

ANNOTATED BIBLIOGRAPHY

IDLE HANDS FOR MONTE CARLO RESIDENTS

Radu V. Craiu

craiu@utstat.utoronto.ca

When travelling through Transylvania's villages, as I did this summer, one feels a certain pang of nostalgia for those simpler times in which one could build by hand an entire way of life. However, once we reach the realm of the Markov chain Monte Carlo universe, the opposite desire prevails: we would *really* like to have the computer take over the daunting task of finding good simulation parameters. Nevertheless, such surrender must be done carefully as pitfalls await the unsuspecting users, be they theorists or practitioners. And one should keep in mind that while adaptation of a sound MCMC algorithm can save time and trouble, it is also true that automation applied to an inefficient algorithm will likely magnify the inefficiency.

Please note that this annotated bibliography mostly covers the period from 2008 until now as I have tried to minimize the overlap with the previous review of Andrieu (2008). In addition, I have not included here the rich literature on adaptive importance sampling (see Cornuet et al, 2011, for example) and stochastic approximation Monte Carlo sampling (see Liang, 2009 and references therein).

Adaptive MCMC

- Atchadè YF, Fort G, Moulines E, Priouret P (2009) Adaptive Markov chain Monte Carlo: Theory and Methods. *Technical Report, University of Michigan*. A review paper in which the authors make the distinction between *internal adaptation* which relies on the past history of the chain and *external adaptation* which relies on auxiliary processes, be they adaptive or not, that are run in parallel with the chain of interest. This creates a natural connection with the world of interacting MCMC.
- Atchadè YF, Fort G (2010) Limit Theorems for some adaptive MCMC algorithms with subgeometric kernels. *Bernoulli*, **16**:

116–154. The authors investigate adaptive MCMC algorithms derived using transition kernels that are not geometrically ergodic. They introduce conditions that guarantee the ergodicity of the chain and prove a strong law of large numbers in this setting. An application for their results is the Adaptive Metropolis (AM) algorithm of Haario et al. (2001) when the target distribution is sub-exponential in the tails.

- Atchadè YF, Fort G (2011) Limit theorems for some adaptive MCMC algorithms with subgeometric kernels: PART II. *Preprint*. The paper contains a proof of a central limit theorem in the case of AMCMC driven by sub-geometrically ergodic transition kernels. The results are applied to a number of existing chains used in the literature.
- Atchadè YF (2010) A cautionary tale on the efficiency of some adaptive Monte Carlo schemes. *The Annals of Applied Probability*, **20**: 841-868. An adaptive algorithm does not necessarily inherit the same good convergence properties that characterize the limiting transition kernel. The article shows that this is the case, in terms of asymptotic variance, for the equi-energy sampler.
- Atchadè YF (2011) Kernel estimators of asymptotic variance for adaptive Markov chain Monte Carlo. *The Annals of Statistics*, **39**: 990-1011. Studies the asymptotic properties of kernel estimators of asymptotic variances for some adaptive MCMC algorithms.
- Bai Y, Roberts GO, Rosenthal JS (2011) On the containment condition for adaptive Markov chain Monte Carlo algorithms. *Advances and Applications in Statistics*, **21**: 1-54. The ergodicity of adaptive MCMC has been framed by Roberts and Rosenthal (2007) in terms of two general conditions: Diminishing Adaptation (DA) and the Containment Condition (CC). The latter is arguably the most difficult one to prove in general adaptive samplers. The authors propose various sufficient conditions for CC

and connect the convergence rates of algorithms with the tail properties of the target distribution. Examples are given to show that Diminishing Adaptation does not ensure ergodicity.

- Bai Y (2009) Convergence of adaptive Markov chain Monte Carlo algorithms. *PhD Dissertation, University of Toronto*. Shows that neither Diminishing Adaptation nor Containment are necessary for ergodicity. Discusses additional conditions that make Containment necessary and find sufficient conditions that ensure Containment.
- Bai Y, Craiu RV, Di Narzo AF (2011) Divide and Conquer: A mixture-based approach to regional adaptation for MCMC. *J. Comput. Graph. Statist.*, **20**:63–79. When the target distribution can be reasonably approximated by a mixture of Gaussians, the authors propose to use an adaptive regional RWM algorithm. The approximating mixture parameters are estimated with an online EM algorithm and the sample space is adaptively partitioned so that in each subregion exactly one component of the mixture is used as the proposal distribution. It extends the work in Craiu et al. (2009) as the regions are evolving along with the mixture parameters.
- Craiu RV, Rosenthal JS, Yang C (2009) Learn from the neighbor: Parallel-Chain Adaptive and Regional MCMC. *Journal of the American Statistical Association*, **104**:1454-1466. The use of AMCMC for sampling from multimodal distributions can be tricky if the chain gets stuck and thus learns/adapts only using draws from a subregion of the sample space. The authors propose using parallel adaptive chains and a tempering-based strategy for learning good starting values of the adaptation parameters. In the case of multimodal targets it is possible that across regions of the sample space one should use different transition kernels. Given a partition of the sample space that approximately reflects the location of the modes, the authors develop an adaptive algorithm which uses a different mixture of Gaussians as the proposal distribution in each region. The parameters of the mixture are adapted as the simulation proceeds. Although ideas are illustrated using the random walk Metropolis (RWM), they applicable to many other MCMC algorithms.
- Ji C, Schmidler SC (2011) Adaptive Markov chain Monte Carlo for Bayesian variable selection *J. Comput. Graph. Statist.*, to appear. Develops AMCMC algorithms for Bayesian variable selection problems in regression models where the mixture posterior distributions have point mass components corresponding to posterior inclusion probabilities .
- Fort G, Moulines E, Priouret P (2010) Convergence of adaptive MCMC algorithms: ergodicity and law of large numbers. *Preprint*. Develops a general theory for handling both internal and external adaptive MCMC schemes (see Atchadè et al., 2009), e.g. the equi-energy sampler or the AM algorithm.
- Fort G, Moulines E, Priouret P, Vandekerckhove P (2011) A central limit theorem for adaptive and interacting Markov chains. *Preprint*. Establish central limit theorems for additive functionals of unbounded functions in the case of general simulation schemes that include both adaptive and interacting MCMC algorithms.
- Garthwaite PH, Fan Y, Sisson SA (2010) Adaptive optimal scaling of Metropolis-Hastings algorithms using the Robbins-Monro process. *Preprint* The authors propose an estimator for the steplength constant used in the Robbins-Monro stochastic search algorithm. The method is applied to RWM and RWM-within-Gibbs samplers.
- Laine M, Tamminen J (2008) Aerosol model selection and uncertainty modelling by adaptive MCMC technique. *Atmospheric Chemistry and Physics Discussions*, **8**:10791-10816. Develops an adaptive reversible jump MCMC algorithm to determine the aerosol model for the data observed.
- Latuszyński K, Roberts GO, Rosenthal JS (2010) Adaptive Gibbs samplers and related MCMC methods. *Annals of Applied Probability*, to appear. Propose adaptive Gibbs sampling algorithms and give theoretical results to prove their validity. The authors also propose an adaptive Metropolis-within-Gibbs algorithm in which the coor-

dinate selection probabilities and the proposal distributions used in the Metropolis steps are updated on the fly.

- Loredo TJ, Berger JO, Chernoff DF, Clyde MA, Liu B (2011) Bayesian Methods for Analysis and Adaptive Scheduling of Exoplanet Observations. *Preprint*. Collaborative effort between astronomers and statisticians to develop Bayesian data analysis tools for planet detection, planetary orbit estimation, and adaptive scheduling of observations. The algorithm designed to sample from the posterior distribution used AMCMC principles.
- Richardson S, Bottolo L, Rosenthal JS (2011) Bayesian models for sparse regression analysis of high dimensional data. *Proceedings of Valencia IX Bayesian Meeting*. Applies adaptive MCMC to Bayesian inference for sparse regression models.
- Schmidler SC, Woodard, DB (2011) Lower bounds on the convergence rates of adaptive MCMC methods. *Preprint*. Develops lower bounds for the mixing time of some well-known adaptive MCMC algorithms including the AM of Haario et al. (2001), the equi-energy sampler of Kou et al. (2006) and the inter-chain adaptation strategy of Craiu et al (2009). This leads to an interesting discussion regarding the gains of adapting and suggests promising directions for additional speed-ups.
- Saksman E, Vihola M (2010) On the ergodicity of the adaptive Metropolis algorithm on unbounded domains. *The Annals of Applied Probability*, 20:2178-2203. Describes sufficient conditions under which the AM algorithm preserves ergodicity without assuming boundedness of the target's support. The approach avoids the use of rejections (see Andrieu and Moulines, 2006).
- Salakhutdinov R (2010) Learning deep Boltzmann machines using adaptive MCMC. *Proceedings of the 27th International conference on machine learning*. Proposes an adaptive simulated tempering algorithm to explore the multimodal energy landscape of the distribution defined by the Deep Boltzmann Machine.
- Vihola M (2010) Robust Adaptive Metropolis algorithm with coerced acceptance rate. *Preprint*. Proposes an adaptive approach designed for multivariate RWM in which the shape of the covariance matrix is adapted while the mean acceptance rate is maintained at a pre-determined level. The method proposed avoids using the sample covariance matrix and is thus more robust.
- Vihola M (2009) On the Stability and Ergodicity of Adaptive Scaling Metropolis Algorithms. *Preprint*. Proves that the ergodicity of the algorithms proposed by Haario et al. (2001) and Vihola (2010) can be proven without restricting the scale parameter to a compact interval.
- Vihola M (2011) GRAPHAM: Graphical models with adaptive random walk Metropolis algorithms. *Preprint*. The paper accompanies an open-source implementation in C of a number of adaptive algorithms including the AM or AM-within-Gibbs.
- Vihola M (2011) Can the adaptive Metropolis algorithm collapse without the covariance lower bound? *Electronic Journal of Probability*, 16:45-75. Discusses variants of the AM algorithm where the eigenvalues of the covariance matrix used in the RWM proposal are not explicitly bound away from zero. Also proves a law of large numbers for super-exponentially decaying distributions with regular contours.
- Vrugt A, ter Braak CJF, Diks CGH, Robinson BA, Hyman JM, Higdon D (2009) Accelerating Markov Chain Monte Carlo Simulation by Differential Evolution with Self-Adaptive Randomized Subspace Sampling. *International Journal of Nonlinear Sciences & Numerical Simulation*, 10:273-290. Uses a self-adaptive Differential Evolution learning strategy within a population-based evolutionary framework. The authors propose running multiple different chains simultaneously for global exploration, and automatically tuning the scale and orientation of the proposal distribution in randomized subspaces during the search.

References

- Andrieu C (2008) Adaptive Monte Carlo methods. *ISBA Bulletin* 15:2-6.

- Andrieu C, Moulines E (2006) On the ergodicity properties of some adaptive MCMC algorithms. *Ann. Appl. Probab.* **16**:1462–1505.
- Cornuet JM, Marin JM, Mira A, Robert CP (2011) Adaptive Multiple Importance Sampling. *Scandinavian Journal of Statistics*, to appear.
- Haario H, Saksman E, Tamminen J (2001) An adaptive Metropolis algorithm. *Bernoulli* **7**:223–242.
- Kou S, Qing Z, and Wong W (2006) Equi-energy sampler with applications in statistical inference and statistical mechanics. *Ann. Statist.* **34**:1581–1619.
- Liang F (2009) On the use of stochastic approximation Monte Carlo for Monte Carlo integration. *Statistics and Probability Letters*, **79**:581-587.
- Roberts GO, Rosenthal JS (2007). Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms. *J. Appl. Probab.* **44**:458–475. ▲

STUDENTS' CORNER

Q & A

Luke Bornn

l.bornn@stat.ubc.ca

In this issue's Students' Corner, we continue our Q & A with a panel of leading Bayesian statisticians. Following the Q & A, find the dissertation abstract of recent Duke University graduates Matthew Heaton, Vinicius Mayrink, Chiranjit Mukherjee, and Hongxia Yang. If you are newly graduated and would like to publish your thesis abstract, don't hesitate to contact me.

What is the biggest and most surprising change in the field of Statistics that you have witnessed, and what do you think will be the next one?

Paul Gustafson

gustaf@stat.ubc.ca

In my own case I came of age during 'the big change' in Bayesian statistics. In 1991 I completed an M.Sc. thesis full of Bayesian inferences computed via Laplace approximation. In 1994 I completed a Ph.D. thesis in which all the computing was done via MCMC. This revolution in Bayesian computation has been well commented on. I also think it is fairly characterized as part of

the larger revolution in statistics as a whole becoming far more interdisciplinary, with the boundary between methodological developments and applications now fruitfully blurred.

What changes might be coming next? Well, both technology and the size and complexity of datasets keep on racing upward. The landscape changes quickly, and my crystal ball is cloudy on where this all might lead. In my own sphere, I am excited to encounter at least some health researchers who can get excited about methodological developments and 'full' modeling of uncertainty, even if their confidence intervals ... whoops, I mean credible intervals of course ... might then be wider at the end of the day. A change I hope for and predict is that science will increasingly strive for quantitative analysis that accounts for all the uncertainties at play, even if that rather often translates to an admission that we don't know very much. Pushes in this direction are already underway. For instance, the need to mitigate high false discovery rates in genome-wide association studies is catching on beyond just the statistical community. As we all preach to those around us, in statistical modeling, as in life, being forthright about what we don't know is of paramount importance. In clichéd terms, honesty is the best policy! And while I don't subscribe to the cliché about ignorance being bliss, I think we all agree that figuring out the extent of our ignorance after having observed the data is indeed blissful.