The Annals of Applied Probability
2001, Vol. 11, No. 2, 438–451

PERFECT SAMPLING OF ERGODIC HARRIS CHAINS[1]

BY J. N. CORCORAN AND R. L. TWEEDIE

University of Colorado and University of Minnesota

We develop an algorithm for simulating "perfect" random samples from the invariant measure of a Harris recurrent Markov chain. The method uses backward coupling of embedded regeneration times and works most effectively for stochastically monotone chains, where paths may be sandwiched between "upper" and "lower" processes. We give an approach to finding analytic bounds on the backward coupling times in the stochastically monotone case. An application to storage models is given.

1. Introduction. There has been considerable recent work on the development and application of algorithms that will enable the simulation of the invariant measure π of a Markov chain, either exactly (that is, by drawing a random sample known to be from π) or approximately, but with computable order of accuracy. These were sparked by the seminal paper of Propp and Wilson [18], and several variations and extensions of this idea have appeared since [7, 9, 10, 12, 11, 13, 14, 16, 17]. These ideas have proved effective in areas such as statistical physics [18, 7] or spatial point processes [13, 14], where they provide simple and powerful alternatives to methods based on iterating transition laws, for example.

In this paper we develop an implementation of the Propp–Wilson algorithm which can be used when the chain is Harris recurrent, even though in standard constructions of such chains the paths of the chain may not actually meet; see [16] for an example. The key idea is to wait until all paths are in some "small set" [15] and by constructing an artificial regeneration for all paths at that time, to ensure that they do couple on an extended state space involving the regeneration time variables. This idea has been used by Murdoch and Green [17] when the whole space is small (so that the chain is uniformly ergodic), and their results can be seen as special cases of our method.

These coupling methods are very effective in implementation when the chain has some monotonicity properties, since then one needs only to construct paths from the maximal and minimal elements, using the fact that all other paths are "sandwiched" between these two. When the top and bottom paths are in the small set then one can attempt to regenerate all paths simultaneously. Many chains of interest in statistical physics or operations research satisfy such stochastic monotonicity properties, and in the absence of maximal

Received February 1999; revised August 2000.

[1]Supported in part by NSF Grants DMS-95-04561 and DMS-98-03682, and CRDF Grant RM1-226.

AMS 2000 *subject classifications*. 60J10, 60K05, 60K30.

Key words and phrases. Irreducible Markov chains, invariant measures, geometric ergodicity, backward coupling, coupling from the past, exact sampling, perfect sampling, queues, storage models.

or minimal elements it is often the case that we can construct stochastically upper and lower dominating processes to give the desired sandwiching result.

After developing the required structure in Section 2 and the general algorithm in Section 3, we show in Section 4, which is the main part of the paper, that this scheme is viable for monotone chains. In Section 5 we implement the method in practice for a storage system with no attainable minimal element and investigate the rate of convergence of the algorithm.

2. Stochastic recursive sequences and backward coupling. Formally, we consider a Markov chain on a state space X, which we assume is a separable metric space (although this restriction could be relaxed as noted on page 18 of [4]). Since we are concerned with sample path behavior, we write $\mathbf{X} = \{X_n\}_{n=0}^{\infty}$ for the version of the chain starting at x_0 (which may be fixed or random); versions starting with other initial values will be distinguished by other notation when they occur. We let $P(x, A)$ denote the transition law of the chain, and assume that there exists an invariant (or stationary) probability measure π; this satisfies

$$\pi(A) = \int_{\mathsf{X}} \pi(dx) P(x, A) \tag{1}$$

for all measurable A. The chain is called *Harris recurrent* if, for every A with $\pi(A) > 0$ we have $\mathsf{P}(X_n \in A \text{ infinitely often}) = 1 \ \forall\ x \in \mathsf{X}$. Note that for Harris recurrent chains π is unique.

Our goal is to draw values from the invariant measure π. In order to do this we first briefly describe the "stochastic recursive sequence" (SRS) coupling construction framework which forms the basis of the Propp–Wilson algorithm in [18]. This was developed in the form below by Borovkov and Foss [2, 8, 5, 3].

The SRS construction enables us to use deterministic sample path arguments which are particularly suited to a simulation environment, and more details are in [9]. We construct a probability space $(\Omega, \mathscr{F}, \mathsf{P})$, where without loss of generality $\Omega = [0, 1]^{\mathbb{Z}}$ is a doubly infinite product space and P is Lebesgue measure, so that there is an independent and identically distributed sequence $\{\xi_n\}_{n=-\infty}^{\infty}$ of uniform $U[0, 1]$ random variables given by $\xi_n(\omega) = \omega_n$ where $\omega = \{\omega_n\}$ for $\omega \in \Omega$; and there is then a measurable function $f\colon \mathsf{X} \times [0, 1] \to \mathsf{X}$ such that $\mathbf{X}$ satisfies the recursion

$$X_0 = x_0, \qquad X_{n+1} = f(X_n, \xi_n), \qquad n \geq 0 \tag{2}$$

and has transition probabilities $P(x, \cdot)$. Note that in (2) we only use $\xi_n,\ n \geq 0$; the doubly infinite construction becomes relevant in constructing backward coupling times below.

There are infinitely many such constructions to choose from and the particular construction used is largely a matter of convenience. However, one special situation worth noting immediately is when $f(x, \xi)$ is monotone in the first variable; such chains are called *stochastically monotone* and the algorithms below are particularly easy to implement in this case.

Now on $\{\Omega, \mathscr{F}, \mathsf{P}\}$, let $\{\theta^m\}_{m=-\infty}^{\infty}$ denote the family of shift transformations; that is, for any $\omega = \{\omega_n\}_{n=-\infty}^{\infty} \in \Omega$, and any $-\infty < m < \infty$, set $\theta^m\omega = \{\omega_{n+m}\}_{n=-\infty}^{\infty}$. It follows from these definitions that $\xi_{n+m} = \theta^m\xi_n$, for any m, n. We also define, for any set $B \in \mathscr{F}$, $\theta^m B = \{\theta^m\omega; \omega \in B\}$, and for any random variable $\psi\colon \Omega \to \mathsf{X}$, the shifted random variable $\psi_m := \theta^m\psi$ is defined as $\psi_m(\omega) = \psi(\theta^m\omega)$.

Using the recursive construction, for any m we introduce a "shifted" Markov chain $\theta^m\mathbf{X} = \{\theta^m X_n\}$, $n \geq 0$ as follows: if the "original" Markov chain $\mathbf{X}$ starts from x_0 at time 0 and takes the value $X_n = f(X_{n-1}, \xi_{n-1})$ at time n, then the "shifted" Markov chain $\theta^m\mathbf{X}$ starts at time m from the value $\theta^m x_0$ and takes the value $\theta^m X_n = f(\theta^m X_{n-1}, \xi_{m+n-1})$ at time $m+n$, $n = 1, 2, \ldots$.

Using these shift operators we define as in [9] the *minimal backward coupling time* $\nu(\mathbf{X})$ by

$$\nu(\mathbf{X}) = \min\{m \geq 0\colon \theta^{-n_1}X_{n_1} = \theta^{-n_2}X_{n_2}\ \forall\ n_1, n_2 \geq m\} \leq \infty. \tag{3}$$

Any integer-valued random variable $\nu \leq \infty$ is a *backward coupling time* for $\mathbf{X}$ if

$$\{\nu \leq m\} \Rightarrow \{\theta^{-n_1}X_{n_1} = \theta^{-n_2}X_{n_2}\ \forall\ n_1, n_2 \geq m\}, \tag{4}$$

and a backward coupling time is *successful* if $\nu < \infty$ almost surely.

Note that the backward coupling time is defined for shifts of the chain starting from x_0. We will see in the next section how this relates to coupling of different copies of the chain from different starting points. If one has a successful backward coupling time, then one can give a constructive approach showing that there exists a stationary version of the chain $\mathbf{X}$. We have from Theorem 3.1 of [9] the following result.

THEOREM 2.1. *Let ν be a successful backward coupling time. Put $\widetilde{X}^0 = \theta^{-\nu}X_\nu$ and define $\widetilde{X}^n = \theta^n\widetilde{X}^0$ for $n \in \mathbb{Z}$. Then the sequence $\widetilde{\mathbf{X}} = \{\widetilde{X}^n\}_{n=-\infty}^{\infty}$ forms a stationary Markov chain with transition probabilities $P(x, \cdot)$, so that in particular when the chain is Harris recurrent, $\theta^{-\nu}X_\nu \sim \pi$.*

Thus, in order to draw from π it suffices to find a backward coupling time ν and then, starting from ν, accept the value $\theta^{-\nu}X_\nu$ which is the value returned at time zero. Theorem 2.1 ensures that this is indeed a draw from π.

The idea behind the Propp–Wilson algorithm is an elegant exploitation of this structure. Suppose that we consider a family of chains $\mathbf{X}^{(x)}$, each with the same laws as $\mathbf{X}$, but with the version $\mathbf{X}^{(x)}$ starting from $x \in \mathsf{X}$. If we can find a time T such that *all* of the chains $\mathbf{X}^{(x)}$ starting, not at time zero, but at time $-T$, have the *same value* at time zero, then as is shown in Theorem 1 of [18], this common value is a perfect draw from π and such a T (called a "vertical" backward coupling time in [9]) is indeed a backward coupling time as described above.

Intuitively, it is clear why this result holds with such a random time T. For consider a chain starting at $-\infty$ with the stationary distribution π. At every iteration it maintains the distribution π. But at time $-T$ it must pick *some* value x, and from then on it follows the trajectory from that value. But

of course it arrives at the same place at time zero no matter what value x is picked at time $-T$ so the value returned by the algorithm at time zero must itself be a draw from π.

For this to be practicable we need to ensure that T is indeed finite. Propp and Wilson [18] show that this occurs for irreducible aperiodic finite space chains, and for a number of stochastically monotone chains possessing maximal and minimal elements. In what follows, we will use minorization methods to develop an implementation of the algorithm that can be applied to more general Harris chains.

3. Algorithms for Harris chains.

3.1. *Uniformly minorized chains.* We first describe a construction described in [17], the "multigamma" coupler. Suppose the chain satisfies the uniform minorization condition

$$P^k(x, \cdot) \geq \varepsilon_{\mathsf{X}}\, \varphi(\cdot), \qquad x \in \mathsf{X} \tag{5}$$

for some probability measure φ and some $0 < \varepsilon_{\mathsf{X}} \leq 1$. Such chains are "uniformly ergodic" and have a number of desirable properties ([15], Chapter 16) and in particular they are Harris chains with finite invariant measures.

When $k = 1$, one way to simulate from such a chain is to draw a sequence ε_n of i.i.d. $U[0, 1]$ random variables, and a sequence V_n of i.i.d. variables with law φ; if $\varepsilon_{n+1} \leq \varepsilon_{\mathsf{X}}$ then we set $X_{n+1} = V_{n+1}$, and if $\varepsilon_{n+1} > \varepsilon_{\mathsf{X}}$ then we choose X_{n+1} from the "residual" distribution

$$R(X_n, \cdot) = [P(X_n, \cdot) - \varepsilon_{\mathsf{X}}\varphi(\cdot)]/[1 - \varepsilon_{\mathsf{X}}]. \tag{6}$$

It is obvious that X_{n+1} has the correct marginal distribution with this construction.

Clearly one can construct an SRS representation for this chain, as given in [17], and although this is more complicated, it is clear how the joint distributions will interact: in particular, if $\varepsilon_{n+1} \leq \varepsilon_{\mathsf{X}}$ it follows that the value of X_{n+1} is V_{n+1} independent of X_n. Thus, to construct a backward coupling time with this construction, as we move back in time one step, we successively draw $U[0, 1]$ variables ε_n and set $T = \min\{n\colon \varepsilon_{-n} \leq \varepsilon_{\mathsf{X}}\}$.

Since this couples chains from all starting points, it is a vertical coupling time. Conversely, in [9] we show that when any backward coupling occurs using a vertical backward coupling time, the chain must in fact be uniformly ergodic, with (5) holding for some ε_{X}, φ and k. We remark that this result has been misinterpreted as saying that backward coupling can only hold for uniformly ergodic chains. This is incorrect. If we can construct an "upper process" $\{U_n\}$ and a "lower process" $\{L_n\}$ such that

$$U_n \geq X_n \geq L_n$$

along any sample path, for all $n \geq 0$, then at the time of coalescence of these processes we also get a successful backward coupling which may not entail uni-

form ergodicity. Subtle constructions of such processes occur in point process models [14] or for storage models [6] for models demonstrably not uniformly ergodic.

3.2. *A general Harris chain algorithm.* If the chain is not uniformly ergodic, one cannot assume that the minorization (5) holds for all x. Let us call a set C *small* if the k-step transition law $P^k(x, \cdot)$ is minorized for some $0 < \varepsilon < 1$ and for some density φ, but only on C; that is,

$$P^k(x, dy) \geq \varepsilon\varphi(dy), \qquad x \in C. \tag{7}$$

It is known that for aperiodic Harris chains, every set with $\pi(A) > 0$ contains a small set of positive π-measure [15]. In what follows we will assume that C is such a set with $k = 1$. This ensures aperiodicity also and is sometimes called the *strongly aperiodic* case.

Now consider the Nummelin splitting construction [15], in which we construct sample paths by using the transition law $P(x, \cdot)$ for $x \in C^c$, but each time $x \in C$ we draw a sequence ε_n of i.i.d. $U[0, 1]$ random variables, and a sequence V_n of i.i.d. variables with law φ. If $\varepsilon_{n+1} \leq \varepsilon$ then we set $X_{n+1} = V_{n+1}$, and if $\varepsilon_{n+1} > \varepsilon$ then we again choose X_{n+1} from the "residual" distribution $R(X_n, \cdot)$ in (6). With this structure we see that that X_{n+1} has the correct marginal distribution, as in the previous section. This suggests a method of backward coupling analogous to that in Section 3.1, as follows. Draw a doubly infinite set of i.i.d. $U[0, 1]$ variables ε_n, independent of the ξ_n. Consider again the family of chains $\mathbf{X}^{(x)}$, with $\mathbf{X}^{(x)}$ starting from $x \in \mathsf{X}$. Suppose we can find a time T such that:

1. All of the values $X_r^{(x)}$ starting at time $-T$ are simultaneously in C at a time r with $-T < r < 0$.
2. The value of $\varepsilon_r \leq \varepsilon$.

Then we can update all of the chains at $r + 1$ with a value drawn from φ; and they will then all have the same value at time zero. Thus T will be a backward coupling time and the common value is a perfect draw from π.

Møller [16] mentions that Murdoch suggests using such a method in the context of models that only get close but do not couple. Murdoch has told us (private communication) that his suggestion came from seeing a preprint of our current paper.

In carrying out this algorithm recursively backward in time, we must, as with the Propp–Wilson algorithm, take care that we reuse the same uniform random variables at each time, and we must in particular take care that if, at a time point m, we *fail* to regenerate using φ because $\varepsilon_m > \varepsilon$, then every path that enters C at time m from an earlier starting time (before $-T$) also fails to regenerate at m. In other words, the chain we must use is $Z_n = (X_n, \varepsilon_n)$, $n \geq 0$ and its shifted versions, if we are to use the coupler above. This is in contrast to the construction in the forward direction, where we need only generate values of ε_n at times when the chain is actually in C.

We note that in practice the implementation will be of value only if we can ensure that the family $\mathbf{X}^{(x)}$ all enter the small set C simultaneously (and in a reasonable time). Trivially, this will happen for the uniformly ergodic case if we take $C = \mathsf{X}$ (the Murdoch–Green algorithm). We see below that this also occurs for monotone chains under some conditions. However, if we take C to be some proper subset, it still follows that the algorithm as described above using the family $\mathbf{X}^{(x)}$ will lead to a "vertical" backward coupling time, and thus from Theorem 4.2 of [9] the chain must in fact be uniformly ergodic (although with the construction of upper and lower processes this may be avoidable).

Even for uniformly ergodic chains, however, we may gain in two ways by introducing the Harris coupling algorithm. First, in practice if the value of ε_{X} in (5) is very close to zero, we can often speed up the Murdoch–Green version by selecting a "smaller" small set, which is entered reasonably often and which has a higher success rate in regenerating through φ using (7). Second, even though vertical coupling can only be successful for uniformly ergodic chains, it does not follow that the value of k in (5) has to be $k = 1$. If (5) only holds for some large value of k, then it may be impractical to use the Murdoch–Green construction on the k-step chain and the Harris coupler we have here will be far more simple to realize. Both of these are illustrated in the example in Section 5.

4. The stochastically monotone algorithm.

4.1. *Bounding the family* $\mathbf{X}^{(x)}$. We next consider a stochastically monotone chain $\mathbf{X} = \{X_n\}$ on a linearly ordered state space X. By this we mean that there is an order on X such that $x \le y$ implies $f(x, \xi) \le f(y, \xi)$ for any realization of ξ, where f is a stochastic recursion generating $\mathbf{X}$.

When the chain is stochastically monotone with an upper and lower bounding process, as described below, we only need to ensure that these processes are in C in order to be able to regenerate using φ. This is analogous to ensuring that the upper and lower processes in the Propp–Wilson algorithm reach a single common point; at that time the "sandwiching" property of the stochastic monotonicity ensures that chains from all starting points have this same common value.

An upper and lower process pair U_n, L_n is defined formally as a pair of processes generated by sequences

$$U_0, \quad U_{n+1} = f_U(U_n, \xi_n),$$
$$L_0, \quad L_{n+1} = f_L(L_n, \xi_n)$$

in such a way that two properties hold: (1) the sandwiching property, that for any $x_l \le x \le x_u$, we have

$$f_L(x_l, \xi) \le f(x, \xi) \le f_U(x_u, \xi),$$

so that

$$\theta^{-n} L_{n-k} \le \theta^{-n} X_{n-k} \le \theta^{-n} U_{n-k}, \qquad n \ge k \ge 0 \tag{8}$$

whenever $L_0 \le X_0 \le U_0$ and (2) a "funneling property", that for all $n \ge m \ge k \ge 0$,

$$\begin{aligned} \theta^{-n} U_{n-k} &\le \theta^{-m} U_{m-k}, \\ \theta^{-n} L_{n-k} &\ge \theta^{-m} L_{m-k}, \end{aligned} \tag{9}$$

so that upper processes starting earlier lie beneath those starting later, and lower processes starting earlier are above those starting later.

If the space has maximal and minimal points $K, 0$, then U_n, L_n can be taken as the chains starting from these points; that is, we can take $f_L = f_U = f$, with $U_0 = K, L_0 = 0$. When the upper or lower process is not simply the chain itself run from a maximal or minimal point then other processes may be designed, as developed in, for example, Kendall [14] and Møller [16]. In the latter, this is exploited to show that if there is a time in the past from which the upper and lower processes are within ε at time 0, then each path (and hence in particular any stationary-start path) is sandwiched into that ε-set.

We shall see that in our case, when we introduce an adjoined coin-tossing process ε_n, more care again is needed to get the appropriate upper process and lower process for the chain Z_n even when the original chain does have maximal and minimal points.

4.2. *Constructing an upper process.* Consider the situation where the small set C is at the "bottom" of the space $\mathsf{X} = [0, K]$, since in this case we do not need to be concerned about the lower process. Let $C = [0, c]$ be a small set on which the one-step transition law $P(x, \cdot)$ is minorized so that

$$P(x, \cdot) \ge \varepsilon\varphi(\cdot), \qquad x \le c. \tag{10}$$

Note that the "residual" transition law $R(x, \cdot)$ given by (6) is stochastically monotone, from the monotonicity of $P(x, \cdot)$. However, R need not be stochastically lower than P nor need φ. We write $f_R(x, \xi)$ for a monotone (in x) function such that the distribution of $f_R(x, \xi)$ is $R(x, \cdot)$ when as usual $\xi \sim U[0, 1]$, and similarly $f_\varphi(\xi)$ as a function which generates variables with distribution φ.

Now let us consider two independent sequences of $U[0, 1]$ variables, ε_n and ξ_n, independent for $n = , \ldots, -1, 0, 1, \ldots,$ and conduct a coin toss H_n where $H_n = 0$ if $\varepsilon_n \le \varepsilon$, with $H_n = 1$ otherwise. The stochastic recursive sequence formulation of the chain using splitting [15] is then given by

$$X_{n+1} = \begin{cases} f(X_n, \xi_n), & X_n > c, \\ f_R(X_n, \xi_n), & X_n \le c, \ H_n = 1, \\ f_\varphi(\xi_n), & X_n \le c, \ H_n = 0, \end{cases} \tag{11}$$

We construct an upper process for $Z_n = \{X_n, \varepsilon_n\}$ by taking

$$U_{n+1} = \begin{cases} \max\{f(U_n, \xi_n), f_R(c, \xi_n)\}, & U_n > c, \ H_n = 1, \\ \max\{f(U_n, \xi_n), f_\varphi(\xi_n)\}, & U_n > c, \ H_n = 0, \\ f_R(U_n, \xi_n), & U_n \le c, \ H_n = 1, \\ f_\varphi(\xi_n), & U_n \le c, \ H_n = 0. \end{cases} \tag{12}$$

It is then obvious that (9) holds if $U_0 = K$, the maximal element in the space. It is also clear from (11) that for all x and all sequences ξ_n, ε_n, if $X_0 = x$ and $U_0 = K$ then $X_n \le U_n$, $n = 0, 1, \ldots$, and so (8) holds. Hence U_n is indeed an upper process.

We can now define the algorithm for generating a perfect sample from π.

MONOTONE COUPLING ALGORITHM.

1. *For each* $n = 1, 2, \ldots$, *generate independent* $U[0, 1]$ *variables* $\xi_{-n}, \varepsilon_{-n}$;
2. *For each* $n = 1, 2, \ldots$, *choose* $\theta^{-n}U_0 = K$ *and simulate the upper process* $\{\theta^{-n}U_k\}_{k=0}^{n}$.
3. *Continue until reaching*

$$T = \min\{n \ge 1\colon \theta^{-n}U_m \in C, \varepsilon_{-n+m} \le \varepsilon \text{ for some } 1 \le m \le n\} \le \infty \tag{13}$$

and let $m(T)$ *denote the smallest value of* m *in* (11).

4. *For* $s = -m(T), -m(T) + 1, \ldots, 0$ *draw* $\theta^{-m(T)}X_{m(T)+s}$ *according to* (11); *so that in particular* $\theta^{-m(T)}X_1$ *has the distribution* φ, *since by definition of* $m(T)$ *we have* $\theta^{-m(T)}X_0 \le c$ *and* $H_{-m(T)} = 0$.

For this algorithm to be successful, it is necessary to choose the functions $f(x, \xi)$ and $f_\varphi(\xi)$ in a compatible manner. One such effective implementation occurs if we choose these to be the "natural" functions corresponding to the inverse distribution functions on the linearly ordered space. We then have the following theorem.

THEOREM 4.1. *Provided a version of the natural inverse distribution functions are used, the monotone coupling algorithm is such that* T *given by* (13) *is a successful backward coupling time, and so* $\theta^{-T}X_T = \theta^{-m(T)}X_{-m(T)} \sim \pi$.

PROOF. Since the marginal stationary distribution of Z_n is π, and since we have already shown in (9) and (8) that T is a backward coupling time of Z_n by construction, it only remains to verify that $T < \infty$ with probability one. First observe that $P(c, [0, c]) > 0$, for otherwise, by stochastic monotonicity, $P(x, [0, c]) = 0$ for all $x \ge c$, and the chain is not Harris recurrent. Hence at least one of $\varphi[0, c]$ and $R(c, [0, c])$ is positive; assume without loss of generality that it is the former, so that there exists δ such that for all $\xi \le \delta$, we have $f_\varphi(\xi) \le c$ using the natural function f_φ.

Next, let N and $\xi'_0, \xi'_1, \ldots, \xi'_{N-1}$ be such that $X'_N = f(f(\cdots f(K, \xi'_0)\cdots \xi'_{N-1}) \le c$. Such a finite sequence exists since the chain is Harris recurrent. Since we are using the natural functions, it follows that for any sequence $\xi_0, \xi_1, \ldots, \xi_{N-1}$ with $\xi_j \le \xi'_j$ for all j, we also have $X_N = f(f(\cdots f(K, \xi_0)\cdots \xi_{N-1}) \le c$. Now consider the set $V_N = \Pi_0^{N-1}[0, \min[\delta, \xi'_j]] \subseteq [0, 1]^N$ and the set $B_{N+1} = [0, \varepsilon]^{N+1}$. Then the event $\{\xi_0, \xi_1, \ldots, \xi_{N-1} \in V_N, \varepsilon_0, \ldots, \varepsilon_N \in B_{N+1}\}$ has positive probability. Thus there is also positive probability η that for $N + 1$ consecutive steps we draw $H_n = 0$ and for the first N of these f_φ

is in C. By the definition of the upper process, then, starting from K we have positive probability that X_N is in C, and at that time N we get regeneration.

Hence, since θ is measure preserving, we have $\mathsf{P}(T \le N | U_0 = K) \ge \eta > 0$. But now, by construction and stochastic monotonicity, $\mathsf{P}(T \le N | U_0 = x) \ge \eta$ for any $x \le K$, and so we have $\mathsf{P}(T \le mN | U_0 = K) \ge 1 - (1 - \eta)^m$, which shows that T is successful. □

Obviously a similar proof will work for many other choices of function. However, note that if we choose the functions $f(x, \xi)$ to be natural but we perversely choose $f_\varphi(\xi)$ to be the reverse of the natural function (so that for some a and $\xi \in [a, 1]$ we generate the renewals in $[0, c]$), then the algorithm may never converge; for then (depending on the model) we could need values of ξ near 0 to drive the chain to $[0, c]$, but these same values might cause $f_\varphi(\xi)$ to be above c, and in that case the upper process would never reach $[0, c]$.

The situation is somewhat more complicated if we take $C = [a, b]$ as our small set, since we now must allow for the possibility of the upper process going below a. We do not spell out more details here, but note that there is one rather surprising drawback to the two-sided version of the algorithm: one cannot always claim that the time when the upper and lower processes reach C will be successful. In somewhat pathological circumstances one can construct chains where the choice of ξ that brings the lower process into C simultaneously drives the upper process out of C, even using natural functions. Thus care needs to be taken to ensure that the algorithm is viable when C is not at the bottom of the space.

4.3. *A simplifying assumption and convergence rates.* The algorithm is considerably simplified and more intuitive when $C = [0, c]$, and we can also take $\varphi = P(c, \cdot)$, that is, when we have

$$P(x, \cdot) \ge \varepsilon P(c, \cdot), \qquad x \in C. \tag{14}$$

The constraint (14) is a different relationship than stochastic monotonicity, but is consistent with such monotonicity and occurs in chains with, for example, appropriate asymptotically random walk behavior, such as the storage model in Section 5. It will be satisfied if the Harnack inequality

$$d^{-1} P(x, \cdot) \le P(y, \cdot) \le d P(x, \cdot)$$

is satisfied for all $x, y \in C$ and some $d > 0$. In this case,

$$R(c, \cdot) = [1 - \varepsilon]^{-1} [P(c, \cdot) - \varepsilon \varphi(\cdot)] = P(c, \cdot);$$

and so, regardless of the value of ε_n, we have that when $U_n = x \in C^c$, by monotonicity the draw of U_{n+1} is carried out according to $\max(f(x, \xi), f(c, \xi)) = f(x, \xi)$: that is, rather than constructing a new upper process U_n we just run the chain X_n itself until it reaches the set C, and only then is the value of ε_n

relevant, exactly as in the forward splitting construction. Thus in this case the upper process U_n^* is generated by

$$U_{n+1}^* = \begin{cases} f(U_n^*, \xi_n), & U_n^* > c, \\ f_R(U_n^*, \xi_n), & U_n^* \le c, \ H_n = 1, \\ f_\varphi(\xi_n), & U_n^* \le c, \ H_n = 0. \end{cases} \tag{15}$$

To differentiate from the general context we write the related backward coupling as

$$T^* = \min\{n \ge 1\colon \theta^{-n} U_m^* \in C \text{ and } \varepsilon_{-n+m} \le \varepsilon \text{ for some } 1 \le m \le n\}. \tag{16}$$

It is tempting to hope that this simplified version should work even without the special choice of φ in (14). However, it can be shown [6] that by selecting the times at which we will use ε_n as in (15), we may induce bias unless $\varphi = P(c, \cdot)$.

In [9] the rates of convergence of successful backward coupling times are related to those of forward coupling times. However, to make these connections one needs to consider the *minimal* coupling times in both directions, and even for the general monotone algorithm it is not clear when we will achieve this minimal time. For the algorithm (15), however, where the bounding process is just X_n itself started from the maximal element, we can express T^* as a minimal time. Consider the extended process

$$W_n = \{X_n, \varepsilon_n, A_n\}, \tag{17}$$

where A_n is the age process since the last regeneration in C. It is clear that T^* is the minimal backward coupling time for this process. Now let $\tilde{\tau}$ denote the forward coupling time of $\mathbf{W}$ (starting at 0) and a stationary chain $\widetilde{\mathbf{W}}$ given by (17). Since the chain is stochastically monotone, if x_0 is the maximal element of the space it follows as in (28) of [9] that

$$\mathsf{P}(\tilde{\tau} \le n) = \mathsf{P}(T^* \le n). \tag{18}$$

Thus if we can estimate the rate of convergence of the forward coupling time $\tilde{\tau}$ then this gives estimates of the rate of convergence of the backward coupling time. We illustrate this in the next section.

5. Storage models. In this section we illustrate our results for a finite storage system on $[0, K]$ with independent and identically distributed exponential replenishments with mean $1/\mu$ at the arrival times of a Poisson process with rate λ. Excessive input above $K < \infty$ is considered overflow and cannot be saved for future use. Between arrivals, content is released deterministically at rate $r(u)$. The Markov chain embedded just prior to arrival times satisfies the PASTA property [1] which ensures that its stationary distribution is identical to the stationary distribution of the continuous time chain.

We shall consider specifically the case $r(u) = \beta u$ for $\beta > 0$. With this release rule, it is easy to see that $\pi(\{0\}) = 0$, so that normal regeneration at zero does

not apply. For $x \in (0, K]$, it can be shown (Lund, private communication) that the density $\pi(x)$ of the stationary distribution π is given by

$$\pi(x) = \frac{x^{(\lambda\beta^{-1}-1)}e^{-\mu x}}{\int_0^K x^{(\lambda\beta^{-1}-1)}e^{-\mu x}\,dx} \quad \text{for } x \in (0, K]. \tag{19}$$

In general the denominator cannot be integrated analytically, and although there are obviously simple numerical approaches to its determination, this example enables us to verify the accuracy of the method. More difficult choices of r do not change the methods below and enable us then to simulate perfectly from many other models.

To execute the algorithm described in Section 4, we first simulate the interarrival times and hold these fixed over all realizations from all starting points. Once this is done we need only minorize the process jumps (replenishments) since the release is deterministic once the interarrival times are fixed. For x in the small set $C = [0, c]$, where $0 < c < K$, one possible minorization is given by

$$\mathsf{P}_x(\text{jump to } dh) \geq e^{-\mu c} \cdot e^{\mu c}\mu e^{-\mu h}\ \mathbb{1}_{\{h \geq c\}}\,dh, \tag{20}$$

where $\mathbb{1}_{\{\cdot\}}$ is the indicator function. Thus we can choose $\varepsilon = e^{-\mu c}$ and $\varphi(dh) = e^{\mu c}\mu e^{-\mu h}\mathbb{1}_{\{h \geq c\}}dh$. In this case we satisfy (14), and (15) can be implemented.

Note that, using this argument, the chain satisifies the uniform minorization 5, with $\varepsilon_{\mathsf{X}} = e^{-\mu K}$. For any large value of μK the Murdoch–Green implementation is therefore very slow, and (as we see below) the approach using regeneration in C can be very much faster.

For this choice of ε and φ, the residual jump probability density $Q(x, x+h)$ is given by

$$Q(x, x+h) = \frac{1}{1-\varepsilon}\begin{cases} \mu e^{-\mu h}, & \text{for } h < c - x, \\ \mu e^{-\mu h}(1 - e^{-\mu x}), & \text{for } h \geq c - x, \end{cases}$$

where x is the current position of the chain. The residual law R is given by jumping with Q and following the deterministic release path until the time of the next jump.

We illustrate the results when $K = 10$, $c = 1$, $\lambda = 1$, $\mu = 2$, $\beta = 1$. These parameter choices allow us to verify that the algorithm is indeed returning samples from π, as illustrated in [6]. We also assess empirically how long one must run the algorithm and compare this with analytic bounds using (18). The mean value of T^* obtained from 10,000 independent backward couplings is 9.2 embedded time steps. In other words, on average, the upper process started from $K = 10$ has reached the small set $C = [0, 1]$ and the ε-coin has turned up a head after approximately nine steps.

To calculate an analytic bound on the rate of convergence, we need to make a number of approximations. Begin by noting that if we are outside $[0, 1]$, the storage model reaches $[0, 1]$ more quickly than the model with constant unit release rate $[r(x) = 1]$, by stochastic comparison. Thus it suffices to analyze this simpler model to get bounds. Second, we note that the time to hit C from

a fixed x if $K = \infty$ bounds the time to hit C if $K < \infty$, so we can make this further simplification, even though we might expect to get rather inefficient (even if analytic) bounds with these changes.

To bound the regeneration times using (18) we need to find a Foster–Lyapunov function $V\colon \mathsf{X} \to [1, \infty)$ satisfying

$$P^*V \le \lambda_C V + b\mathbb{1}_C, \tag{21}$$

where $C = [0, 1]$, $b < \infty$, and $\lambda_C < 1$; here P^* is the law of the embedded chain of the infinite K, unit release model. We assume, without loss of generality, that $V(x) \equiv 1$ on C. In this case we have that

$$P^*(x, dy) = \frac{\lambda\mu}{\lambda+\mu} e^{-\mu(y-x)} e^{-(\lambda+\mu)d}\, dy, \qquad y > 0,$$

$$P^*(x, 0) = \frac{\mu}{\lambda+\mu} e^{-\lambda x},$$

where $d = \max(0, x - y)$. If we let $V(x) = \gamma^x$ for $x \notin C = [0, 1]$, then for $1 < \gamma < e^{\mu}$, we have that

$$\begin{aligned} P^*V(x) &= \left[\frac{\lambda\mu}{\lambda+\mu}\left(\frac{1}{\log\gamma+\lambda} + \frac{1}{\mu-\log\gamma}\right)\right]V(x) + \frac{\mu}{\lambda+\mu}\left(\frac{\log\gamma+\lambda-\lambda\gamma}{\log\gamma+\lambda}\right)e^{-\lambda(x-1)} \\ &\le \left[\frac{\lambda\mu}{\lambda+\mu}\left(\frac{1}{\log\gamma+\lambda} + \frac{1}{\mu-\log\gamma}\right) + \frac{\mu}{\lambda+\mu}\left(\frac{\log\gamma+\lambda-\lambda\gamma}{\log\gamma+\lambda}\right)\frac{1}{\gamma}\right]V(x) \\ &:= \lambda_C(\lambda, \mu, \gamma)V(x). \end{aligned}$$

When $\lambda = 1$ and $\mu = 2$, we have that $\lambda_C(1, 2, \gamma)$ achieves a minimum of $\lambda_C \approx 0.84$ at $\gamma \approx 1.88$. After similar computations for $x \in C = [0, 1]$, we get that $b \approx 2.4$.

Now if we define $\tilde{\tau}$ to be the coupling time of a chain started from K and a chain started with the stationary distribution π, then $\mathsf{E}_K(\beta^{\tilde{\tau}}) = \mathsf{E}_K(\beta^{\widetilde{T}})$, where we define $\widetilde{T}$ as the time of the first regeneration according to the distribution φ in the forward setting,

$$\widetilde{T} = \min\{n \ge 1\colon X_n \in C \text{ and } X_{n+1} \sim \varphi\}.$$

By Theorems 2.1 and 2.2 of Roberts and Tweedie [19], we have an upper bound on the generating function of $\widetilde{T}$ given by

$$\mathsf{E}_K\left[\beta^{\widetilde{T}}\right] \le \frac{\varepsilon V_K^{\phi(\beta)}}{1 - (1-\varepsilon)([\lambda_C + b - \varepsilon]/[\lambda_C(1-\varepsilon)])^{\phi(\beta)}} \quad \text{for } 1 \le \beta \le e^{\beta^*}, \tag{22}$$

where $V_K := \mathsf{E}_K[V] + \mu(C)b/\lambda_C$, $\phi(\beta) = \log\beta/\log(1/\lambda_C)$, ε is from the minorization condition, $\mu = \delta_K$ is the measure concentrated at K and

$$\beta^* = \frac{\log\left[(1-\varepsilon)^{\log\lambda_C}\right]}{\log([\lambda_C + b - \varepsilon]/[\lambda_C(1-\varepsilon)])}$$

In the unit release case, $r(x) \equiv 1$, we may use the same jump process minorization given in (20).

Since $K \notin C$, from (21) we have that

$$V_K = \mathsf{E}_K[V] = PV(K) \leq \lambda_C V(K) = (0.84)(1.88)^K,$$

and we have a bound on the generating function of $\widetilde{T}$ given by

$$\mathsf{E}_K\left(\beta^{\widetilde{T}}\right) \leq \frac{e^{-2}[(0.84)(1.88)^K]^{5.8\log\beta}}{1-(1-e^{-2})(4.3)^{5.8\log\beta}} \qquad \text{for } 1 \leq \beta \leq 1.01. \tag{23}$$

Note that due to the regenerations and residual jumps of the process, our model is not necessarily stochastically monotone for chains started in the time interval $[-T^*, 0]$ but does maintain monotonicity for chains started at earlier times. However, by an argument similar to (18) applied to the tails of $\tilde{\tau}$ and T^* one can show that $\tilde{\tau} \sim T^*$. Hence, we have that $\mathsf{E}_K(\beta^{T^*}) = \mathsf{E}_K(\beta^{\tilde{\tau}}) = \mathsf{E}_K(\beta^{\widetilde{T}})$. Therefore, we can conclude that

$$\mathsf{E}_K(\beta^{T^*}) \leq \frac{e^{-2}[(0.84)(1.88)^K]^{5.8\log\beta}}{1-(1-e^{-2})(4.3)^{5.8\log\beta}} \quad \text{for } 1 \leq \beta \leq 1.01.$$

Using Theorem 4.1 of [19], we also have

$$\begin{aligned}\mathsf{E}_K[T^*] &\leq \left[\varepsilon^{-1}\log((1-\varepsilon)^{-1}J) + \log V_K\right]/\log\lambda_C^{-1} \\ &= 5.74[10.73 + \log(0.84) + K\log(1.88)],\end{aligned}$$

where $J = 1 + (b-\varepsilon)/\lambda_C$.

For $K = 10$, $\mathsf{E}_{10}[T^*] \leq 97.4$ which can be compared with the computed means of around 9.2 for these parameters. The bounds above are clearly larger than the empirical results, but nonetheless they give us one rigorous approach to answering the question of convergence rates.

Finally, we note that the restriction on having a finite bound K can be removed by suitably constructing a random upper process. Details of one approach are in [6]; an earlier sketch of related ideas was given by D. B. Wilson (personal communication).

Acknowledgments. We are grateful to Serguei Foss and Wilfrid Kendall for valuable insights and discussions and to referees for careful reading and comments on earlier versions of the paper.

REFERENCES

[1] ASMUSSEN, S. (1987). *Applied Probability and Queues*. Wiley, New York.
[2] BOROVKOV, A. A. (1984). *Asymptotic Methods in Queueing Theory*. Wiley, New York.
[3] BOROVKOV, A. A. (1998). *Ergodicity and Stability of Stochastic Processes*. Wiley, New York.
[4] BOROVKOV, A. A. and FOSS, S. G. (1992). Stochastically recursive sequences and their generalizations. *Siberian Adv. Math.* **2** 16–81.
[5] BOROVKOV, A. A. and FOSS, S. G. (1994). Two ergodicity criteria for stochastically recursive sequences. *Acta Appl. Math*. **34** 125–134.
[6] CORCORAN, J. N. and TWEEDIE, R. L. (2000). Perfect sampling of queueing models. In *Proceedings of the 38th Allerton Conference on Computing, Control and Communication* To appear.

[7] FILL, J. A. (1998). An interruptible algorithm for perfect sampling via Markov chains. *Ann. Appl. Probab.* **8** 131–162.
[8] FOSS, S. G. (1983). On ergodicity conditions in multi-server queues. *Siberian Math. J.* **24** 168–175.
[9] FOSS, S. G. and TWEEDIE, R. L. (1998). Perfect simulation and backward coupling. *Stochastic Models* **14** 187–203.
[10] FOSS, S. G., TWEEDIE, R. L. and CORCORAN, J. (1998). Simulating the invariant measures of Markov chains using horizontal backward coupling at regeneration times. *Probab. Engrg. Inform. Sci*. **12** 303–320.
[11] HÄGGSTRÖM, O. and NELANDER, K. (1998). Exact sampling from anti-monotone systems. *Statist. Neerlandica* **52** 360–380.
[12] HÄGGSTRÖM, O., VAN LIESHOLT, M. N. M. and MØLLER, J. (1999). Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. *Bernoulli* **5** 641–658.
[13] KENDALL, W. S. (1997). On some weighted Boolean models. In *Advances in Theory and Applications of Random Sets* (D. Jeulin, ed.) 105–120. World Scientific, Singapore.
[14] KENDALL, W. S. (1998). Perfect simulation for the area-interaction point process. In *Probability Towards 2000* (L. Accardi and C. C. Heyde, eds.) 218–234. Springer, New York.
[15] MEYN, S. P. and TWEEDIE, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer, London.
[16] MØLLER, J. (1999). Perfect simulation of conditionally specified models. *J. Roy. Statist. Soc. Ser. B* **61** 251–264.
[17] MURDOCH, D. J. and GREEN, P. J. (1998). Exact sampling from a continuous state space. *Scand. J. Statist.* **25** 483–502.
[18] PROPP, J. G. and WILSON, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures Algorithms* **9** 223–252.
[19] ROBERTS, G. O. and TWEEDIE, R. L. (1999). Bounds on regeneration times and convergence rates of Markov chains. *Stochastic Process. Appl.* **80** 211–229.

DEPARTMENT OF APPLIED MATHEMATICS
UNIVERSITY OF COLORADO
BOULDER, COLORADO 80309
E-MAIL: corcoran@stat.uga.edu

DIVISION OF BIOSTATISTICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
E-MAIL: tweedie@biostat.umn.edu