Po pierwsze, musi być spełnione centralne twierdzenie graniczne dla $\hat{I}_n$, czyli twierdzenie postaci

$$\frac{\hat{I}_n - I}{\sqrt{n}} \xrightarrow{d} N(0, \sigma_f^2), \tag{3}$$

gdzie $\sigma_f^2 < \infty$ nazywana jest wariancją asymptotyczną. Po drugie, niezbędny jest mocno zgodny estymator $\hat{\sigma}_f^2$ wariancji asymptotycznej σ_f^2. Praca [Jones et al. 2006] przedstawia aktualny stan wiedzy na temat tej metodologii.

Wyniki rozdziałów 3 i 4 w pracy doktorskiej dotyczą tego właśnie podejścia.

W rozdziale 3, opartym na pracy [Bednorz, Latała & Łatuszyński 2008], dla łańcuchów ergodycznych uzyskany został w terminach regeneracji warunek konieczny i dostateczny dla centralnego twierdzenia granicznego w sensie (3). Okazuje się, że centralne twierdzenie graniczne jest spełnione wtedy i tylko wtedy gdy skończone są wartości oczekiwane kwadratów wycieczek pomiędzy kolejnymi regeneracjami. Dodatkowym wynikiem rozdziału 3 jest rozwiązanie problemu otwartego postawionego w pracy [Roberts & Rosenthal 2005], mianowicie regeneracyjny dowód centralnego twierdzenia granicznego dla jednostajnie ergodycznych łańcuchów Markowa przy warunku $E_\pi f^2 < \infty$.

Rozdział 4, oparty na pracy [Bednorz & Łatuszyński 2007], poświęcony jest założeniom potrzebnym do konstrukcji mocno zgodnych estymatorów wariancji asymptotycznej σ_f^2. Osłabione są założenia przyjęte do tej konstrukcji w pracy [Jones et al. 2006].

Najobszerniejszy rozdział 5 poświęcony jest uzyskaniu nieasymptotycznych wyników typu (2) dla przypadku niezwartej przestrzeni stanów $\mathcal{X}$ oraz nieograniczonej funkcji f.

Dokładniej, celem tego rozdziału jest analiza estymatora

$$\hat{I}_{t,n} = \frac{1}{n} \sum_{i=t}^{t+n-1} f(X_i) \tag{4}$$

nieznanej wartości I przy założeniu dryfu w kierunku małego zbioru następującej postaci.

(A.1) Mały zbiór. Istnieje $C \subseteq \mathcal{X}$, $\tilde{\beta} > 0$ oraz miara probabilistyczna ν na $\mathcal{X}$, taka, że dla wszystkich $x \in C$ oraz $A \subseteq \mathcal{X}$

$$P(x, A) \geqslant \tilde{\beta}\nu(A).$$

(A.2) Dryf. Istnieje funkcja (Lapunowa) $V : \mathcal{X} \to [1, \infty)$ oraz stałe $\lambda < 1$ i $K < \infty$ spełniające warunek

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

(A.3) Nieokresowość. Istnieje stała $\beta > 0$ taka, że $\tilde{\beta}\nu(C) \geqslant \beta$.

Przy tych założeniach uzyskane są jawne dolne oszacowania na t oraz n, które gwarantują $(\varepsilon - \alpha)-$aproksymację. Oszacowania te zależą w sposób jawny od parametrów estymacji ε i α, parametrów dryfu $\tilde{\beta}, \beta, \lambda, K$ oraz normy $|f^2|_V = \sup_x f^2(x)/V(x)$.

Ponadto, oprócz estymatora danego przez (4) przeanalizowany został estymator gwarantujący $(\varepsilon - \alpha)-$aproksymację oparty na medianie kilku estymatorów postaci (4) o krótszej trajektorii. Dla małych wartości α oszacowania całkowitego kosztu symulacji dla estymatora medianowego okazały się lepsze.

Wyniki teoretyczne rozdziału 5 zostały zastosowane do próbników Gibbsa dla hierarchicznego modelu komponentów wariancyjnych dając $(\varepsilon - \alpha)-$aproksymację. Jest to model o bardzo rozległych zastosowaniach w różnych gałęziach statystyki. Dotychczasowa analiza procedur MCMC dla modelu komponentów wariancyjnych ograniczała się do górnych oszacowań na czas *burn-in*, czyli t w estymatorze (4), uzyskanych w pracy [Jones & Hobert 2004].

Ostatni, 6 rozdział porusza nieco odmienną tematykę, mianowicie metody adaptacyjne. Idea polega na modyfikowaniu operatora przejścia procesu w zależności od pewnych informacji zgromadzonych na podstawie dotychczasowego przebiegu procedury. Z reguły prowadzi to do procesów stochastycznych, które nie mają własności Markowa i są przez to znacznie trudniejsze do badania teoretycznego. Z drugiej strony, w procedurach tych wykorzystuje się dodatkową informację o przebiegu symulacji i jednocześnie rozszerza się dostępną klasę procesów stochastycznych, na których oparta jest estymacja. Pozwala to oczekiwać znacznego polepszenie własności estymatorów przy umiejętnym projektowaniu adaptacji. Symulacje potwierdzają te oczekiwania, przykłady numeryczne dla wielu konkretnych algorytmów wskazują na istotną poprawę efektywności estymacji [Roberts & Rosenthal 2006], [Kohn & Nott 2005]. Szczególnie ważnym przykładem zastosowania procedur adaptacyjnych jest algorytm Metropolisa z wielowymiarowym rozkładem normalnym jako rozkładem propozycji, w którym adaptacja pozwala na automatyzację doboru macierzy wariancji i kowariancji dla rozkładu propozycji [Atchadé & Rosenthal 2005]. Wiedza teoretyczna na temat własności algorytmów adaptacyjnych jest jak na razie bardzo ograniczona i nawet wyniki dotyczące zbieżności i twierdzeń asymptotycznych znane są tylko w bardzo szczególnych przypadkach i zakładają między innymi stabilizację operatorów przejścia w czasie ([Atchadé & Rosenthal 2005], [Kohn & Nott 2005]), co intuicyjnie odpowiada zbliżaniu się procesu do jednorodnego ergodycznego łańcucha Markowa.

W rozdziale 6 zostały udowodnione dwa twierdzenia, jedno w postaci oszacowania tempa zbieżności, a drugie w postaci prawa wielkich liczb dla procedury adaptacyjnej. W obu twierdzeniach dla operatorów przejścia zakłada się warunek *stabilizacji względem trajektorii* słabszy od warunku *stabilizacji w czasie*, przyjmowanego w pracy [Atchadé & Rosenthal 2005] do dowodu analogicznych twierdzeń, i wynikającego z warunku *stabilizacji w czasie* poprzez nierówność trójkąta. Warunek *stabilizacji względem trajektorii* intuicyjnie oznacza, że proces adaptacyjny może się zbliżać również do niejednorodnego ergodycznego łańcucha Markowa i ta niejednorodność nie popsuje zbieżności procedury adaptacyjnej.

Literatura

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

[Atchadé & Rosenthal 2005] Atchadé Y. F., Rosenthal J. S., 2005. *On Adaptive Markov Chain Monte Carlo Algorithms.* Bernoulli 11, 815–828.

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

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

[Bednorz, Latała & Łatuszyński 2008] Bednorz W., Latała R., Łatuszyński K. (2008). *A Regeneration Proof of the Central Limit Theorem for Uniformly Ergodic Markov Chains.* Electronic Communications in Probability, 13, 85–98.

[Bednorz & Łatuszyński 2007] Bednorz, W., Łatuszyński, K., (2007), *A few Remarks on "Fixed-Width Output Analysis for Markov Chain Monte Carlo" by Jones et al.* Journal of the American Statatistical Association 102 (480), 1485-1486.

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

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

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

[Glynn & Ormoneit 2002] Glynn P. W., Ormoneit D. 2002 *Hoeffding's Inequality for Uniformly Ergodic Markov Chains.* Statist. and Probab. Lett. 56, 143-146.

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

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

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

[León & Perron 2004] León C. A., Perron F., 2004. *Optimal Chernoff Bounds for Finite Reversible Markov Chains.* Ann. Appl. Prob. 14, 958-970.

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

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

[Kohn & Nott 2005] Kohn R., Nott D., 2005. *Adaptive Sampling for Bayesian Variable Selection.* To appear, Biometrika.

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

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

[Nummelin 1978] Nummelin E. (1978). A splitting technique for Harris recurrent chains. *Z. Wahrscheinlichkeitstheorie und Verw. Geb.* **43** 309–318.

[Nummelin 1984] Nummelin E. (1984). *General Irreducible Markov Chains and Nonnegative Operators.* Cambridge University Press, Cambridge.

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

[Roberts & Rosenthal 2005] Roberts G. O., Rosenthal J. S., 2005. *General state space Markov chains and MCMC algorithms.* Probability Surveys 1:20-71.

[Roberts & Rosenthal 2006] Roberts G.O., Rosenthal J.S., *E*xamples of Adaptive MCMC, Preprint (2006).

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