Authors' rejoinder to

Markov chain Monte Carlo: Some practical implications of theoretical results

Gareth O. Roberts* and Jeffrey S. Rosenthal**

We thank the two discussants for their interesting and thought-provoking comments. Here, we present a brief reply, organised by topic.

1. Dirichlet process priors.

Ishwaran has presented an analysis of a model (the Rasch model) using MCMC with Dirichlet process priors, following the approach of Escobar (1994) and MacEachern (1994). The analysis is a good example of how applied statisticians should approach such problems. Ishwaran is well informed about available theoretical results, and makes use of them where possible. At the same time, he is aware of the current limitations of such results, and is not afraid to use ad-hoc techniques and intuitive reasoning where necessary. We wish that more applied users would strike this same balance.

Ishwaran concludes his analysis by discussing how to monitor the convergence of his MCMC algorithm. We certainly agree with him when he says "The ideal solution would have been to derive quantitative rates of convergence". Coincidentally, we have recently begun working (Petrone, Roberts, and Rosenthal, 1997) on convergence results for models similar to that which he considers. However, this work has encountered many obstacles, and currently available results are not good enough to be of practical value. This highlights the need for a pragmatic approach to implementation, as both discussants have argued. Because of these difficulties, it is perfectly reasonable that Ishwaran instead uses convergence diagnostics, following Gelman and Rubin (1992), a popular approach (see Brooks

* Statistical Laboratory, University of Cambridge, Cambridge CB2 1SB, U.K. Internet: `G.O.Roberts@statslab.cam.ac.uk`.. Supported in part by EPSRC of the U.K.

** Department of Statistics, University of Toronto, Toronto, Ontario, Canada M5S 3G3. Internet: `jeff@utstat.toronto.edu`. Supported in part by NSERC of Canada.

and Roberts (1996) for a detailed discussion of this method, and Brooks and Gelman (1996) for a useful multivariate extension). While such diagnostics are always fraught with risks, we feel that Ishwaran has done a good job of implementing them.

2. Central limit theorems.

Madras discusses the relation between geometric ergodicity and central limit theorems. It is indeed true, as he says, that while geometric ergodicity *implies* the existence of central limit theorems, the converse is *not* true, and "It is possible to have chains which are not geometrically ergodic but in which the central limit theorem does hold." Such matters are considered in some detail in Roberts (1996). Note however that in our example in Section 4, practical problems are caused not only by the fact that the limit of the scaled variances is finite. The heavily skewed nature of the distribution of the sample path mean makes detection of poor convergence particularly difficult, since most sample paths give adequate looking autocorrelation plots (see our Figure 2).

As we say in the paper, we still feel that proving geometric ergodicity is by far the easiest and most common method of establishing central limit theorems. Furthermore geometric ergodicity is an important property in its own right. It is indeed possible, as Madras notes, that non-geometric chains could still be "pretty efficient" (in some sense) for some situations. However, we do not know of many cases where this has been established.

3. Periodicity.

We completely agree with Madras's observation that periodicity (or near-periodicity) is not a major concern in practice, since a user will typically average a number of observations so that periodic behaviour will cancel out. Moreover, certain algorithms are provably *spectrally positive*, such as the random scan Gibbs sampler, ruling out any kind of periodicity. However, from the point of view of theoretical analysis of convergence rates, periodicity has been a major inconvenience. A number of authors (e.g. Meyn and Tweedie, 1994; Diaconis and Stroock, 1993) have developed complicated analyses to control periodic behaviour, while others (e.g. Jerrum and Sinclair, 1989; Rosenthal, 1996b) have used tricks such as insisting that the chain have fixed positive probability of not moving at each step.

Essentially, the difficulty with direct theoretical analysis was that to prove the chain was converging to stationarity, it was necessary to *couple* the chain with a second, stationary chain, which happens to be at the same place *at the same time*.

This difficulty was removed by applying the notion of shift-coupling in Roberts and Rosenthal (1994). This allowed the two chains to be at the same place at *different* times, a much easier requirement. This led to substantially improved bounds, when considering the average distribution $\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}(X_i)$. We felt that with this analysis we were, in some sense, getting to the heart of the matter by avoiding concerns about near-periodic behaviour, in exchange for averaging over many different observations.

4. Convergence diagnostics.

Both Ishwaran and Madras place great emphasis on the use of empirical convergence diagnostics to assess convergence of MCMC algorithms. They both argue (and we agree) that rigorous quantitative bounds are too difficult to be used routinely, so that diagnostics may be all that are left. Nevertheless, they both recognise the risks inherent in such an approach, and in particular the risk of prematurely diagnosing convergence of the chain.

Madras urges caution in this regard, and provides some helpful tips (don't cut corners, estimate autocorrelations, try different starting points, etc.) to avoid making unfortunate mistakes. However, we do note that such procedures are not sufficient to avoid premature diagnosis of convergence in all cases. For example, the "witch's hat example" discussed in our paper, simple though it may be, can fool most convergence diagnostics.

On the other hand, Madras also urges us to "Understand your simulation, and try to guess where it might have problems". Perhaps this is the most important advice to give: Don't rely on *any* routine diagnostics, rather try somehow (through theoretical analysis, through empirical observation, or through intuitive reasoning) to figure out exactly what your algorithm is doing and how it might fail to converge quickly.

5. Communication.

Both Madras and Ishwaran urge greater communication – between statisticians and physicists, and between theoretical and applied users of MCMC algorithms. We completely support this goal. The more we can learn from each other, the more progress we can make, and the less time we will waste re-discovering what is already known. In this regard, we commend Madras in particular for his excellent work at moving between these disparate groups.

We further urge all researchers to make efforts to learn about others' ideas, and to present their own ideas in such a way that others can easily understand them. This is particularly important since one of the major communication problems between the physics and statistics literatures is one of language. For instance, it is unfortunate that the phrase "Gibbs sampler" has caught on in the statistics literature when the term "Glauber dynamics" was already firmly established in physics. Such anomalies only exacerbate attempts to bridge these two groups.

We also commend the editors of the *Canadian Journal of Statistics* for inviting this discussed paper – the first of many we hope – to help improve communication among different research communities.

ADDITIONAL REFERENCES

S.P. Brooks and A. Gelman (1996), General methods for monitoring convergence of iterative simulations. J. Comp. Graph. Stat., to appear.

S. Petrone, G.O. Roberts, and J.S. Rosenthal (1997), Convergence rates of Dirichlet-type processes. Work in progress.