Comment

Jeffrey S. Rosenthal

The papers by Geyer and Gelman-Rubin discuss the question of whether one long run or several shorter runs are preferable when using Markov chain Monte Carlo (MCMC). We feel that there are good arguments on both sides. We agree with Gelman-Rubin that multiple runs allow for more effective use of diffuse, well-chosen starting distributions. Furthermore, they can help identify situations (Section 4.8) where a Markov chain happens to get "stuck" in a low-probability region for too long. On the other hand, they require throwing away more initial samples to allow for multiple "burn-in periods." Furthermore, we share Geyer's concern (Section 4) about the difficulty of obtaining useful starting distributions in the first place. The debate will continue to rage, as these pages show. We merely wish to observe that spectacular successes have been achieved with each strategy.

Each of these papers presents useful methods for monitoring convergence of estimates to their correct values. We appreciate Geyer's careful discussion (Section 3) about estimating the variance of a function $\hat{\mu}_n$, using a version (Section 2) of the central limit theorem (CLT). (Unfortunately, Geyer does not appear to have considered the rate of convergence of this CLT; in other words, how long does the MCMC have to be run before the given normal approximation is valid?) We also appreciate the methods described by Gelman-Rubin (Section 2) for estimating convergence by using multiple runs with an overdispersed starting distribution and then monitoring the "potential scale reduction." However, despite the authors' admirable efforts to present their methods straightforwardly and directly, each method still seems to involve certain "heuristics," such as difficult choices about weight functions and bandwidths (Geyer, Section 3) or "simply discarding" certain undesirable sequences (Gelman-Rubin, Section 4.8). In general, we feel that MCMC can be used with greater confidence if it is more automated and requires less "poking around."

We agree with Geyer (Section 5) that "guarantees can only come from theoretical calculations. . . ." In this spirit, in Rosenthal (1991, 1991a), ideas related to Harris Recurrence are used to get specific, sharp theoretical bounds on time to convergence for MCMC for certain specific models, including (1991) the Gibbs sampler applied to a standard variance components model (where the required run length is shown to increase only logarithmically with increasing numbers of parameters). These bounds allow an MCMC to be run for a prespecified number of iterations, without any need for difficult or controversial monitoring techniques. Similarly, in Diaconis and Hanlon (1992), a Metropolis algorithm on the set of permutations is explicitly diagonalized, giving results on convergence rate. We feel that further theoretical results such as these could provide solid, quantitative bases for running MCMCs, thereby eliminating some of the difficulties and heuristics that are often encountered.

Jeffrey S. Rosenthal is Assistant Professor, School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455.

Comment

Bruce Schmeiser

Estimating high-dimensional volumes is analogous to estimating steady-state performance of computer, communications or manufacturing systems. The issues now attracting wide attention in the statistics community—alleviating initial bias, estimating precision and improving point-estimator quality—are long-standing problems in the operations-research literature. Based on the operations-research community's decades-old, continuing debate of one long replication versus many shorter replications, I doubt that the statistics community will soon reach a consensus. Having little hope of aiding a consensus, I only briefly discuss the number of

Bruce Schmeiser is Professor, School of Industrial Engineering, Purdue University, West Lafayette, Indiana 47907-1287.

replications before listing several points and references that I hope are new to some readers.

I enjoyed both papers, but my general preference remains unchanged: Use a single long replication, except in special cases such as parallel computing or when stratified or antithetic initial states happen to be easy to determine. This preference for a single replication is due to its robustness to analyst lack of sophistication or time. Fifteen years ago substantial background, insight and effort were required for simulation and for statistics practitioners to analyze complex problems. Commercial software has blossomed in both fields, allowing relatively naive practitioners to expect something good to happen when they give their problem to the computer. Similarly, one day we will expect software to evaluate posterior distributions with little practitioner insight. The single long replication makes negligible the initial bias, thereby alleviating the difficult initial-data-deletion problem.

Glynn and Heidelberger (1992) and Kelton (1989) are recent additions to the extensive literature that discusses initial deletion of warm-up data and choice of initial states.

Glynn (1987), Whitt (1990) and Damerdji (1991) discuss the choice of number of replications, the extreme cases being a single long run and many short runs.

Smith (1984) discusses Monte Carlo sampling from doubly stochastic Markov chains. The motivation is the need to identify nonredundant constraints in mathematical programming. The methods can be used to sample from a density by sampling uniformly within the region defined by the density and the zero plane. The Hit-and-Run sampler (Belisle, Romeijn and Smith 1992) is a generalization to nonuniform distributions.

Since essentially all point estimators are asymptotically normal, sampling error is well summarized by point-estimator standard error. The method of nonoverlapping adjacent batch means (NBM) is extended to overlapping batch means (OBM) in Meketon and Schmeiser (1984). OBMs are highly dependent, which is acceptable since batches are sufficiently large not when the batch means are essentially independent but (loosely) when each batch subsumes the autocorrelation structure.

Except for end effects, OBM is the Bartlett-window spectral estimator with lag-window length equal to the batch size. Therefore, OBM has the same bias but only two-thirds the variance of NBM. Both NBM and OBM estimator are easily computed in $O(n)$ time; therefore OBM dominates NBM for Markov chain sampling.

For NBM, OBM and some other estimators based on batching, the mse-optimal batch size is asymptotically

$$m^* = \left[2n\left(\frac{c_b^2}{c_v}\right)\left(\frac{\gamma_1}{\gamma_0}\right)^2\right]^{1/3},$$

where c_b and c_v are the estimator's bias and variance constants, respectively, and γ_1 / γ_0 is the center of gravity of the absolute values of the autocorrelation lags. For OBM $c_b = 1$ and $c_v = 4/3$. Geyer's Theorem 3.1 helps to estimate the autocorrelation center of gravity, which is problem dependent, but the goal is to estimate optimal batch size without estimating individual autocorrelations.

An advantage of batch-means methods is that they extend directly to point estimators that are not means. Schmeiser, Avramidis and Hashem (1990) discuss sufficient assumptions and provide a code for overlapping batch variances and overlapping batch quantiles.

Nelson (1989) quantifies the additional number of batches needed when estimating optimal weights for control variates.

For random-number and random-variate generation, see Fishman and Moore (1986) and Devroye (1986), respectively.

Glasserman (1991) discusses single-replication methods for estimating derivatives of performance measures with respect to system design parameters. Could similar methods be used to estimate the change, for example, in the posterior mean caused by a unit change in the prior mean?

A variety of other simulation-experiment issues are discussed in Schmeiser (1990).

Comment

Luke Tierney

Both papers make some interesting contributions to the discussion of issues related to Markov chain Monte Carlo. Geyer's variance estimates that take advantage of the Markov chain structure appear to be particularly promising and worthy of further investigation. As these methods require a reversible chain, they are not directly applicable to the fixed scan Gibbs sampler. But several simple devices are available for making Gibbs samplers reversible, including random scans,

Luke Tierney is Professor, School of Statistics, University of Minnesota, Vincent Hall, Minneapolis, Minnesota 55455.