Regeneration in Markov Chain Samplers

Per MYKLAND, Luke TIERNEY, and Bin YU*

Markov chain sampling has recently received considerable attention, in particular in the context of Bayesian computation and maximum likelihood estimation. This article discusses the use of Markov chain splitting, originally developed for the theoretical analysis of general state-space Markov chains, to introduce regeneration into Markov chain samplers. This allows the use of regenerative methods for analyzing the output of these samplers and can provide a useful diagnostic of sampler performance. The approach is applied to several samplers, including certain Metropolis samplers that can be used on their own or in hybrid samplers, and is illustrated in several examples.

KEY WORDS: Gibbs sampling; Hybrid sampler; Markov chain Monte Carlo; Metropolis algorithm; Simulation output analysis; Split chain.

1. INTRODUCTION

In Markov chain Monte Carlo, a distribution π is examined by obtaining sample paths from a Markov chain constructed to have equilibrium distribution π. This approach, introduced by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953), has recently received considerable attention as a method for examining posterior distributions in Bayesian inference and for approximating the relative likelihood function in maximum likelihood estimation (Besag and Green 1993; Gelfand and Smith 1990; Geyer 1994; Geyer and Thompson 1992; Gilks et al. 1993; Liu, Wong, and Kong, in press; Smith and Roberts 1993; Tanner and Wong 1987; Tierney 1991, in press; Yu 1993).

The analysis of the output produced by Markov chain samplers is more challenging than for other Monte Carlo methods, such as importance sampling, that are based on independent observations. The dependence in the samples makes estimating standard errors of Monte Carlo estimates more difficult. Furthermore, because it is usually not possible to start a Markov chain sampler with its equilibrium distribution, it may take some time for it to reach equilibrium, and thus it may be useful to discard some initial portion of the sample to reduce the effect of the initial distribution used.

One approach that can help reduce these problems is to try to identify *regeneration times* at which the chain restarts itself. The *tours* of the chain between regenerations are then independent and identically distributed. If the chain is observed for a fixed number of tours, then initialization issues are eliminated, and standard errors of sample path averages can be computed using methods based on iid observations. This approach is known as *regenerative simulation* (see, for example, Ripley 1987, sec. 6.4).

Regeneration times are easy to find for discrete Markov chains, if we fix a particular state, then the chain starts over every time it returns to that state. In general state-space Markov chains, where the transition densities and the stationary distribution may be continuous, the chain may never return to any particular state. Nevertheless, several authors have developed ways of introducing regeneration times into general state-space Markov chains (Athreya and Ney 1978; Nummelin 1978). The method of Nummelin (1978), called *splitting,* is well suited for use in regenerative simulation (Asmussen, Glynn, and Thorisson 1992; Kalashnikov 1992). Splitting is particularly easy to apply to a class of Metropolis samplers that can be used on their own or as components in hybrid samplers.

The article is organized as follows. Section 2 reviews the regenerative simulation method, and Section 3 introduces the general splitting technique of Nummelin. Section 4 discusses the application of splitting to some Metropolis chains and Gibbs samplers. Section 5 illustrates these approaches using several examples, and Section 6 presents some final comments. Proofs of several results are given in an Appendix.

2. REGENERATIVE SIMULATION ANALYSIS

A stochastic process $\{X_n : n = 0, 1, \ldots\}$ is regenerative if there are times $T_0 \le T_1 \le \cdots$ such that at each T_i, the future of the process is independent of the past and identically distributed. Then the tours of the process between these times are iid, and the times themselves form a renewal process. The renewal process is delayed if $T_0 \neq 0$.

Suppose that a regenerative process has equilibrium distribution π, that we wish to estimate $\theta = E_\pi[f]$ for some function f, and that the process is ergodic in the sense that sample path averages of f converge almost surely to θ. Also assume for the moment that we observe the process for a fixed number n of complete tours. Let $N_i = T_i - T_{i-1}$ and

$$Y_i = \sum_{j=T_{i-1}+1}^{T_i} f(X_j)$$

for $i = 1, \ldots, n$. Then the pairs (N_i, Y_i) are iid, and if $E[|Y_i|] < \infty$ and $E[N_i] < \infty$, then $\hat{\theta}_n = \sum Y_i / \sum N_i = \bar{Y}/\bar{N} \to \theta$ by the strong law of large numbers. If the Y_i and N_i have finite variances, then the distribution of $\sqrt{n}(\hat{\theta}_n - \theta)$ converges to a $N(0, \sigma^2)$ distribution, and σ^2 can be estimated using the variance estimation formula for a ratio estimator,

$$\hat{\sigma}^2 = \frac{\frac{1}{n}\sum (Y_i - \hat{\theta}_n N_i)^2}{\bar{N}^2}. \tag{1}$$

* Per Mykland is Assistant Professor, Department of Statistics, University of Chicago, IL 60637. Luke Tierney is Professor, School of Statistics, University of Minnesota, MN 55455. Bin Yu is Assistant Professor, University of California, Berkeley, CA 94720. Research of P. Mykland was supported in part by Grants DMS-8902667, DMS-9204504 and DMS-9305601 from the National Science Foundation. Research of L. Tierney was supported in part by Grant DMS-9005858 and DMS-9303557 from the National Science Foundation. Research of B. Yu was supported in part by Grant DAAL03-91-G-007 from the Army Research Office.

The accuracy of the ratio estimator variance formula (1) depends on the variability of its numerator and on the error in the Taylor series approximation of $\bar{Y}/\bar{N} - \theta$ by $(\bar{Y} - \theta\bar{N})/E[N_i]$ used in the delta method derivation of this formula. Both depend on the sample size and the variability of the tour lengths. In particular, the Taylor series error depends on the relative error of $\bar{N}$ as an estimator of $E[N_i]$; the expected absolute value of the Taylor series error is bounded by $(\mathrm{MSE}(\hat{\theta}_n, \theta)\mathrm{CV}(\bar{N}))^{1/2}$, where $\mathrm{MSE}(\hat{\theta}_n, \theta) = E[(\hat{\theta}_n - \theta)^2]$ and $\mathrm{CV}(X) = \mathrm{var}(X)/(E[X])^2$ denotes the coefficient of variation of X. As a result, many simulation texts (such as Bratley, Fox, and Schrage 1987 and Ripley 1987 recommend using formula (1) only when the relative error of $\bar{N}$ is small; jackknife estimators are recommended as a possible alternative when the relative error is not small. The coefficient of variation $\mathrm{CV}(\bar{N})$ can be estimated by $\widehat{\mathrm{CV}}(\bar{N}) = \widehat{\mathrm{CV}}(N_i)/n = \sum (N_i - \bar{N})^2/(n\bar{N})^2$. Formula (1) should be used with caution if $\widehat{\mathrm{CV}}(\bar{N})$ is larger than, say, 1%; this corresponds to an estimated bound on the Taylor approximation error in (1) of .1 $\mathrm{MSE}(\hat{\theta}_n, \theta)^{1/2}$.

In addition to computing $\widehat{\mathrm{CV}}(\bar{N})$, it is useful to examine the pattern of regeneration times graphically; for example, by plotting T_i/T_n against i/n. We will refer to this as a *scaled regeneration quantile* (SRQ) plot. An SRQ plot can provide a useful diagnostic for examining the performance of a sampler. If the total run length is long enough, then this plot should be close to a straight line through the origin with unit slope. This follows from the law of large numbers and also reflects the asymptotic uniform distribution of regenerations predicted by renewal theory; that is, the proportion of renewals that fall in a fraction α of the observation period converges to α. Deviations from this straight line suggest that the observation period is not long enough for the sampler to have reached equilibrium. Deviations occur in particular when some tours are substantially larger than others. Examining the states visited by the process during these longer tours might suggest improvements in the sampler. Otherwise, it may be necessary to use a longer run to reduce the impact of these longer tours.

Another interpretation of the SRQ plot is that its reflection, the plot of i/n against T_i/T_n is a scaled plot of an estimate of the renewal function over the observation period $[0, T_n]$.

The deviations of an SRQ plot from a straight line are related to $\widehat{\mathrm{CV}}(\bar{N})$ through the fact that $\widehat{\mathrm{CV}}(\bar{N})$ can be computed as the sum of the squares of the increments of the deviations $T_i/T_n - i/n$. The estimated coefficient of variation $\widehat{\mathrm{CV}}(\bar{N})$ is asymptotic to $\mathrm{CV}(N_i)n$. This provides a useful relationship between the coefficient of variation of the tour lengths, the number of observed tours, and the departure from uniformity of the regeneration time distribution. Once an estimate of $\mathrm{CV}(N_i)$ is available from a preliminary run, this relationship can be used to estimate a minimal number of tours for which formula (1) should produce acceptable results. Other numerical measures of uniformity of the renewal time distribution are of course possible and may be worth exploring. Asymptotic properties of such measures can be derived from the fact that the normalized deviations process converges to a Brownian bridge.

The assumption that the simulation is run for a fixed number of complete tours implies that the process is started with a regeneration and that the total run length T_n is random. Starting with a regeneration can be accomplished by running from an arbitrary starting point until a regeneration occurs and discarding the sample path up to this first regeneration. It is also possible to run the simulation for a fixed period and either treat the first and last observations as regenerations or use only the portion of the sample path between the first and last observed regenerations. This produces a bounded run length but a random number of observed tours. An intermediate option is to set a desired total run length t and continue until the first regeneration time greater than or equal to t. This has the technical advantage that the last observation time is a stopping time. All these approaches lead to sample path averages with the same asymptotic distribution, and formula (1) remains asymptotically valid. Biases can occur because of the effect of the waiting time paradox on the final tour, but these biases should be small if the simulation run length is long enough to produce an approximately uniform regeneration pattern. Ripley (1987, sec. 6.4) and Bratley et al. (1987, secs. 3.3.2 and 3.7) discuss these issues and give further references.

3. SPLITTING MARKOV CHAINS

The terminology used in this section is based on work of Nummelin (1984) and also has been defined by Tierney (in press). Let $\{X_n: n = 0, 1, \ldots\}$ be an irreducible Markov chain on a state-space $(E, \mathcal{E})$ with transition kernel $P = P(x, dy)$ and invariant distribution π. The sigma algebra $\mathcal{E}$ is assumed to be countably generated. These assumptions imply that X_n is positive recurrent (see, for example, Tierney in press, thm. 1). Assume in addition that X_n is Harris recurrent; this is satisfied by most Markov chain samplers (Tierney, in press, cors. 1 and 2; Chan and Geyer, in press, thm. 1).

A set $A \in \mathcal{E}$ is a *proper atom* for the Markov chain if $\pi(A) > 0$ and $P(x, \cdot) = P(y, \cdot)$ for all $x, y \in A$. If a chain has a proper atom, then the times at which the chain enters the atom are regeneration times. Few chains contain proper atoms, but it is often possible to construct a related chain that does. Suppose that it is possible to find a function $s(x)$ and a probability measure $\nu(dy)$ such that $\pi(s) = \int s(x)\pi(dx) > 0$ and

$$P(x, A) \geq s(x)\nu(A) \tag{2}$$

for all $x \in E$ and all $A \in \mathcal{E}$. A pair (s, ν) satisfying these conditions is called an *atom* for the transition kernel P. Atoms represent a generalization of proper atoms: If A is a proper atom, then $(1_A(x), P(\tilde{x}, \cdot))$ for some $\tilde{x} \in A$ is an atom. Condition (2) implies that we can write $P(x, dy) = s(x)\nu(dy) + (1 - s(x))Q(x, dy)$, where Q is a transition kernel defined as $Q(x, dy) = (P(x, dy) - s(x)\nu(dy))/(1 - s(x))$ if $s(x) < 1$ and, arbitrarily, as $Q(x, A) = 1_A(x)$ if $s(x) = 1$. Thus we can imagine generating X_{n+1}, given $X_n = x$, in two stages: First, generate a Bernoulli variable S_n with success probability $s(x)$. If $S_n = 1$, then generate X_{n+1} from ν. Otherwise, generate X_{n+1} from $Q(x, \cdot)$. The marginal sequence $\{X_n\}$ is a Markov chain with transition kernel P.

The pairs (X_n, S_n) form a Markov chain, the split chain, with a proper atom $E \times \{1\}$; the times at which $S_n = 1$ are regenerations times for this chain. This construction is given in Nummelin (1984, sect. 4.4).

Sampling from the kernel Q as required by the split chain construction may not be particularly convenient. Fortunately an alternative is available: We can generate the marginal sequence X_n as usual from P, and then generate the splitting variables S_n from their conditional distribution given the $\{X_n\}$ sequence. Conditionally, given the entire sequence $\{X_n, n = 0, 1, \ldots\}$, the Bernoulli variables $\{S_n\}$ are independent. Furthermore, conditionally on X_n and X_{n+1} the variable S_n is independent of all other X_i, and $P(S_n = 1 \mid X_n = x, X_{n+1} = y) = r(x, y)$, where

$$r(x, y) = \frac{s(x)\nu(dy)}{P(x, dy)} \tag{3}$$

is the Radon–Nikodym derivative, which exists by the absolute continuity implied by (2). Thus we can generate S_n as soon as X_{n+1} is available. This construction is summarized in the following theorem.

Theorem 1. Suppose that the pairs (X_n, S_n) are generated by choosing X_0 from E according to an arbitrary initial distribution and for each $n = 0, 1, \ldots,$ generating first X_{n+1} from $P(X_n, \cdot)$ and then S_n as a Bernoulli variable with success probability $r(X_n, X_{n+1})$. Then (X_n, S_n) is a Markov chain, the times when $S_n = 1$ are regeneration times (with probability 1 that the tours are all finite), and the expected tour length is $E[N_i] = 1/\pi(s)$, where $\pi(s) = \int s(x)\pi(dx)$.

The construction of the split chain depends on an atom (s, ν) only through the product $s(x)\nu(dy)$. Thus it is not necessary to determine the normalizing constant needed to make ν into a probability measure. It is sufficient to find a finite nonzero measure ν' and a function s' such that $s'(x)\nu'(dy) \le P(x, dy)$ and $\pi(s') > 0$; then the atom (s, ν) is given by $\nu(dy) = \nu'(dy)/\nu'(E)$ and $s(x) = s'(x)\nu'(E)$.

Atoms of a given transition kernel P need not exist. Nummelin (1984) showed that if $\mathcal{E}$ is countably generated, under the assumptions listed at the beginning of this section it is always possible to find an atom for the m-step transition kernel P^m for some $m \ge 1$. But in Markov chain simulations P^m is rarely available in closed form for any $m > 1$, so we consider only the case $m = 1$.

If an atom (s, ν) of P does exist, then it is not unique. For any s' and ν' with $\pi(s') > 0$ and $s'\nu' \le s\nu$ in the sense that $s'(x)\nu'(A) \le s(x)\nu(A)$ for all $x \in E$ and $A \in \mathcal{E}$, the pair (s', ν') is also an atom of P. The relation $s'\nu' \le s\nu$ provides a partial ordering on atoms. If two atoms (s', ν') and (s, ν) satisfy $s'\nu' \le s\nu$, then we would prefer to use the larger atom (s, ν), because it produces more regenerations in the sense that a split for (s', ν') can be constructed by first generating a split based on (s, ν) and then randomly deleting renewals. If two atoms are not comparable in terms of this relation, then it is not clear which one is preferred. If a choice has to be made, one approach is to pick a particular criterion, say mean tour length, and then choose the atom with the larger mean tour length as long as other characteristics, such as the coefficient of variation of the tour lengths, are reasonable. This is the approach that we adopt in the examples in Section 5. The mean tour length can be estimated by the average of observed tour lengths or as the inverse of an estimate of the regeneration rate $\pi(s)$ using a preliminary sample. The inverse regeneration rate has the advantage that an estimate of the form $\widehat{\pi(s)} = (1/n)\sum_{i=0}^{n-1} r(X_i, X_{i+1})$ may provide a better estimate than the proportion of observations resulting in regenerations.

The conditional regeneration probability (3) may not always be easy to calculate. But in many examples the generation of variables from P may produce additional information that allows generation of the Bernoulli variables S_n even if $r(x, y)$ is not explicitly available. Two examples are the Metropolis kernel and hybrid kernel splits discussed in Sections 4.1 and 4.3.

The split process is constructed by adding randomizations to an observed X_n sample path. Given the sample path, one could repeat the determination of the split variables several times and combine the resulting variance estimates. Sample path averages will not be affected, assuming that the entire observed sample path of the X_n process is always used. But variance estimates do change with the splitting variables, and repeated sampling can reduce the contribution of the splitting randomization to the variability in variance estimates. Conditional resampling with the first, last, or total number of regenerations kept fixed is also possible.

4. SPLITTING SOME MARKOV CHAIN SAMPLERS

Two general approaches to incorporating regeneration into a Markov chain sampler are available. The first approach is to attempt to find an atom for the sampler itself. This is possible for certain special Metropolis and Gibbs samplers. If it is not possible to find an atom for the original sampler, then the second approach is to form a hybrid sampler that incorporates steps from a sampler for which an atom is available. An atom for a component of a hybrid sampler can be used to produce an atom for the combined sampler.

4.1 Splitting Metropolis Chains

Suppose that the distribution π we wish to sample has a density, also denoted by π, with respect to a measure μ, $\pi(dx) = \pi(x)\mu(dx)$. Hastings's (1970) version of the Metropolis algorithm originally introduced by Metropolis et al. (1953) generates the next step X_{n+1} in a Markov chain from the current state X_n by first generating a candidate step Y from a transition kernel $Q(X_n, dy) = q(X_n, y)\mu(dy)$. This candidate is accepted with probability $\alpha(X_n, Y)$, where

$$\alpha(x, y) = \min\left\{\frac{\pi(y)q(y, x)}{\pi(x)q(x, y)}, 1\right\}, \tag{4}$$

and X_{n+1} is set equal to Y. Otherwise, the candidate is rejected and X_{n+1} is set equal to X_n.

It is natural to find an atom for a Metropolis kernel by finding an atom for the subprobability transition density $q(x, y)\alpha(x, y)$; that is, by finding a pair (s', ν') such that

$$q(x, y)\alpha(x, y)\mu(dy) \ge s'(x)\nu'(dy). \tag{5}$$

Because the Metropolis kernel P satisfies $P(x, dy) \ge q(x, y)\alpha(x, y)\mu(dy)$, this provides an atom of the kernel P. In many cases there is no loss in this approach.

Proposition 1. Let (s, ν) be an atom for the Metropolis chain and assume that $\nu(\{x\}) = 0$ for all $x \in E$. Then ν has a density with respect to μ, denoted by $\nu(y)$, such that $q(x, y)\alpha(x, y) \geq s(x)\nu(y)$; that is, (s, ν) is an atom for $q(x, y)\alpha(x, y)\mu(dy)$.

For a Metropolis chain, the splitting variables S_n of Theorem 1 can be generated by allowing a split to occur only when a candidate step is accepted.

Theorem 2. Suppose that the Metropolis chain satisfies (5), and suppose that X_{n+1} and S_n are generated as follows: (1) draw X_{n+1} conditionally on X_n by taking candidate steps from $q(X_n, y)\mu(dy)$ and accepting or rejecting according to $\alpha(X_n, y)$; (2) if the candidate in step (1) is rejected, then set $S_n = 0$; otherwise, generate S_n as a Bernoulli random variable with success probability $r_A(X_n, X_{n+1})$, where

$$r_A(x, y) = \frac{s'(x)\nu'(dy)}{q(x, y)\alpha(x, y)}. \tag{6}$$

Then the process (X_n, S_n) has the same distribution as the process described in Theorem 1.

The success probability $r_A(x, y)$ is the conditional probability of a regeneration, given $X_n = x$ and $X_{n+1} = y$ and given that the candidate is accepted. In principle, it is possible to generate the splitting variables S_n directly using the success probability (3). But when $\nu'(\{x\}) > 0$ for some x, the expression for $r(x, x)$ can be complicated to derive, due to the fact that $X_n = X_{n+1}$ can occur with positive probability when the candidate step is accepted. An estimate of the regeneration rate $\pi(s)$ can be computed using $r_A(x, y)$ as $\widehat{\pi(s)} = (1/n)\sum_{i=0}^{n-1} Z_{i+1}r_A(X_i, X_{i+1})$, where the $Z_i = 1$ if the candidate for X_i is accepted and $Z_i = 0$ otherwise.

This leaves the question of finding a pair (s, ν) that satisfies (5). This can be done by finding an atom (s_q, ν_q) for the kernel Q, provided that there exists a positive function h such that

$$h(x)q(x, y) = h(y)q(y, x) \tag{7}$$

for all x and y. This condition is formally similar to a reversibility condition, but h is not required to be a probability density. Under (7), the candidate acceptance probability becomes $\alpha(x, y) = \min\{w(y)/w(x), 1\}$, where $w(x) = \pi(x)/h(x)$.

Theorem 3. Suppose that a Metropolis chain satisfies (7), and let (s_q, ν_q) be an atom for Q. For any $c > 0$, set $s'(x) = s_q(x)\min\{c/w(x), 1\}$ and $\nu'(dy) = \nu_q(dy)\min\{w(y)/c, 1\}$. Then (5) holds.

The ν' given in Theorem 3 is not a probability measure, but this does not matter in (5), as $\nu'(E)$ can be absorbed into s'.

The tour length distribution for this atom depends on the choice of the constant c. The product $s'(x)\nu'(dy) = s_q(x)\nu_q(dy)\min\{c/w(x), 1\}\min\{w(y)/c, 1\}$ will be small if c is far above or below typical values of $w(x)$. This suggests that a good choice for c will usually be in the center of the distribution of the weights $w(x)$ under π.

The most important case where condition (7) holds is an independence Metropolis chain (Tierney, in press, sec. 2.3). In an independence chain, candidates are generated from a fixed density f, regardless of the current state of the chain; thus $q(x, y) = f(y)$. Equation (7) holds for any h proportional to f. To apply Theorem 3, choose $s_q = 1$ and $\nu_q(dy) = f(y)\mu(dy)$. Independence chains are useful primarily as components in hybrid chains, as discussed in Section 4.3.

For an independence chain with the split of Theorems 2 and 3, the distribution ν' has density proportional to $f(y)\min\{w(y)/c, 1\}$. This can be sampled by rejection sampling to obtain an initial value X_0 for the chain corresponding to a regeneration. The conditional probability (6) of a regeneration at step n, given $X_n = x$, $X_{n+1} = y$ and no rejection, simplifies to

$$\begin{aligned} r_A(x, y) &= \max\{c/w(x), c/w(y)\} \quad \text{if } w(x) > c \text{ and } w(y) > c \\ &= \max\{w(x)/c, w(y)/c\} \quad \text{if } w(x) < c \text{ and } w(y) < c \\ &= 1 \quad \text{otherwise.} \end{aligned} \tag{8}$$

Thus a regeneration is certain to have occurred if $w(x)$ and $w(y)$ are on opposite sides of c.

As a special case, suppose the candidates for an independence chain are produced by a rejection algorithm with envelope function g (Tierney, in press, sec. 2.3); the function g is usually chosen to give the set $C = \{x : \pi(x) \leq g(x)\}$ high probability under π. Then the candidate generation density f is proportional to $\min\{g(x), \pi(x)\}$. Taking $h = \min\{g(x), \pi(x)\}$ in (7) and $c = 1$, the regeneration probability 4.5 becomes

$$\begin{aligned} r_A(x, y) &= 1 \quad \text{if } x \in C \text{ or } y \in C \\ &= \min\{g(x)/\pi(x, g(y)/\pi(y)\} \quad \text{otherwise.} \end{aligned}$$

The set C, where g dominates π, is a proper atom for this chain. It would be possible to base a regenerative analysis entirely on this proper atom alone, but the atom based on Theorem 2 is larger in the sense described after Theorem 1 and thus produces more regenerations.

A simple calculation shows that the regeneration rate for an independence chain using this atom can be written as $\pi(s) = (\int \min\{\sqrt{c}/w(x), 1/\sqrt{c}\}\pi(x)\mu(dx))^2/\int \pi(x)/w(x)\mu(dx)$. The effect of c on the regeneration rate or the mean tour length thus can be assessed by comparing values of $(1/n+1)\sum_{i=1}^{n} \min\{\sqrt{c}/w(X_i), 1/\sqrt{c}\}$, for a preliminary sample.

Another case where (7) holds is the original version of the Metropolis algorithm of Metropolis et al. (1953), where it is assumed that the candidate generation kernel is symmetric, $q(x, y) = q(y, x)$. In this case, Equation (7) holds with h a constant. To apply Theorem 3, an atom (s_q, ν_q) of q must be found. One approach to finding such an atom is to choose a point $\bar{x} \in E$ and a set $D \in \mathcal{E}$, usually a compact set, and to set $\nu_q(dy) = q(\bar{x}, y)1_D(y)\mu(dy)/\int_D q(\bar{x}, u)\mu(du)$ and $s_q(x) = \inf\{q(x, y)/q(\tilde{x}, y) : y \in D\}$. It is possible to start the chain with a regeneration by rejection, sampling the initial state X_0 from a density proportional to $q(\tilde{x}, y)1_D(y)$.

As an example, consider $q(x, y) \propto \exp\{-\frac{1}{2}(y - x)^T(y - x)\}$, a random walk chain with normal increments, and

let $D = \{y: |y| \le d\}$ for some $d > 0$. Then for $\tilde{x} = 0$, we have $s_q(x) = \exp\{-\frac{1}{2}x^T x - d|x|\}$. A similar approach can be used for any random walk chain based on a spherically symmetric increment distribution.

4.2 Splitting a Gibbs Sampler

Suppose that the state space E is a product of d components, $E = E_1 \times \cdots \times E_d$, an element of E is written as $x = (x_1, \ldots, x_d)$ with $x_i \in E_i$, and $\pi(x)$ is a density with respect to a product measure $\mu(dx) = \mu_1(dx_1) \times \cdots \times \mu_d(dx_d)$. Let $\pi_i(x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_d)$ denote the conditional density of the ith component given all the others. The Gibbs sampler (Gelfand and Smith 1990) starting with $X_n = x$ generates X_{n+1} by successively replacing the components of X_n by draws from the conditional distributions π_i for $i = 1, \ldots, d$. Even though examples of very strong dependence are available, experience suggests that for many problems, the dependence in the Gibbs sampler sequence drops off very quickly, often within 10 to 20 cycles. As a result, its seems reasonable that a regeneration scheme with mean tour lengths on the order of 10 to 20 should be available in these cases.

The transition kernel of the Gibbs sampler has transition density

$$p(x, y) = \pi_1(y_1 | x_2, \ldots, x_d)\pi_2(y_2 | y_1, x_3, \ldots, x_d) \cdots \pi_d(y_d | y_1, \ldots, y_{d-1}). \quad (9)$$

In some cases it may be possible to find an atom for this density by direct examination. In others, it may be possible to follow the strategy used to find an atom for the standard Metropolis transition kernel by choosing a distinguished point $\tilde{x}$ and a set $D \in \mathcal{E}$, taking $\nu(dy)$ to have density $p(\tilde{x}, y)$, and setting

$$s(x) = \inf_{y \in D} \frac{p(x, y)}{p(\tilde{x}, y)}. \quad (10)$$

In many problems the minimization required to compute $s(x)$ can take advantage of the exponential family structure often present in problems where a Gibbs sampler is used; this is the case for the first example discussed in Section 5.

Computation of $s(x)$ may also be simplified by a suitable choice of the ordering of the components. For example, the final factor in (9) does not depend on x and thus cancels from the ratio in the definition of $s(x)$ in (10). For the purposes of computing an atom, it thus is useful to place the most complicated conditional distribution last in the update sequence. In the case of two components (i.e., for $d = 2$), an atom for π_1 or π_2 thus provides an atom for the Gibbs sampler.

Once again, starting the chain with a regeneration is easy, because sampling from $\nu(dy)$ corresponds to taking one Gibbs sampler cycle starting at $\tilde{x}$.

4.3 Hybrid Samplers

If an atom of a particular kernel is not available, then it may be possible to form a hybrid (Tierney in press, sec. 2.4) with another kernel, such as an independence kernel, for which an atom is available. If P_1 and P_2 are transition kernels with invariant distribution π, then the *cycle hybrid* kernel P_1P_2 and the *mixture hybrid* kernel $\alpha P_1 + (1 - \alpha)P_2$ for $0 \le \alpha \le 1$ are also transition kernels with invariant distribution π. The cycle hybrid corresponds to alternately using P_1 and P_2 to generate a new state; for a mixture, at each step kernel P_1 is used with probability α and kernel P_2 is used with probability $1 - \alpha$. If an atom of one of the kernels in a hybrid is available, then an atom of the hybrid chain is available.

Proposition 2. Suppose that P_1 and P_2 are transition kernels with invariant distribution π and that (s, ν) is an atom for P_1. Then $(s, \nu P_2)$ is an atom for the cycle kernel P_1P_2, and $(\alpha s, \nu)$ is an atom for the mixture kernel $\alpha P_1 + (1 - \alpha)P_2$ if $0 < \alpha \le 1$.

Computing the conditional regeneration probability (3) for the hybrid sampler may be quite difficult, but it is not necessary. Instead, the splitting variables S_n can be constructed by applying the appropriate construction to the P_1 transitions alone. For a cycle, this requires using the value of the intermediate state produced by applying P_1; for a mixture, it requires using the variable that indicates whether P_1 was used.

As shown in Section 4.1, it is particularly easy to find atoms for independence chains. To introduce regeneration into a chain with kernel P, we can choose a suitable independence kernel P_1 and take $P_2 = P^m$ for some $m \ge 1$ in a mixture or a cycle. If the candidate generation density of P_1 approximates π well, then the hybrid sampler may mix faster than a sampler based only on P. Split t distributions (Geweke 1989) or the overdispersed distributions of Gelman and Rubin (1990) may be useful as candidate generation densities. Alternatively, if the transition probabilities $P(x, \cdot)$ have densities, then we can generate candidates for P_1 from $P(\tilde{x}, \cdot)$ for some reasonable initial point $\tilde{x}$. The resulting hybrid will usually not mix any faster than a pure P sampler, but it will be possible to split the chain by identifying regeneration times that roughly correspond to returns of the chain to states that can be reached in a single step from $\tilde{x}$.

5. EXAMPLES

5.1 A Hierarchical Poisson Model

One of the examples presented by Gelfand and Smith (1990) is a hierarchical Poisson model. Failures in ten pumps at a nuclear power plant are assumed to occur according to independent Poisson processes, with each pump having its own failure rate $\lambda_1, \ldots, \lambda_{10}$. The pumps were observed for periods t_i of varying lengths, and the numbers of observed failures s_i for each pump were recorded. The data were originally analyzed by Gaver and O'Muircheartaigh (1987) and are also reproduced by Tierney (in press, Table 1). Conditional on a hyperparameter β, the individual pump failure rates are assumed to be independent random variables with a gamma distribution $G(\alpha, \beta)$ with density proportional to $x^{\alpha-1}e^{-\beta x}$. The hyperparameter β has a gamma distribution $G(\gamma, \delta)$, with $\gamma = .01$ and $\delta = 1$. For the gamma exponent of the rate distribution, Gelfand and Smith (1990) used the method of moments estimator, $\alpha = 1.802$.

For the resulting posterior distribution, given β, the λ_i are independent $G(\alpha + s_i, t_i + \beta)$ random variables, and, given $\lambda_1, \ldots, \lambda_{10}$, the distribution of β is $G(\gamma + 10\alpha, \sum \lambda_i + \delta)$. For constructing an atom of the Gibbs sampler, suppose that we first generate β, then generate $\lambda_1, \ldots, \lambda_{10}$. Then the new values of β and λ_i depend only on the previous values through $\Lambda = \sum \lambda_i$, and the ratio of the Gibbs sampler transition densities started with two different values of Λ depends only on the next state through its value of β:

$$\frac{p(x, y)}{p(\tilde{x}, y)} = \left(\frac{\Lambda(x) + \delta}{\Lambda(\tilde{x}) + \delta}\right)^{\gamma+10\alpha} \exp\{(\Lambda(\tilde{x}) - \Lambda(x))\beta(y)\}.$$

In this equation x and y represent different combinations of β and λ's. To apply the approach outlined in Section 4.2, we need to choose a distinguished value $\tilde{x}$, or its corresponding value of Λ, $\tilde{\Lambda} = \Lambda(\tilde{x})$, and a set D, which need only depend on β, and compute

$$s(\Lambda) = P\{\beta \in D | \tilde{\Lambda}\} \inf_{\beta \in D} \left(\frac{\Lambda + \delta}{\tilde{\Lambda} + \delta}\right)^{\gamma+10\alpha} \exp\{(\tilde{\Lambda} - \Lambda)\beta\}.$$

For an interval $D = [d_1, d_2]$, the minimization produces

$$s(\Lambda) = P\{d_1 \le \beta \le d_2 | \tilde{\Lambda}\} \left(\frac{\Lambda + \delta}{\tilde{\Lambda} + \delta}\right)^{\gamma+10\alpha} \times \exp\{(\tilde{\Lambda} - \Lambda)d(\Lambda)\},$$

where $d(\Lambda) = d_1$ if $\Lambda < \tilde{\Lambda}$ and $d(\Lambda) = d_2$ if $\Lambda \ge \tilde{\Lambda}$. The corresponding conditional probability of a regeneration, given $X_n = x$ and $X_{n+1} = y$, is $r(x, y) = \exp\{(\tilde{\Lambda} - \Lambda(x))(d(\Lambda(x)) - \beta(y))\}$ if $d_1 \le \beta(y) \le d_2$ and $r(x, y) = 0$ otherwise.

Based on examining a short preliminary run of a Gibbs sampler, a reasonable approach to choosing the three parameters $\tilde{\Lambda}$, d_1, and d_2 of this atom is to set $\tilde{\Lambda}$ equal to 6.7, the approximate posterior mean of Λ based on the preliminary sample, and to choose d_i of the form $\tilde{\beta} \pm k\tilde{s}_\beta$, where $\tilde{\beta} = 2.35$ and $\tilde{s}_\beta = .69$ are the approximate posterior mean and standard deviation of β, again based on the preliminary sample. We chose $k = 1.1$, because this value produced the largest estimated regeneration rate. Using this atom to split a Gibbs sampler run of length 5,000 produced 1,967 complete tours, an average tour length of $\bar{N} = 2.56$, and an estimated coefficient of variation of $\widehat{\mathrm{CV}}(\bar{N}) = .03\%$. The SRQ plot for the observed regeneration times is shown in Figure 1.

As a second approach, we used an alternating sampler in which a Gibbs cycle was followed by an independence step. The independence step candidates were generated by a single Gibbs cycle starting with $\tilde{\Lambda} = 6.7$, the approximate posterior mean of Λ, and generating first β and then the λ_i. The weight function for this independence kernel is $w(x) = e^{\tilde{\Lambda}\beta(x)} \prod (t_i + \beta(x))^{-(s_i+\alpha)}$. This can be standardized to be near 1 by dividing by its value at $\tilde{\beta} = 2.35$, the estimated posterior mean of β based on the preliminary sample. By maximizing the estimated regeneration rate for the independence steps based on a preliminary Gibbs sample of 100, we chose a value of $c = 1.1$ for the constant in Theorem 3. A run of 5,000 from this alternating chain, consisting of 2,500 Gibbs cycles and 2,500 independence steps, produced 2,069 complete tours, an average tour length of $\bar{N} = 2.41$, and an estimated coefficient of variation of $\widehat{\mathrm{CV}}(\bar{N}) = .01\%$. The SRQ plot for this run is also shown in Figure 1.

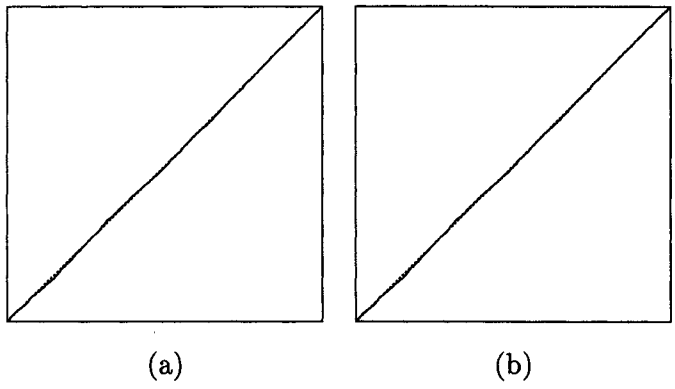

Figure 1. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Pump Failure Data Based on Runs of Length 5,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

Both the split Gibbs sampler and the alternating sampler have very small average tour lengths and uniform regeneration patterns, suggesting that the Gibbs sampler works very well in this problem. Because the independence steps used here only use a single Gibbs step to generate their candidates, they do not significantly accelerate convergence of the algorithm; but additional acceleration does not seem necessary in this case.

5.2 An Artificial Example

As an artificial example that illustrates the diagnostic value of a regenerative analysis, we considered a distribution π that is a mixture of a bivariate standard normal distribution and a bivariate standard normal distribution shifted to have its center at the point (μ, μ). The mixing probability was .5. For sufficiently large μ, the density π is bimodal with the modes displayed along the diagonal; the Gibbs sampler thus should have some difficulty in moving from one mode to the other. An alternating Gibbs/independence hybrid sampler was constructed with a single Gibbs step from the origin as the candidate generation density for the independence steps. This example is intended to model a situation where preliminary exploration has revealed one mode, the mode at the origin; but a second, and equally important, mode is in fact present at (μ, μ).

Runs of length 5,000 were performed for μ equal to 1, 3, 5, and 7. Table 1 shows the number of complete tours, the mean tour lengths $\bar{N}$, the estimated coefficients of variation $\widehat{\mathrm{CV}}(\bar{N})$, and the sample means of the two coordinates. Two estimated standard errors are given for each sample mean: a batch mean estimate based on batches of size 50, and a regenerative estimate based on (1). The means of the marginal distributions of the two coordinates under π are equal to $\mu/2$. Figure 2 shows the SRQ plots for the observed regeneration times. As expected, the performance of the sampler deteriorates as μ increases. At $\mu = 5$, there are several large gaps in the regeneration times, corresponding to periods

when the sampler is in the mode at (μ, μ). For $\mu = 7$, the sampler starts in the mode at the origin, moves to the second mode after approximately 800 observations, and returns to the mode at the origin after a total of approximately 3,800 observations. The estimated standard errors are also considerably larger for $\mu = 5$ and $\mu = 7$.

The regenerative simulation analysis clearly reveals that the sampler is not behaving well for $\mu = 5$ and $\mu = 7$. In a real example, further exploration should reveal the second mode. Incorporating this mode into a candidate generation density for independence steps in a hybrid sampler should produce a sampler with much better properties.

5.3 Splitting and the Swendsen–Wang Algorithm

Even though our work is primarily motivated by applications in Bayesian and maximum likelihood computations, the ideas can also be used in other Markov chain Monte Carlo problems. As an illustration, we show how they can be applied to the Swendsen–Wang algorithm.

Swendsen and Wang (1987) proposed a method for sampling the Potts (1952) model (Besag and Green 1993), the multicolor generalization of the Ising model. This model assumes the vertices $V = \{1, \ldots, M\}$ of a graph (V, E) are each given one of L colors, x_i. The distribution of the colors is assumed proportional to $\exp\{-\beta\eta(x)\}$, where $\eta(x)$ is the number of edges $(i, j) \in E$ for which $x_i \neq x_j$ and β is a nonnegative constant.

Given the colors x_i, the algorithm adds auxiliary *bond variables*, b_{ij}. No bonds are placed between vertices with different colors. If $x_i = x_j$ and (i, j) is an edge in the graph, then with probability $1 - \exp(-\beta)$, a bond is placed between vertices i and j, and $b_{ij} = 1$. Otherwise, no bond is placed between the vertices, and $b_{ij} = 0$. A set of bonds partitions the vertex set V into connected components. Conditional on the bonds b_{ij}, the x_i's are the same within components, and the component colors are selected independently and uniformly from the available L colors. The joint distribution of (x, b) is proportional to $e^{-\beta(|E|-\Sigma b_{ij})}(1 - e^{-\beta})^{\Sigma b_{ij}}$ on the set of (x, b) values such that $b_{ij} = 0$ whenever $x_i \neq x_j$ and is zero elsewhere; here $|E|$ is the number of edges in the graph, and sums are over edges. The algorithm is a two-coordinate Gibbs sampler that alternates between selecting bonds and colors from these conditional distributions.

It is possible to find an atom for this Gibbs sampler along the lines of Section 4.2. A natural choice for the distinguished state $\tilde{x}$ is the state where all vertices have the same color; the sampler then generates a new set of bonds as iid. Bernoulli random variables and then selects a new set of colors for the resulting components. In computing the infimum needed to find the splitting probability $s(\cdot)$ in Equation (10), taking $D = E$, it is easy to see that any difference in color produces an infimum of zero. So $s(\cdot)$ will be just the indicator of whether all vertices have the same color or not, and the split will occur each time the sampler returns to a configuration in which all states have the same color. The set of configurations that are all the same color forms a proper atom.

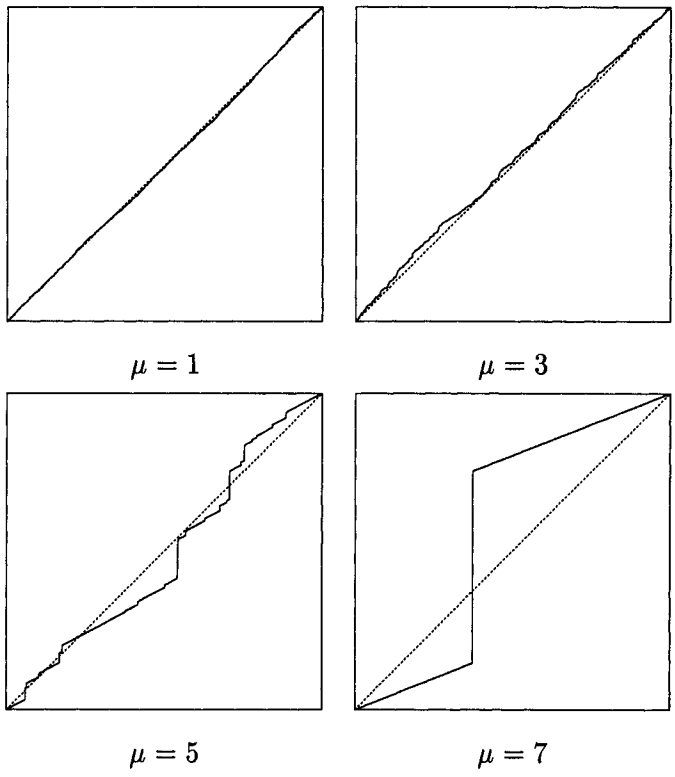

Figure 2. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Bivariate Normal Mixture Example Based on Runs of Length 5,000 From Alternating Gibbs/Independence Samplers. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

An alternating sampler using independence steps can also be constructed using Gibbs steps started at a uniform color configuration as the candidate generator. Because the conditional distribution of the colors given bonds is just uniform on the $L^{c(b)}$ possible component colorings, where $c(b)$ is the number of components, the weight function for this independence step is proportional to $L^{c(b)}$.

Both the split of the Gibbs sampler itself and the split of the alternating chain should work reasonably well if β is not too small and the graph not too large. As a simple illustration,

Table 1. Summary Statistics for Gibbs/Independence Samplers for Mixtures of Bivariate Normal Densities

μ	Tours	$\bar{N}$	$\widehat{CV}(\bar{N})$	$\bar{X}_1$	$\bar{X}_2$	$SE_B(\bar{X}_1)$	$SE_R(\bar{X}_1)$	$SE_B(\bar{X}_2)$	$SE_R(\bar{X}_2)$
1	1458	3.42	0.03%	0.516	0.493	0.018	0.019	0.020	0.020
3	1289	3.88	0.16%	1.449	1.458	0.061	0.070	0.062	0.068
5	1311	3.81	2.05%	2.374	2.355	0.218	0.375	0.219	0.381
7	988	5.06	36.52%	4.233	4.214	0.342	1.671	0.346	1.691

NOTE: Standard errors were computed using batch means (SE_B) and the regenerative method (SE_R).

we used an $m \times m$ grid with $m = 32$ and two colors, $L = 2$. The parameter β was chosen to make the bond placement probability $(1 - e^{-\beta})$ equal to .8. This corresponds to a temperature well below the freezing point of the infinite Ising lattice; a more elaborate candidate generation density would be needed for lower values of β or higher values of m. Both samplers were started with all vertices the same color.

For the pure Gibbs sampler with a split on returns to all one color, using a run of 20,000 gave 780 complete tours, an average tour length of $\bar{N} = 25.57$, and an estimated coefficient of variation of $\widehat{\mathrm{CV}}(\bar{N}) = .20\%$. Using a preliminary Gibbs sampler run of length 100, the choice of c that maximized the estimated regeneration rate for the independence split was found to be $c = e^4$. In a run of 20,000 of the alternating chain, 1,857 complete tours, a mean tour length of $\bar{N} = 10.76$, and an estimated coefficient of variation of $\widehat{\mathrm{CV}}(\bar{N}) = .05\%$ were obtained. The SRQ plots for the two sampler runs are given in Figure 3.

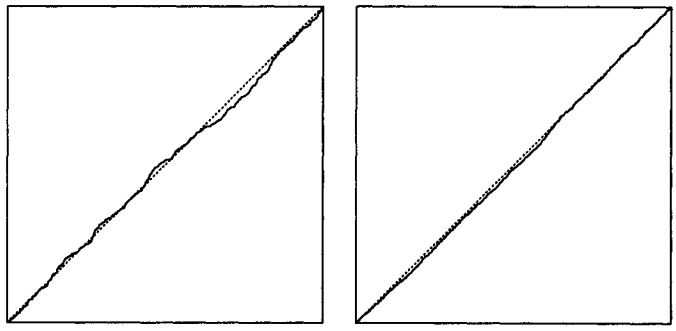

Figure 3. SRQ Plots of T_i/T_n (Vertical Axes) Against i/n (Horizontal Axes) for the Swendsen–Wang Gibbs Sampler (a) and an Alternating Gibbs/Independence Sampler (b) for the Ising Model Based on Runs of Length 20,000. Lines through the origin with unit slope are shown dashed; axis ranges are from 0 to 1 for all axes.

6. CONCLUSIONS

Markov chain splitting provides a useful way of introducing regenerations into a Markov chain simulation. This allows the use of the regenerative simulation method, which can be used to avoid initialization issues and allow variance estimates to be computed based on iid observations. In addition, it allows Markov chain sampling to take advantage of a parallel computing environment without the problems created by many short Markov chain runs when regeneration points are not available. Examining the pattern of regenerations can also give useful diagnostic information about the performance of the sampler.

It is, however, important to emphasize that regenerative simulation is essentially only a method of analysis. By itself it does not improve a sampler. A sampler that mixes very slowly will still mix very slowly even if regeneration points have been identified. The tours of such a sampler will be independent, but the slow mixing rate results in heavy tails for the tour length distribution. The regenerative approach can help reveal problems with a sampler; for example, by showing a nonuniform regeneration pattern. But, as with any method based on examining sample paths from Markov chains, absence of a problem signal does not guarantee that the sampler is working properly. In the $\mu = 7$ case of the artificial example of Section 5, the pattern of regenerations looks perfectly uniform for the first 800 iterations, up to the transition into the second mode. Had sampling been terminated before this jump, then the problem would have gone undetected.

In principle, the methods outlined in this article can be used to split samplers whenever the single step transition density is available; this is the case for most samplers proposed for exploring posterior distributions. The resulting splits will not always be satisfactory—there are many reasonable samplers for which the basic mixing rate is too slow to provide reasonable splits based on a single transition. As pointed out by Tierney (in press), it may be possible to improve the mixing rate of a sampler by forming a hybrid that uses independence chain steps based on a distribution f every m iterations. This has advantages over using multiple short runs of length m started independently from the initial distribution f, as advocated by some authors. If the initial distribution is well chosen, then most independence steps will be accepted and result in regenerations, thus producing sufficient independent tours to allow examination of diagnostics based on many short runs. In addition, by preserving the invariant distribution through the use of the Metropolis algorithm, a hybrid sampler produces a single long sample path that gets closer to equilibrium and can be averaged to produce estimates with smaller bias than many independent short sample paths. Finally, the acceptance rate for the independence steps and the rate and pattern of regenerations in a hybrid sampler may be able to detect problems with the independence candidate distribution f. The only additional cost of a hybrid algorithm over short runs started with the initial distribution f is the cost of the Metropolis accept/reject steps, which will usually be small in relation to the total sampling cost if m is of moderate size.

APPENDIX: PROOFS

Proof of Theorem 1.

The construction of (X_n, S_n) was given by Nummelin (1984, pp. 61–62). Nummelin's corollary 4.2 shows that the recurrence of X_n implies that the renewal sequence T_i is recurrent; that is, all regeneration times are finite. The expression for the mean time between regenerations was given by Nummelin (1984, p. 76).

Proof of Proposition 1.

Let $A \in \mathcal{E}$ and $x \in E$ be arbitrary, and set $B = A - \{x\}$. Because $\nu(\{x\}) = 0$,

$$s(x)\nu(A) = s(x)\nu(B) \le \int_B q(x, y)\alpha(x, y)\mu(dy)$$

$$\le \int_A q(x, y)\alpha(x, y)\mu(dy).$$

This yields the required result.

Proof of Theorem 2.

Let $P(x, dy)$ denote the Metropolis kernel, and let A_{n+1} be the event that the candidate for X_{n+1} is accepted. The conditional probability of A_{n+1}, given $X_n = x$ and $X_{n+1} = y$, is given by $P(A_{n+1} \mid X_n$

$= x, X_{n+1} = y) = q(x, y)\alpha(x, y)\mu(dy)/P(x, dy)$. Thus the conditional probability that $S_n = 1$, given $X_n = x$ and $X_{n+1} = y$, is

$$\begin{aligned} &P(S_n = 1 \mid X_n = x, X_{n+1} = y) \\ &\quad = P(\{S_n = 1\} \cap A_{n+1} \mid X_n = x, X_{n+1} = y) \\ &\quad = P\{S_n = 1 \mid X_n = x, X_{n+1} = y, A_{n+1}\} \\ &\qquad \times P(A_{n+1} \mid X_n = x, X_{n+1} = y) \\ &\quad = r_A(x, y)\frac{q(x, y)\alpha(x, y)\mu(dy)}{P(x, dy)} \\ &\quad = \frac{s'(x)\nu'(dy)}{q(x, y)\alpha(x, y)\mu(dy)}\frac{q(x, y)\alpha(x, y)\mu(dy)}{P(x, dy)} \\ &\quad = \frac{s'(x)\nu'(dy)}{P(x, dy)}, \end{aligned}$$

which is the conditional regeneration probability (3) used in the construction of Theorem 1.

Proof of Theorem 3.

It is sufficient to show that $\min\{c/w(x), 1\}\min\{w(y)/c, 1\} \le \min\{w(y)/w(x), 1\}$ for all x and y and for any $c > 0$. To see this, note that if $c/w(x) \le 1$, then

$$\begin{aligned} \min\left\{\frac{c}{w(x)}, 1\right\}\min\left\{\frac{w(y)}{c}, 1\right\} &= \frac{c}{w(x)}\min\left\{\frac{w(y)}{c} 1\right\} \\ &= \min\left\{\frac{w(y)}{w(x)}, \frac{c}{w(x)}\right\} \le \min\left\{\frac{w(y)}{w(x)}, 1\right\}, \end{aligned}$$

whereas $c/w(x) \ge 1$ implies $\min\{c/w(x), 1\}\min\{w(y)/c, 1\} = \min\{w(y)/c, 1\}$ and $\min\{w(y)/c, 1\} \le \min\{w(y)/w(x), 1\}$.

Proof of Proposition 2.

For the cycle kernel,

$$\begin{aligned} (P_1P_2)(x, dy) &= \int P_1(x, du)P_2(u, dy) \\ &\ge s(x)\int \nu(du)P_2(u, dy) = s(x)\nu(P_2)(dy); \end{aligned}$$

thus $(s, \nu P_2)$ is an atom of P_1P_2. For the mixture kernel,

$$\alpha P_1(x, dy) + (1 - \alpha)P_2(x, dy) \ge \alpha P_1(x, dy) \ge \alpha s(x)\nu(dy).$$

So $(\alpha s, \nu)$ is an atom of $\alpha P_1 + (1 - \alpha)P_2$ as long as $\alpha > 0$.

[Received November 1992. Revised May 1994.]

REFERENCES

Asmussen, S., Glynn, P. W., and Thorisson, H. (1992), "Stationarity Detection in the Initial Transient Problem," *ACM Transactions on Modeling and Computer Simulation,* 2, 130–157.

Athreya, K. B., and Ney, P. (1978), "A New Approach to the Limit Theory of Recurrent Markov Chains," *Transactions of the American Mathematical Society,* 245, 493–501.

Besag, J., and Green, P. J. (1993), "Spatial Statistics and Bayesian Computation," *Journal of the Royal Statistical Society,* Ser. B, 55, 25–38.

Bratley, P., Fox, B. L., and Schrage, L. E. (1987), *A Guide to Simulation* (2nd ed.), New York: Springer-Verlag.

Chan, K. S., and Geyer, C. J. (in press), Discussion of "Markov Chains for Exploring Posterior Distributions," by L. Tierney, *The Annals of Statistics.*

Gaver, D. P., and O'Muircheartaigh, I. G. (1987), "Robust Empirical Bayes Analysis of Event Rates," *Technometrics,* 29, 1–15.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association,* 85, 398–409.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science,* 7, 457–472.

Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrics* 57, 1317–1339.

Geyer, C. J. (1994), "On the Convergence of Monte Carlo Maximum Likelihood Calculations," *Journal of the Royal Statistical Society,* Ser. B, 56, 261–274.

Geyer, C. J., and Thompson, E. A. (1992), "Constrained Monte Carlo Maximum Likelihood for Department Data" (with discussion), *Journal of the Royal Statistical Society,* Ser. B, 54, 657–700.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D., and Kirby, A. J. (1993), "Modeling Complexity: Applications of Gibbs Sampling in Medicine," *Journal of the Royal Statistical Society,* Ser. B, 55, 39–52.

Hastings, W. K. (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika,* 57, 97–109.

Kalashnikov, V. V. (1992), "Statistical Estimates of Transient Periods by the Coupling Method," *Mathematical Methods of Statistics,* 1, 39–48.

Liu, J., Wong, W., and Kong, A. (in press), "Correlation Structure and Convergence Rate of the Gibbs Sampler With Various Scans," *Journal of the Royal Statistical Society,* Ser. B.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., and Teller, E. (1953), "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics,* 21, 1087–1091.

Nummelin, E. (1978), "A Splitting Technique for Harris Recurrent Markov Chains," *Zeitschrift für Wahrscheinlichkeitstheorie und Vervandte Gebiete,* 43, 309–318.

——— (1984), *General Irreducible Markov Chains and Non-Negative Operators,* Cambridge, U.K.: Cambridge University Press.

Potts, R. B. (1952), "Some Generalized Order-Disorder Transformations," *Proceedings of the Cambridge Philosophical Society,* 48, 106–109.

Ripley, B. D. (1987), *Stochastic Simulation,* New York: John Wiley.

Smith, A. F. M., and Roberts, G. O. (1993), "Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society,* Ser. B, 55, 3–24.

Swendsen, R. H., and Wang, J.-S. (1987), "Nonuniversal Critical Dynamics in Monte Carlo Simulations," *Physical Review Letters,* 58, 86–88.

Tanner, M. A., and Wong, W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association,* 82, 528–550.

Tierney, L. (1991), "Exploring Posterior Distributions Using Markov Chains," in *Computing Science and Statistics: 23rd Symposium on the Interface,* pp. 563–570.

——— (in press), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics.*

Yu, B. (1993), "Density Estimation in the L^∞ Norm for Department Data With Applications to the Gibbs Sampler," *The Annals of Statistics,* 21, 711–735.