Comment

Neal Madras

Professors Gelman, Rubin and Geyer have presented us with many interesting ideas to think about. These two papers have been billed as representing opposing sides in a debate, but to my mind they complement more than they contradict each other. This is probably because I find the "debate" somewhat artificial, with each side having its merits as well as its limitations. It is true that one long run gives more statistical information per CPU minute than do several shorter runs (because of warm-up time); however, if you suspect slow mixing for some particular reason, then you should do your best to investigate its likely sources – and multiple runs with intelligently selected starting points are probably the most natural way to detect such a problem. I think that most researchers who use Monte Carlo simulation will find useful things in both papers.

Guarantees are hard to come by in iterative simulation. At present, few Markov chains are sufficiently tractable to yield good rigorous upper bounds on the amount of time necessary to run simulations (although the last paragraph of Geyer's Section 3.5 is a nice observation that may make even relatively weak bounds useful in some cases). What can we do when rigorous analysis eludes us?

Geyer claims that guarantees can come from "experiments with a range of sampling schemes proceeding in small steps from schemes known to mix rapidly to the scheme of interest, making sure at each step that the run is long enough by comparing it to the runs already done." This is a very appealing idea, and it may work very well sometimes, but it is hard for me to believe that this method comes with general guarantees. The problem is that two Markov chains whose transition probabilities are very close may have very different properties. For example, the Ising model in statistical physics is a family of distributions indexed by a parameter $T > 0$ ("temperature") and having a *critical value* T_c (corresponding to a *phase transition* in the associated infinite system). There is a standard implementation of the Gibbs sampler (or the Metropolis algorithm) to the Ising model, giving a Markov chain at each temperature T. We need not be concerned with the details of the model and implementation [which may be found in Geman and Geman (1984) or Ripley and Kirkland (1990)], except to say that the Markov chain is rapidly mixing if T is large and slowly mixing if T is small, and that the change occurs fairly abruptly in a small neighborhood of T_c. Consider a sequence of schemes in which the ith scheme corresponds to the Markov chain at temperature T_i, where $T_1 > T_2 > \cdots$. If $T_3 > T_c > T_4$, say, then the small value of σ^2 at T_3 may provide a false sense of security at T_4, and the entire run at T_4 may be too short to permit the chain to escape some subset of the state space, which is metastable at T_4 but not at T_3. The resulting estimate of σ^2 at T_4 would be far too small. Similar abrupt changes can occur in simulated annealing, perhaps at more than one "critical value" [see p. 677 of Kirkpatrick, Gelatt and Vecchi (1983)].

It is tempting to say that we can avoid such pitfalls if we are careful enough, but the point remains that important questions must be answered before accepting Geyer's claim. First of all, what can be said rigorously about a continuum of sampling schemes? How small should the steps be to ensure that σ^2 does not change too much from one scheme to the next, if it is indeed possible to make such assurances at all? Second, is the procedure feasible in practice? Perhaps for real problems it is just too time-consuming to run lots of simulations of different sampling schemes, and it would be just as efficient and informative to run the scheme of interest for ten times as long. These are intriguing problems that are worth investigating, both theoretically and experimentally.

To provide a context for my remaining remarks, I shall briefly describe one Monte Carlo study of a simple lattice model of polymers known as the *self-avoiding walk*. A linear polymer is a molecule that consists of many "monomers" (groups of atoms) joined sequentially by chemical bonds. The spatial configuration of a linear polymer with N (monomer–monomer) bonds can be modeled as a random walk path $W \equiv w(0), w(1), \ldots, w(N)$ on the three-dimensional integer lattice $\mathbf{Z}^3$; here $w(i)$ represents the location of the ith monomer, and $w(i)$ and $w(i+1)$ are always nearest neighbours in the lattice. But two monomers cannot occupy the same position in space; the simplest model that captures this effect is the *self-avoiding walk*, which is defined by requiring $w(0), \ldots, w(N)$ to be *distinct* sites of $\mathbf{Z}^3$.

Let S_N^o be the set of all N-step random walk paths in $\mathbf{Z}^3$ that start at the origin, and let S_N^* be the subset of walks in S_N^o that are self-avoiding. Let P_N^o and P_N^* denote the uniform probability measures on S_N^o and S_N^*,

Neal Madras is Associate Professor and NSERC University Research Fellow, Department of Mathematics and Statistics, York University, 4700 Keele Street, North York, Ontario M3J 1P3, Canada.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *Statistical Science*. www.jstor.org