On the efficiency of pseudo-marginal Markov chain Monte Carlo algorithms

Christophe Andrieu[1], Lasse Leskelä[2] and Matti Vihola[3]
[1] University of Bristol, UK
[2] Aalto University, Finland
[3] University of Jyväskylä, Finland

Pseudo-marginal Markov chain Monte Carlo (MCMC) (Andrieu & Roberts, 2009) is a generic emerging class of algorithms for challenging Bayesian inference. The methods are generally applicable when the density of interest $\pi(x)$ cannot be evaluated, or it is computationally expensive to evaluate, but non-negative and unbiased estimates T_x of $\pi(x)$ are available. Surprisingly, a simple method relying on a bivariate Markov chain, including the density estimates as auxiliary variables, turns out to be exact in the sense that it is reversible with respect to an invariant distribution with the correct marginal distribution $\pi(x)$. This guarantees that the estimates produced by the algorithm converge almost surely to the correct value, with minimal irreducibility assumption.

First methods of the pseudo-marginal type were those involving importance sampling estimates, termed Grouped Independence Metropolis-Hastings by Beaumont (2003), followed soon by Approximate Bayesian Computation (ABC) MCMC algorithms due to Marjoram et al. (2003). The more recent important instances include the algorithm for inference of discretely observed diffusions (Beskos et al., 2006) and the so-called particle marginal Metropolis-Hastings algorithm (Andrieu et al., 2010). The latter involves likelihood estimates of sequential Monte Carlo algorithms. All the above mentioned methods have attracted substantial interest.

Usually in practice, there is some freedom how to implement the estimates T_x of $\pi(x)$. The talk focuses on providing informative criteria which kind of estimators should be preferred. More specifically, suppose $T_x^{(1)}$ and $T_x^{(2)}$ are two non-negative unbiased estimators of π. We discovered that if $T_x^{(1)}$ is less than $T_x^{(2)}$ in the convex stochastic order (cf. Shaked & Shanthikumar, 2007), then the corresponding pseudo-marginal algorithms are guaranteed to be ordered in terms of efficiency measures such as the asymptotic variance (Andrieu & Vihola, 2014).

The result has similarities with that discovered by Peskun (1973) and later extended, for example, by Tierney (1998). There is, however, a fundamental difference in that the Markov chains that are compared are reversible with respect to different invariant measures. A key ingredient of the proof involves on the existence of conditional martingale couplings (Leskelä & Vihola, 2014), which generalise the well-known martingale characterisation of convex orders due to Strassen (1965).

The literature on convex orders is rich, and allows for example finding extremal behaviour incorporating estimators which are 'best' or 'worst' in

the convex order given certain constraints. There is also an immediate consequence to the implementation of certain ABC MCMC algorithms, which states that under general conditions, a stratified estimator provides more efficient algorithm than an estimator involving indepdendent samples.

Keywords: Asymptotic variance, Convex order, Efficiency, Pseudo-marginal algorithm

References:

Andrieu, C., A. Doucet, R. Holenstein (2010). Particle Markov chain Monte Carlo methods (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol. 72* 269–342.

Andrieu, C., G.O. Roberts (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist. 37*, 697–725.

Andrieu C., M. Vihola (2014). Establishing some order amongst exact approximations of MCMCs. In preparation.

Beaumont, M.A. (2003). Estimation of population growth or decline in genetically monitored populations. *Genetics 164*, 1139–1160.

Beskos, A., O. Papaspiliopoulos, G.O. Roberts, P. Fearnhead (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *J. R. Stat. Soc. Ser. B Stat. Methodol. 68*, 333–382.

Leskelä, L., M. Vihola (2014). Conditional convex orders and measurable martingale couplings. Preprint *arXiv:1404.0999*

Marjoram, P., J. Molitor, V. Plagnol, S. Tavaré (2003), Markov chain Monte Carlo without likelihoods. *Proc. Nat. Acad. Sci. 100*, 15324-15328.

Peskun, P. (1973). Optimum Monte-Carlo sampling using Markov chains. *Biometrika 60*, 607–612.

Shaked, M., J.G. Shanthikumar (2007). *Stochastic orders*. Springer.

Strassen., V. (1965). The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 423–439.

Tierney, L. (1998). A note on Metropolis-Hastings kernels for general state spaces. *Ann. Appl. Probab. 8*, 1–9.