Comment on the article by Stephens and Donnelly

Stephen Brooks and Andrew Gelman

1 March 2000

First of all, we would like to congratulate the authors on a very stimulating paper in which they tackle a very interesting problem. Our attention was drawn in particular to section 6.4 where we see some overlap with our own work.

One 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 $\mathrm{E}U(\theta)$ for any computable function U. Many diagnostic techniques are based upon monitoring functions that 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 it is the *true* value (see, e.g., Gelman and Rubin, 1992a). With Markov chain Monte Carlo algorithms, it is often possible to diagnose convergence with multiple overdispersed sequences, but this approach does not work with algorithms such as importance sampling that do not have local dependence (see Brooks and Gelman, 1998). This is one reason why we have found convergence monitoring to be easier for MCMC than for importance sampling. For this reason, we welcome the suggestion in Section 6.1 to use the new importance sampling methods in an expanded MCMC framework.

To return to importance sampling: the difficulties in monitoring convergence of functions $\mathrm{E}(U(\theta))$ would be removed if we knew what the true expectation of U was under the stationary distribution,

and there are in fact some functions U for which this is the case. One such function is the score function. If $\theta \in R^K$, and we let $\pi(\theta)$ denote the target distribution for the simulations, then we might take

$$U_k(\theta) = \frac{\partial \log \pi(\theta)}{\partial \theta_k}, \quad k = 1, \ldots, K.$$

Under fairly general conditions on the density π, $\mathrm{E}_\pi U_k(\theta) = 0$ for all $k = 1, ..., K$.

Thus, one might monitor the sample mean of each of these U_k functions as the simulations proceed, until they appear to settle to around zero. One can estimate the standard error of the $U_k(\theta)$ from parallel independent runs of the importance sampling or MCMC procedure, so as to determine whether or not observed values are "significantly" different from zero.

Incidentally, it is not necessarily true that, as claimed in the paragraph before Section 5.1, that parallel simulation runs are computationally more expensive. For one thing, the results of the parallel runs can themselves be averaged, so that inference from several runs is more efficient than from any single run. Even more important, parallel runs can give confidence about accuracy of simulation results that can sometimes allow one to stop the simulations far earlier than might be done under the condition of insecurity arising from using a single simulation run (see Gelman and Rubin, 1992b, and accompanying discussion).

References

Brooks, S., and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*.

Gelman, A., and Rubin, D. B. (1992a). A single sequence from the Gibbs sampler gives a false sense of security. In *Bayesian Statistics 4*, ed. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, 625–631. New York: Oxford University Press.

Gelman, A., and Rubin, D. B. (1992b). Inference from iterative simulation using multiple sequences (with discussion). *Statistical Science* **7**, 457–511.