Some Issues in Monitoring Convergence of Iterative Simulations

Stephen Brooks and Andrew Gelman

Abstract:

In this paper, we discuss some recent results and open questions concerning monitoring convergence of iterative simulations. We begin by discussing the various approaches to convergence assessment proposed in the literature, grouping the methods according to their underlying principles. We then discuss how MCMC simulations can be constructed so that convergence monitoring is simplified. Finally, we discuss some new convergence assessment ideas that are the focus of current work.

Key Words: Markov chain Monte Carlo; Convergence Diagnosis; Inference.

1. Introduction

Iterative simulations, especially Markov chain Monte Carlo (MCMC) methods, have been increasingly popular in statistical computation, most notably for drawing simulations from Bayesian posterior distributions, see Gilks *et al* (1996) and Brooks (1998a) for example. In addition to any implementational difficulties and computing resources required, iterative simulation presents two problems beyond those of traditional statistical methods. First, when running an *iterative* algorithm, one must decide when to stop the iterations or, more precisely, one must judge how close the algorithm is to convergence after a finite number of iterations. Secondly, MCMC simulation converges to a target *distribution*, rather than a target point. This leads to many practical difficulties; see Kass *et al* (1997) for example.

1.1 Monitoring Convergence of Iterative Simulation in General, not just MCMC

Convergence problems apply quite generally to iterative simulation algorithms, not just to MCMC algorithms. For example Gelman (1992) discusses how importance sampling methods are in fact iterative and, in general, result in draws from the target distribution only in the limit as the number of iterations approaches infinity. One way of seeing this approximate nature of importance sampling is to note that ratio estimates of importance-weighted means, $\sum_{i=1}^{n} w_i h(\theta_i) / \sum_{i=1}^{n} w_i$ are unbiased only in the limit as $n \to \infty$, and that this convergence (as well as more practical issues of the variance of the estimate in a finite sample) depends upon the upper tail of the distribution of the weights w_i. Liu *et al* (1998) note the duality between this "importance weight infinity" and the "waiting time infinity" of MCMC and rejection sampling.

1.2 Categorisation of Methods for Monitoring Convergence

In this paper, we discuss some recent results and open questions in monitoring the convergence of iterative simulations. We begin by briefly categorising the techniques that can be used to monitor convergence.

1. Methods for assessing convergence without analysis of simulation output. This can be done in several ways, including:

   (a) Designing the simulation algorithm to produce independent draws from the target distribution. Examples include (i) rejection sampling using a proposal function that uniformly dominates the target density, (ii) coupling and regeneration methods in MCMC (Nummelin, 1978, Mykland *et al*, 1995 and Johnson, 1998) and (iii) the "perfect simulation" method of Propp and Wilson (1996), in cases where it is computationally feasible. In each of these approaches, the time required to wait until the next independent draw is a random variable, which can limit the effectiveness of these methods if the waiting time is too long.

   (b) Theoretical (analytic) results bounding the difference between the simulation and target distributions after some specified number of iterations. Reasonable results of this type have appeared only for some very simple models, see Rosenthal (1995) and Cowles and Rosenthal (1998), for example. Though the most recent work has

Stephen Brooks, Department of Mathematics, University Walk, Bristol, BS8 1TW, UK. (Steve.Brooks@bris.ac.uk). Andrew Gelman, Department of Statistics, Columbia University, New York, USA. (gelman@stat.columbia.edu). We thank the U.S. National Science Foundation for grants SBR-9708424 and Young Investigator Award DMS-9796129.

improved considerably the state of the art in this area (see for example Roberts and Tweedie, 1998a,b), it is still unrealistic to expect this approach to become widely applicable in MCMC simulation except in certain special cases (most notably the so-called *slice* sampler, see Roberts and Rosenthal, 1998). It is also worth noting that the time to convergence of these methods generally depends upon the starting points of the simulation.

2. Methods for detecting mixing of Markov chain samplers. Probably the most commonly-used convergence diagnostics make use of the fact that most MCMC algorithms have a random-walk behaviour in which a simulated chain gradually spreads out from its starting point to ergodically cover the space of the target distribution. Convergence occurs when the chain has fully spread to the target distribution, which can be judged in three basic ways:

   (a) Monitoring trends. Given a single MCMC sequence, one can judge mixing by looking for trends in the simulation (Yu, 1995, Yu and Mykland, 1998, and Brooks 1998b); unfortunately, such an approach will not necessarily detect lack of convergence of a slowly-moving sequence (Gelman and Rubin, 1992b).

   (b) Monitoring autocorrelation. Efficiency of simulations can be judged by autocorrelations, and this approach can also be used to obtain approximately independent simulation draws (Raftery and Lewis, 1992). This approach however can also be fooled by very slow-moving series and thus is perhaps most effective as a measure of efficiency for an MCMC algorithm for which convergence has already been judged by other means.

   (c) Monitoring mixing of sequences. Gelman and Rubin (1992a) proposed directly monitoring the mixing of simulated sequences by comparing the variance within each sequence to the total variance of the mixture of the sequences. This is an adaptation of statistical analysis of variance to the standard multiple-sequence approaches in statistical physics (see, e.g., Fosdick, 1959).

   Interestingly, the approaches based upon detecting a lack of mixing are ineffective in monitoring convergence of non-Markov-chain iterative simulation methods such as importance sampling, for which successive draws are not nearby in the parameter space. This is another argument in favour of the use of MCMC in preference to other iterative simulation methods. It is interesting that autocorrelation or locality of random-walk or state-space algorithms, which is generally perceived as a drawback (since it decreases the efficiency of simulations), is actually an advantage in convergence monitoring.

3. Methods based upon sequential testing of portions of simulation output in order to determine whether or not they could be considered to have been drawn from the same distribution. Methods of this sort sequentially discard an increasing proportion of the early simulated values and divide the remaining observations into three blocks. The observations in the first and third block are then compared and a formal procedure used to test the null hypothesis that the simulated observations are drawn from the same distribution. If the test is rejected then more of the early values are discarded and the testing procedure is repeated. If the test is accepted, then it is assumed that the discarded observations covered the burn-in period and that the remaining observations are all generated from the same (assumed to be the stationary) density. See Geweke (1992) and Heidelberger and Welch (1983), for example. Such approaches can be considered as a special case of methods for detecting trends, as discussed in item 2(a) above.

4. Methods based upon functions of the simulation output that are related to the simulation algorithm in a known way, such as importance ratios, acceptance probabilities, transition probabilities, and posterior densities. Importance ratios and acceptance probabilities have been useful in approximately evaluating the efficiency of importance sampling (Kong, 1992) and Metropolis algorithms (Gelman *et al*, 1996) once convergence has been reached, but they do not seem very powerful in detecting poor convergence if used alone. More effective approaches combine importance ratios with other information, as in the methods of Cui *et al* (1992), Roberts (1992), Liu *et al* (1993) and Brooks *et al* (1997).

5. Methods based upon estimating some aspect of the target distribution in more than one way, with the knowledge that the estimates should

be identical (within sampling variation) at convergence. One such method is to compare empirical joint density estimates to the target density function itself, which seems to work well in some moderate-dimensional problems (Ostland and Yu, 1997). Another approach is to compare empirical estimates of marginal densities to estimates obtained from path sampling (Gelman and Meng, 1998). We imagine that much more work could be done in this area, especially if combined with the methods discussed in item 4 above, and considering the inherent redundancy of the information available in the unnormalised density function and the simulations themselves (see also O'Hagan, 1987).

2. Designing the Simulations to Make Convergence Monitoring More Reliable

We can think of structuring the simulations as a problem in experimental design for sequential analysis, in which we wish to design our simulation procedure so as to achieve the following goals using the minimum amount of simulation time: (a) we wish to obtain inference for functionals of the target distribution (for example, accurate estimates for expectations $\mathbb{E}[h(\theta)]$), (b) we wish to obtain inference from the target distribution (for example, 95% intervals for functions $h(\theta)$ that should have approximately 95% coverage under the target distribution), and (c) we wish to reliably monitor convergence for the summaries of interest. In the simulation literature, there is much written about the effect of design upon the precision of estimates (a), some written about inferences (b), and very little about (c), which we shall consider here (see also the first paragraph of Gelfand, 1992).

Consider the following design factors that we can control:

1. Type of simulation algorithm (importance sampling, Gibbs, Metropolis-Hastings, hybrid MCMC, etc.; see Neal, 1993, for a statistical review)

2. Structure of implementation (schedule of approximate distributions for importance sampling; parameterisation and blocking in Gibbs; form of jumping rule in Metropolis-Hastings; choice of auxiliary variables and possible tempering schemes; etc.

3. Details of implementation (scaling of jumping rules, updating schedules, etc.)

4. Starting points of the simulations and number of parallel sequences

Some of the most effective or promising approaches to diagnosing convergence are based upon design considerations.

- Most familiarly, it is desirable to design an efficient implementation of the simulation algorithm to avoid difficulties such as MCMC samplers getting stuck near local modes.

- More sophisticated approaches such as coupling and regeneration involve designing the simulation algorithm in order that they work effectively (e.g., with reasonably short regeneration times).

- Auxiliary-variable methods can allow parameterisations (Liu *et al*, 1998) and model indicators (Green, 1995) to be random variables, which gives additional flexibility in design.

- Simulating multiple sequences allows the use of between and within variance components to monitor convergence, and is also useful in coupling schemes (Johnson, 1996, 1998).

- As noted above, the local property of most Markov chain simulation algorithms allows the user to identify "convergence" with "mixing."

- Overdispersed starting points, along with the local property, allows the user to compare the increasing within-sequence variance to the decreasing between-sequence variance (see Brooks and Gelman, 1998).

3. New Methods of Detecting Mixing using Multiple Sequences

Brooks and Gelman (1998) and Brooks and Giudici (1998) propose some generalisations of the method of Gelman and Rubin (1992a) for monitoring convergence using output analysis of multiple sequences of Markov chain simulation. We can categorise the generalisations as follows:

1. Monitoring "mixing" in terms of increasing within-sequence variability and decreasing between-sequence variability, rather than simply the ratio of the two. Brooks and Gelman (1998) illustrate with an example that lack of convergence can sometimes be detected by examining the empirical within and between-sequence variances on a single plot.

2. More effective construction of overdispersed starting points. In some examples, these can be obtained simply by sampling from the prior distribution, but in more complicated problems this will not work because, if the parameters are so overdispersed that they have very low probability in the target distribution, then the first few steps of an MCMC algorithm may pull them all the way to the centre of the distribution, thus removing the overdispersion.

   A. Zaslavsky (*pers. comm.*) proposes the following strategy: set some of the hyperparameters at overdispersed values, then run the Markov chain for a short time with these hyperparameters fixed so that each sequence settles down to a relatively stable position. These are then used as overdispersed starting points.

   More elaborate schemes are possible using simulated annealing or tempering ideas (Geyer, 1991), for example. At first it may seem excessive to consider using such elaborate approaches just to obtain overdispersed starting points—but in many examples, reliable convergence monitoring is as important as efficiency in simulation, and it is an important research goal to merge these two aims, rather than taking either of the two extremes of (a) blindly relying upon output analysis to judge convergence in complicated problems, or (b) designing a simulation that might be efficient but does not have the information to allow one to diagnose convergence problems.

3. Summarising mixing using more general statistics than variances. Brooks and Gelman (1998) report successes (and some counterintuitive results) using interval widths and coverage probabilities. Both these summaries are appealing because they closely correspond to the methods used to summarise inference after convergence.

4. Generalising the analysis-of-variance approach to measure different levels of mixing. Brooks and Giudici (1998) apply the analysis of variance approach to output from MCMC algorithms that move in a varying-dimensional parameter space (as in the reversible jump MCMC algorithm of Green, 1995). In this case, there are variance components between and within *models* as well as sequences (essentially a 3-way ANOVA). In addition, one could apply similar hierarchical analysis to elaborate structures of starting points (for example, 3 sequences starting from each of 5 overdispersed starting points), or for more complicated algorithms that involve the splitting of multiple sequences (e.g., Gilks and Berzuini, 1998). In the latter case, one would expect mixing to begin after each split.

5. Extending the variance-based approach to consider vector statistics of interest. Brooks and Gelman (1998) generalise the original univariate diagnostic to consider groups of parameters simultaneously. They show how sample variance-covariance matrices can be compared via their principal eigenvalues in order to gain a scalar *potential scale reduction factor* (PSRF) providing an upper bound to the original PSRF's formed from monitoring any of the corresponding parameters individually. This generalisation improves the interpretability of the original diagnostic, for problems where many parameters need to be monitored simultaneously.

4. Monitoring Convergence using Invariants of the Target Distribution

Another approach to detecting lack of convergence is to estimate, using simulation, quantities that have known values under the target distribution. If θ denotes the parameter vector sampled via iterative simulation, then we can use simulation draws to estimate $\mathbb{E}[h(\theta)]$ for any computable function h. Many diagnostic techniques are based upon monitoring functionals which converge to some specific value. However, in general this value is not known and so the resulting diagnostic is rather hard to interpret in that it may have settled to some value, but it is unclear whether or not it is the *true* value. Of course, these problems would be removed if we knew what the true expectation of h was under the stationary distribution, and current work on this idea is focussed upon trying to find functions, or families of functions, for which this is the case.

As an example, one such function is the score function. If $\theta \in E \subseteq \mathbb{R}^K$, and we let $\pi(\theta)$ denote the target distribution for the simulations, then we might take

$$h_k(\theta) = \frac{d \log \pi(\theta)}{d\theta_k}, \quad k = 1, ..., K.$$

It is simple to show that, under fairly general conditions on the density π, $\mathbb{E}_\pi[h_k(\theta)] = 0$ for all $k = 1, ..., K$. Thus, we might monitor each of these h_k functions until they appear to settle to around zero.

In addition, because it focuses upon univariate summaries, this approach might be combined with the convergence check based on path sampling (Gelman and Meng, 1998), which in turn is related to checking the marginal distribution of simulated tempering (Geyer and Thompson, 1995). All of these approaches have the virtue of being applicable to iterative simulations in general, not just MCMC, because they do not make use of the locality property.

5. Monitoring Convergence using Expected Perturbations

Consider the following situation: you have 5000 simulation draws, obtained by subsampling from long MCMC runs, and that you are willing to believe have approximately converged to the posterior distribution of a particular model. Data are now added or removed, creating a new posterior distribution that you wish to summarise. (Adding data is natural; removing data could occur in a study of optimal design or for cross-validation.) You can write the unnormalised posterior density for the new distribution, but you would rather not take the time to run a full MCMC simulation from this model. Rather, you will compute the importance ratios (easy to compute: they depend only on the added or subtracted data) of the new versus the old model for the 5000 draws, and then summarise the new posterior distribution using importance-weighted averages.

As discussed in Section 1.1, importance-weighted averages are in reality iterative simulation estimates, and it is vital in this situation to assess the convergence given the finite sample of size $n = 5000$.

We first note that, if necessary, we can improve the simulations by using the importance-weighted draws as starting points for an MCMC scheme such as the Metropolis-Hastings algorithm, or perhaps as part of a more elaborate method combining importance weights with MCMC jumps; see Gilks and Berzuini (1998), Wong and Liang (1997) and Liu *et al* (1998). This additional MCMC is only necessary if the desired target distribution is shifted in some way from the importance-weighted approximate distribution. This suggests the following diagnostic: run the Metropolis-Hastings algorithm for one step and see if the expected shift in the weighted means of the simulations is significantly different (statistically) from zero. Note that by computing only one step, we can average over the acceptances and rejections of the Metropolis-Hastings jump and thus achieve higher precision in our diagnostic with no extra computing cost.

A simple differential analysis shows that, in the limit of infinitely small spherical Metropolis jumps, this method is equivalent in expectation to the score-function convergence diagnostic in Section 4. This suggests that the score-function diagnostic is superior (as it eliminates the variability due to the Metropolis jumping itself); however, the expected-perturbations test may be more effective in some examples as it allows larger jumps and is thus not a purely local measure.

6. Discussion

No discussion of the issue of convergence assessment techniques could be complete without some more general discussion of the wider context of their use.

One issue relating to convergence assessment that is rarely discussed in the literature is the fact that deciding to stop the simulation on the basis of an output-based diagnostic can induce a bias in the resulting estimates. Cowles *et al* (1997) illustrate this idea for a number of simple models and diagnostic techniques. A simple illustration of the general idea can be seen by observing that stationarity is less likely to be diagnosed on occasions when the sample path is out in the tails of the distribution, and so variances (for example) are likely to be underestimated when many of the standard convergence diagnostics are used. Of course, the effect of this bias can be minimised by using overdispersed starting points and generating large post-convergence samples. However, the existence of a bias in such simple cases raises the question of what may happen for more complicated problems where both the sampling algorithm and posterior surface may be less well understood.

Another issue, discussed by Brooks and Gelman (1998), is that the question of convergence depends, in general, upon what the simulations will be used for. For example, when computing posterior intervals, there is a natural limit on the necessary precision of inferences (e.g., the 95% interval $[3.5, 8.4]$ is as good, in practice, as $[3.51345, 8.37802]$). In contrast, when estimating functionals such as posterior expectations (which are generally unnecessary in Bayesian inference, but are needed in decision analyses and in certain physical models such as those for which MCMC methods were originally designed), the required precision of inferences must be given externally. Thus, no automatic convergence test could work in such a setting without some input as to the desired precision level.

In conclusion, the goal of this article is to briefly discuss some of the important ideas, principles, and assumptions underlying convergence diagnostics for

iterative simulation, with the hope that future methods can be developed combining features of existing approaches. Most notably, we believe that existing analysis-of-variance methods (which implicitly rely on locality of MCMC simulations) can be made more effective by systematically exploiting the design of simulation algorithms (interpreting "design" quite generally, to include not just the algorithm and the number of simulated sequences, but also implementation issues such as updating and heating/cooling schedules). There also seems to be potential in the use of available "redundant" information such as importance ratios and unnormalised density functions.

References

Brooks, S. P. (1998a), Markov Chain Monte Carlo Method and its Application. *The Statistician* **47**, 69–100

Brooks, S. P. (1998b), Quantitative Convergence Diagnosis for MCMC via CUSUMS. *Statistics and Computing*. In press

Brooks, S. P., P. Dellaportas and G. O. Roberts (1997), A Total Variation Method for Diagnosing Convergence of MCMC Algorithms. *Journal of Computational and Graphical Statistics* **6**, 251–265

Brooks, S. P. and A. Gelman (1998), Alternative Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*. In press

Brooks, S. P. and P. Giudici (1998), Diagnosing Convergence of Reversible Jump MCMC Algorithms. Technical report, University of Bristol

Cowles, M. K., G. Roberts and J. S. Rosenthal (1997), Possible Biases Induced by MCMC Convergence Diagnostics. Technical report, Department of Statistics and Actuarial Science, University of Iowa

Cowles, M. K. and J. S. Rosenthal (1998), A Simulation Approach to Convergence Rates for Markov Chain Monte Carlo. Technical report, Harvard School of Public Health

Cui, L., M. A. Tanner, D. Sinha and W. J. Hall (1992), Monitoring Convergence with the Gibbs Sampler: Further Experience with the Gibbs Stopper. (Discussion of Gelman and Rubin 1992a). *Statistical Science* **7**, 483–486

Fosdick, L. D. (1959), Calculation of Order Parameters in a Binary Alloy by the Monte Carlo Method. *Physical Review* **116**, 565–573

Gelfand, A. E. (1992), Discussion of Gelman and Rubin (1992). *Statistical Science* **7**, 486–487

Gelman, A. (1992), Iterative and non-iterative simulation algorithms. *Computing Science and Statistics* **24**, 433–438

Gelman, A. and X. L. Meng (1998), Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling. *Statistical Science*. In press

Gelman, A., G. O. Roberts and W. R. Gilks (1996), Efficient Metropolis Jumping Rules. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics 5*, New York: Oxford University Press

Gelman, A. and D. B. Rubin (1992a), Inference from Iterative Simulation using Multiple Sequences. *Statistical Science* **7**, 457–511

Gelman, A. and D. B. Rubin (1992b), A Single Series from the Gibbs Sampler Provides a False Sense of Security. In J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics 4*, pp. 625–631, New York: Oxford University Press

Geweke, J. (1992), Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. In J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.), *Bayesian Statistics 4*, pp. 169–193, New York: Oxford University Press

Geyer, C. J. (1991), Markov Chain Monte Carlo Maximum Likelihood. In E. M. Keramidas (ed.), *Computing Science and Statistics: Proceedings of the 23rd Symposium of the Interface*, pp. 156–163, Interface Foundation

Geyer, C. J. and E. A. Thompson (1995), Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *Journal of the American Statistical Association* **90**, 909–920

Gilks, W. R. and C. Berzuini (1998), Following a Moving Target – Monte Carlo Inference for Dynamic Bayesian Models. Technical report, MRC Biostatistics Unit, Cambridge

Gilks, W. R., S. Richardson and D. J. Spiegelhalter (1996), *Markov Chain Monte Carlo in Practice*. Chapman and Hall

Green, P. J. (1995), Reversible Jump MCMC Computation and Bayesian Model determination. *Biometrika* **82**, 711–732

Heidelberger, P. and P. D. Welch (1983), Simulation Run Length Control in the Presence of

an Initial Transient. *Operations Research* **31**, 1109–1144

Johnson, V. E. (1996), Studying Convergence of Markov chain Monte Carlo Algorithms using Coupled Sample Paths. *Journal of the American Statistical Association* **91**, 154–166

Johnson, V. E. (1998), A Coupling-Regeneration Scheme for Diagnosing Convergence in Markov Chain Monte Carlo Algorithms. *Journal of the American Statistical Association* pp. 238–248

Kass, R. E., B. P. Carlin, A. Gelman and R. M. Neal (1997), MCMC in Practice: A Roundtable Discussion. Technical report, Carnegie Mellon University

Kong, A. (1992), A Note on Importance Sampling using Standardised Weights. Technical report, Department of Statistics, University of Chicago

Liu, C., J. Liu and D. B. Rubin (1993), A Control Variable for Assessment the Convergence of the Gibbs Sampler. In *Proceedings of the Statistical Computing Section of the American Statistical Association*, pp. 74–78

Liu, C. Rubin, D. B. and Y. Wu (1998), The PX-EM Algorithm. *Biometrika*. In press

Liu, J. S., F. Liang and W. H. Wong (1998), Dynamic Weighting in Markov chain Monte Carlo. Technical report, Stanford University

Mykland, P., L. Tierney and B. Yu (1995), Regeneration in Markov Chain Samplers. *Journal of the American Statistical Association* **90**, 233–241

Neal, R. M. (1993), Probabilistic inference using Markov Chain Monte Carlo methods. Technical report, Department of Computer Science, University of Toronto, Technical Report No. CRG-TR-93-1

Nummelin, E. (1978), A Splitting Technique for Harris Recurrent Markov chains. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **43**, 309–318

O'Hagan, A. (1987), Monte Carlo is Fundamentally Unsound. *The Statistician* **36**, 247–249

Ostland, M. and B. Yu (1997), Exploring Quasi Monte Carlo for Marginal Density Approximation. *Statistics and Computing* **7**, 217–228

Propp, J. G. and D. B. Wilson (1996), Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics. *Random Structures and Algorithms* **9**, 223–252

Raftery, A. E. and S. M. Lewis (1992), How Many Iterations in the Gibbs Sampler? In J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.), *Bayesian Statistics 4*, Oxford University Press

Roberts, G. O. (1992), Convergence Diagnostics of the Gibbs sampler. In J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger (eds.), *Bayesian Statistics 4*, Oxford University Press

Roberts, G. O. and J. S. Rosenthal (1998), Convergence of Slice Sampler Markov Chains. *Journal of the Royal Statistical Society, Series B*. In press

Roberts, G. O. and R. L. Tweedie (1998a), Bounds on Regeneration Times and Convergence Rates for Markov Chains. Technical report, University of Cambridge

Roberts, G. O. and R. L. Tweedie (1998b), Rates of Convergence of Stochastically Monotone and Continuous Time Markov Models. Technical report, University of Cambridge

Rosenthal, J. S. (1995), Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *Journal of the American Statistical Association* **90**, 558–566

Wong, W. H. and F. Liang (1997), Dynamic Weighting in Monte Carlo and Optimisation. *Proceedings of the National Academy of Sciences* **94**, 14220–14224

Yu, B. (1995), Discussion to Besag *et al* (1995). *Statistical Science* **10**, 3–66

Yu, B. and P. Mykland (1998), Looking at Markov Sampler through Cusum Path Plots: A Simple Diagnostic Idea. *Statistics and Computing*. In press