ELSEVIER

Journal of Statistical Planning and Inference 104 (2002) 297–314

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Perfect sampling from independent Metropolis-Hastings chains ☆

J.N. Corcoran *, R.L. Tweedie

Department of Applied Mathematics, University of Colorado at Boulder, Boulder CO 80309-0526, USA

Received 14 January 2000; received in revised form 14 April 2001; accepted 5 June 2001

Abstract

"Perfect sampling" enables exact draws from the invariant measure of a Markov chain. We show that the independent Metropolis-Hastings chain has certain stochastic monotonicity properties that enable a perfect sampling algorithm to be implemented, at least when the candidate is overdispersed with respect to the target distribution. We prove that the algorithm has an optimal geometric convergence rate, and applications show that it may converge extreme rapidly.

MSC: 60J10; 60K05; 60K30

Keywords: Metropolis algorithms; Hastings algorithms; Markov chains; Invariant measures; Backwards coupling; Coupling from the past; Exact sampling; Perfect sampling

1. Introduction

1.1. Perfect sampling

The development of algorithms that enable "perfect" sampling of the invariant measure π of a Markov chain, following work in the seminal paper of Propp and Wilson (1996), provides an important new set of tools for simulation approaches to inference. Given the availability of recent Markov chain Monte Carlo methodology, which allows many problems of interest in Bayesian and frequentist settings to be couched in terms of such invariant measures, perfect sampling is of particular interest in a statistical context.

In this paper we show that certain monotonicity properties of the "independent" Metropolis-Hastings (IMH) scheme are such that perfect sampling is feasible. Moreover, even in practical multidimensional situations it appears that one may need only

☆ Work supported in part by NSF Grant DMS-9504561.

* Corresponding author. Dept. of Applied Mathematics, University of Colorado at Boulder, Campus Box 526, Boulder, Colorado 80309-0526, USA.

E-mail addresses: corcoran@colorado.edu (J.N. Corcoran), tweedie@biostat.umn.edu (R.L. Tweedie).

0378-3758/02/$ - see front matter
PII: S0378-3758(01)00243-9

very limited knowledge of the structure of π to implement the proposed algorithm, and that the convergence of the IMH algorithm is likely to be very rapid.

We first sketch the ideas of perfect sampling, which we give more formally in the Appendix. In the next section, we show how these can be applied to the IMH chain, and then we give a number of examples of the use of the techniques.

To describe perfect sampling, we consider a Markov chain $\{X_n, n \geqslant 0\}$ on a rather arbitrary state space X, which will typically be (a subset of) k-dimensional Euclidean space $\mathbb{R}^k$ (although it can be any separable metric space: see the Appendix). We let $P(x,A) := \mathsf{P}(X_n \in A \,|\, X_{n-1} = x)$ denote the transition law of the chain, and assume that there exists an invariant (or stationary) probability measure π for P: this satisfies

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

for all measurable sets A.

Our goal then is to draw values from the invariant measure π.

Such chains are normally considered to run from time zero, and under standard conditions (Meyn and Tweedie, 1993) the distribution of X_n tends to π as $n \to \infty$. Perfect sampling is based on the same limit result, but with the chain running from the distant past to the current time. That is, we consider moving as usual from x at time $-n$ to A at time $-n+1$ according to the probability $P(x,A)$: if we can start the chain at time $-\infty$ with the stationary distribution π, then at every iteration it maintains the distribution π, due to (1). Then the value that this scheme returns at time zero is also distributed according to π.

Indeed one can show (see the Appendix) that if the distribution at $-\infty$ is any fixed (or even random) value x_0, then under fairly standard conditions the value at time zero is still distributed according to π, and this observation is crucial in what follows.

Obviously in practice, we cannot start at time $-\infty$, with either a fixed or a random start. The essential idea of the Propp–Wilson algorithm is to find a random but finite epoch $-T$ in the past such that, if we construct sample paths from *every point* in X starting at $-T$, then all these paths will have "coalesced" successfully by time zero, so that there will be a unique value of the paths at time zero. When the chain conceptually started at $-\infty$ reaches time $-T$ it must take *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 where it starts, by the definition of T: so the value returned by the algorithm at time zero must itself be a draw from π.

More formally, and indeed in order to run this "backward coupling" algorithm, we use a recursive approach for constructing the paths of the Markov chain. Suppose at time n we draw a Uniform $[0,1]$ random variable ξ_n, and can find a measurable function $f : \mathsf{X} \times [0,1] \to \mathsf{X}$ such that we can construct the next value of the chain deterministically from the current value X_n and the value of ξ_n, using

$$X_{n+1} = f(X_n, \xi_n). \tag{2}$$

It is known (see the Appendix) that one can always find such a function f, and in applications the construction of f is usually fairly clearcut.

Now if we have an independent and identically distributed sequence $\{\xi_n\}_{n=-\infty}^{\infty}$ of Uniform[0, 1] random variables, then we can construct any segment of the chain using (2). That is, for a given (and fixed) sequence of outcomes ξ_n, $n < 0$, and any given starting point x and time $-K$, we can construct the path sequence

$$X_{-K}^{(x)} = x,\ X_{-K+1}^{(x)} = f(x, \xi_{-K}),\ X_{-K+2}^{(x)} = f(X_{-K+1}^{(x)}, \xi_{-K+1}), \ldots, X_0^{(x)}. \tag{3}$$

Thus, if we define the random time T by

$$T = \min\{K\colon \exists z\colon X_{-K}^{(x)} = x,\ X_0^{(x)} = z\ \forall x \in \mathsf{X}\},$$

so that T is the backward coalescence time when *all* of the outcomes $X_0^{(x)}$ based on the fixed sequence of ξ_n have a common value, then we have a *verifiable* algorithm for drawing from π. (Holding the sequence ξ_n fixed is often called "using the same random bits".)

For such an algorithm to be practicable we need to ensure that T is indeed finite. Propp and Wilson (1996) show that this occurs for irreducible aperiodic finite space chains, and for stochastically monotone chains as in the next section. Various extensions of this idea have appeared since the Propp–Wilson paper (Corcoran and Tweedie, 2001; Fill, 1998; Foss and Tweedie, 1998; Foss et al., 1998; Häggström and Nelander, 1998; Häggström et al., 1999; Kendall, 1998; Møller, 1999; Murdoch and Green, 1998), and perfect sampling has already proven effective in areas such as statistical physics, spatial point processes, Gibbs sampling and operations research, where it provides simple and powerful alternatives to methods based on iterating P.

1.2. Stochastic monotonicity

Probably the most difficult step in perfect sampling lies in showing that, for any non-trivial space, paths from *all* starting points have coalesced before time zero. As pointed out in Propp and Wilson (1996), backward coupling methods are much easier to implement when the space is ordered and the chain has the monotonicity property that paths from lower starting points stay below those from higher starting points. This occurs if f in (2) is monotone in the first variable.

In that case, if the state has maximal and minimal elements u and l, one can construct paths from u and l, and use the fact that all other paths are "sandwiched" between these two, so that at all times (for a particular driving sequence ξ_n) we have for any $m \geqslant 0$

$$X_{-m}^{(l)} \leqslant X_{-m}^{(x)} \leqslant X_{-m}^{(u)}, \quad x \in \mathsf{X}$$

and of course, in particular

$$X_0^{(l)} \leqslant X_0^{(x)} \leqslant X_0^{(u)}, \quad x \in \mathsf{X}.$$

It is then clear that the coalescing time for the upper and lower paths is the coalescing time for the whole set of starting points. Many chains of interest in statistical physics or operations research satisfy such stochastic monotonicity properties, and in Section 2.1, we show that with an appropriate ordering, the IMH chain is also monotone.

However, many ordered chains have no minimal and/or maximal points. To overcome this we may be able to invoke a more subtle type of construction introduced by Kendall (1998): we introduce an "upper process" U_n and a "lower process" L_n such that paths started from a fixed x_0 are always below the paths of the upper process, and above those of the lower process. If we then construct a path from the values of the upper process U_{-n}, in the sense that

$$X_{-K}^{(U)} = U_{-K},\ X_{-K+1}^{(U)} = f(U_{-K}, \xi_{-K}),\ X_{-K+2}^{(U)} = f(X_{-K+1}^{(U)}, \xi_{-K+1}), \ldots, X_0^{(U)} \qquad (4)$$

and analogously form $X_n^{(L)}$ from the lower process L_n, then again we get a verifiable algorithm by taking $-T$ to be the first time that such paths coalesce, in the sense that $X_0^{(U)} = X_0^{(L)}$.

In Section 2, we show that under certain conditions, the IMH chain will have a minimal element l and we can take $L_n \equiv l$; moreover, we can construct an upper process in this case, and thus backward coupling is a viable way to derive perfect samples.

2. Perfect sampling for independent MH chains

2.1. Monotonicity of the IMH chain

The Metropolis-Hastings algorithm (Tierney, 1994) allows simulation of a probability density $\pi(x)$ (with respect to Lebesgue measure, say, when $\mathsf{X} = \mathbb{R}^k$) which is known only up to a factor: that is, when only $\pi(x)/\pi(y)$ is known.

We will consider the so-called "independent" MH chain, where we have a given candidate distribution Q which we will assume to have a density q, positive everywhere for convenience, with which we can generate potential values of an i.i.d. sequence. A "candidate value" generated according to q is then accepted with probability $\alpha(x, y)$ given by

$$\alpha(x, y) = \begin{cases} \min\{\frac{\pi(y)}{\pi(x)}\frac{q(x)}{q(y)}, 1\} & \pi(x)q(x) > 0, \\ 1 & \pi(x)q(x) = 0. \end{cases}$$

Thus, actual transitions of the IMH chain take place according to a law P with transition density

$$p(x, y) = q(y)\alpha(x, y), \quad y \neq x$$

and with probability of remaining at the same point given by

$$P(x, \{x\}) = \int q(y)[1 - \alpha(x, y)]\mu(\mathrm{d}y),$$

where μ is Lebesgue measure. With this choice of α we have that π is invariant for P.

We now use the ratios in the acceptance probabilities $\alpha(x, y)$ to reorder the states in such a way that we always accept moves to the left (or downwards). We will show that this produces a stochastically monotone chain.

If we write $\pi(x) = kh(x)$ where k is unknown, define the IMH ordering,

$$x \succcurlyeq y \;\Leftrightarrow\; \frac{\pi(y)q(x)}{\pi(x)q(y)} \geqslant 1 \;\Leftrightarrow\; \frac{h(y)}{q(y)} \geqslant \frac{h(x)}{q(x)}. \tag{5}$$

Theorem 2.1. *If* X *is ordered according to* (5) *then the chain is stochastically monotone*: *that is, if we put* $[z,\infty) = \{w \succcurlyeq z\}$,

$$x \preccurlyeq y \Rightarrow P(x,[z,\infty)) \leqslant P(y,[z,\infty)) \tag{6}$$

for all $z \in \mathsf{X}$.

Proof. We distinguish three cases.

Case (i): $z \preccurlyeq x \preccurlyeq y$

In this case,

$$P(x,[z,\infty)) = 1 - P(x,[0,z)) = 1 - Q([0,z)) = P(y,[z,\infty)),$$

since, we always accept moves to the left, and so (6) is verified.

To prove the other cases we show that for any $z \in \mathsf{X}$, we have $p(x,z) \leqslant p(y,z)$ whenever $x \preccurlyeq y$, where p is the transition density corresponding to the law P. To see this, we note

$$\begin{aligned} p(x,z) \leqslant p(y,z) &\Leftrightarrow \alpha(x,z)q(z) \leqslant \alpha(y,z)q(z) \\ &\Leftrightarrow \min\left(1, \frac{\pi(z)q(x)}{\pi(x)q(z)}\right) \leqslant \min\left(1, \frac{\pi(z)q(y)}{\pi(y)q(z)}\right) \end{aligned} \tag{7}$$

and verify (7) in the following cases.

Case (ii): $x \preccurlyeq z \preccurlyeq y$

In this case

$$\frac{\pi(x)}{q(x)} \geqslant \frac{\pi(z)}{q(z)} \geqslant \frac{\pi(y)}{q(y)}$$

and the inequality (7) becomes

$$\frac{\pi(z)q(x)}{\pi(x)q(z)} \leqslant 1,$$

which follows from the ordering given in (5). In this case,

$$\begin{aligned} P(x,[z,\infty)) &= \int_z^\infty p(x,w)\mathrm{d}w \\ &\leqslant \int_z^\infty p(y,w)\,\mathrm{d}w + P(y,\{y\}) \\ &= P(y,[z,\infty)). \end{aligned}$$

Case (iii): $x \preccurlyeq y \preccurlyeq z$

Here

$$\frac{\pi(x)}{q(x)} \geqslant \frac{\pi(y)}{q(y)} \geqslant \frac{\pi(z)}{q(z)}$$

and the inequality (7) becomes

$$\frac{\pi(z)q(x)}{\pi(x)q(z)} \leqslant \frac{\pi(z)q(y)}{\pi(y)q(z)},$$

which holds if and only if

$$\frac{\pi(x)}{q(x)} \geqslant \frac{\pi(y)}{q(y)}.$$

This again follows from the ordering given in (5). In this case,

$$\begin{aligned} P(x,[z,\infty)) &= \int_z^\infty p(x,w)\,\mathrm{d}w \\ &\leqslant \int_z^\infty p(y,w)\,\mathrm{d}w \\ &= P(y,[z,\infty)) \end{aligned}$$

as required, and monotonicity is established. □

It is known (Mengersen and Tweedie, 1996) that the IHM chain has desirable convergence properties (and for example is uniformly ergodic) if there exists $\beta > 0$ such that

$$\frac{q(x)}{\pi(x)} \geqslant \beta \tag{8}$$

and we will call q overdispersed with respect to π if this holds. Under (8) we have that

$$\sup_{x\in\mathsf{X}} \frac{\pi(x)}{q(x)} = \frac{1}{\beta_0} < \infty.$$

If there exists a state l such that this supremum is attained, so that

$$\frac{\pi(l)}{q(l)} = \frac{1}{\beta_0} \tag{9}$$

then l is a minimal element under the IMH ordering.

Remark. As far as we know the stochastic monotonocity of the IHM chain was first observed by the second author and Gareth Roberts at the Mount Holyoke Conference on MCMC in 1994. This monotonicity has also been used in (Cai, 1997) for developing couplings of MH chains.

2.2. A backward coupling algorithm

We will assume that (8) holds and that an l satisfying (9) exists. If the infimum is not attained then our results hold essentially unchanged by introducing a state l with $q(l) = 0$ such that $\alpha(l, y) = \beta_0\pi(y)/q(y)$.

For an upper process in this context, we can simply use an i.i.d. sequence drawn with density q. It is obvious that if we take the starting point of the generic chain in the theorem in the Appendix as l, then the chain at the first step either accepts the draw from q or rejects it and stays below the upper process.

We can now define the algorithm for generating a perfect sample from π when q is overdispersed.

IMH (Backward coupling) Algorithm

1. Draw a sequence of random variables $Q_n \sim Q$ for $n = 0, -1, -2, \ldots$, and a sequence $\alpha_n \sim \text{Uniform}[0,1]$ for $n = -1, -2, \ldots$.
2. For each time $-n = -1, -2, \ldots$, start a lower path L at l, and an upper path, U at Q_{-n}, according to (4).
3. (a) For the lower path: accept a move from l to Q_{-n+1} at time $-n+1$ with probability $\alpha(l, Q_{-n+1})$, otherwise remain at state l. That is, accept the move from l to Q_{-n+1} if $\alpha_{-n} \leqslant \alpha(l, Q_{-n+1})$.
   (b) For the upper path: similarly, accept a move from Q_{-n} to Q_{-n+1} at time $-n+1$ if $\alpha_{-n} \leqslant \alpha(Q_{-n}, Q_{-n+1})$; otherwise remain at state Q_{-n}. (Note that this step is a formality and need not be actually implemented in practice; the IMH monotonicity requires that this move is accepted any time the lower path move described in (a) is accepted.)
4. Continue until T defined as the first n such that at time $-n+1$ each of these two paths accepts Q_{-n+1}.

We then have

Theorem 2.2. *The IMH algorithm produces a perfect sample X_0^l starting from X_{-T}^l, and the distribution of T is geometric with parameter β_0.*

Proof. It is clear that the stream of independent samples from q form an upper process as described above as they are proposals for the moves of the chain and moves "down" are always accepted by the IMH ordering. Also, it is not difficult to check that T is a backward coupling time in the sense of the Appendix.

Since coupling is achieved when both paths accept the next draw from q, and since by monotonicity the less likely path to accept the draw from q is that starting from l, it follows that coupling occurs at the first successful acceptance of a draw from q when the chain is in l. This is a geometric series of trials, and the probability of success is, integrating out the value drawn from q,

$$\int \alpha(l,y) q(y) \mu(\mathrm{d}y) = \int \frac{\pi(y)}{q(y)} \frac{q(l)}{\pi(l)} q(y) \mu(\mathrm{d}y) = \frac{q(l)}{\pi(l)},$$

which is β_0 as claimed. □

Note that as in this proof, when carrying out the algorithm we do not actually need to take step 3(b); when 3(a) is successful so is 3(b), and if 3(a) does not succeed we move to the next iteration.

Remarks. This result has interesting connections with other properties of the IMH chain. For example, it is known (cf. Foss and Tweedie, 1998) that a successful "vertical" backward coupling time T^* (in the sense that one can start from all initial points at the backward time $-T^*$) exists if and only if P is uniformly ergodic, so that

$$\lim_{n\to\infty} \|P^n(x,\cdot) - \pi\| = 0 \tag{10}$$

uniformly in $x \in \mathsf{X}$, where $\|\mu\| := \sup_A |\mu(A)|$ is the total variation norm. It is also known (Mengersen and Tweedie, 1996) that if β_0 is non-zero, as we have assumed, then the chain is uniformly ergodic in the sense above, with the rate of convergence in (10) being $(1-\beta_0)^n$.

Murdoch and Green (1998) use the uniform ergodicity to produce a "vertical" backward coupling time T^*, by sampling a minorizing distribution φ at each time point $-1,-2,\ldots$ and then using the minorization to accept the value from φ regardless of the current state of the chain.

But for the IMH chain, to do this using the minorization in (Mengersen and Tweedie, 1996) which would indeed achieve the rate β_0, we would need to choose φ to be π itself! Thus our algorithm, which instead samples from q at each time using the standard accept–reject ratios, achieves this same conceptual result, which can be shown to be best possible (Chen and Tweedie, 1998).

The convergence of the IMH chain in total variation norm at rate $(1-\beta_0)^n$ follows also from the detailed eigenanalysis of (Liu, 1992) and (Smith and Tierney, 1997); the general results of Foss and Tweedie for stochastically monotone chains (Foss and Tweedie, 1998) then imply that the backward coupling has this same rate, although the direct proof above is clearly more illuminating.

3. Implementation and examples

We consider three examples where the stationary distribution is known in order to evaluate performance of the IMH algorithm. In the first, we simulate a geometric distribution using a candidate transition law q such that the states need no reordering. In the second and third example, we exploit the fact that there are no added complications when simulating continuous or multivariate distributions, provided we can identify the minimal point l. Indeed this is the crux of problem and a drawback to this algorithm. We point out however that we only need the state at which the supremum of π/q is attained and not the value of the supremum $(1/\beta_0)$. This would be useful for target densitites for which the normalizing constant is unknown as in the example in Section 3.2.

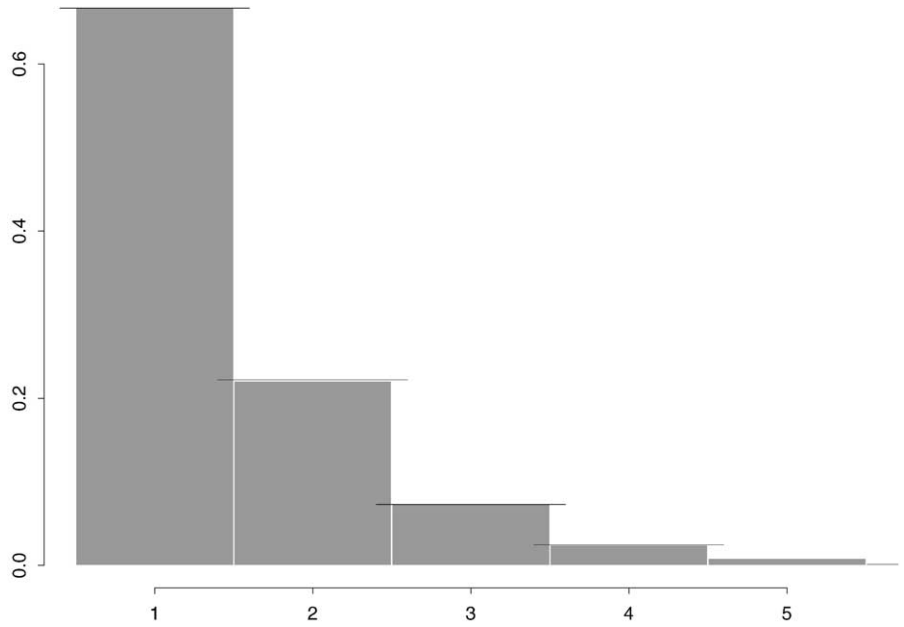


Fig. 1. 100,000 draws from $\pi(x) \propto (\frac{1}{3})^k$.

3.1. Example: a geometric density

We first apply this algorithm to simulate a simple geometric density $\pi(k) \propto h(k) = (\frac{1}{3})^k$ using the candidate density $q(k) = (\frac{1}{2})^k$ where $k = 1, 2, \ldots$

Since $h(k)/q(k) = (\frac{2}{3})^k$ is decreasing for all k, the states need no reordering, and so at each time $-1, -2, \ldots$ we start a lower path from state 1 and an upper path from a draw from q.

We carried out 100,000 draws from π (that is, the backward coupling time $-T$ was simulated 100,000 times and each time the simulated sample paths were followed forward from $-T$ to their common value at time zero). This gives the histogram in Fig. 1. The horizontal lines drawn at the tops of the bars indicate the true values of π.

All estimates are subject to sampling variability but are within two standard errors of the true values.

This trivial example can be used to illustrate a small speedup mechanism. If $q(l) > 0$, then when we have a path with $U_0 = l$, we know by monotonicity that the draw at time zero will be l; hence we can set $T = 0$ with no further calculation. Similarly, if we have any $U_{-n+1} = 0$, we do not need to check at time $-n$ for coupling.

This means that we can have $E(T) = 0q(l) + (1 - q(l))\beta_0$ rather than β_0. In this example $\beta_0 = \frac{3}{4}$ and so we reduce $E(T)$ from $\frac{4}{3}$ to $\frac{2}{3}$. In the simulation example the mean backward coupling time was indeed 0.666 with a minimum and maximum of 0 and 8, respectively.

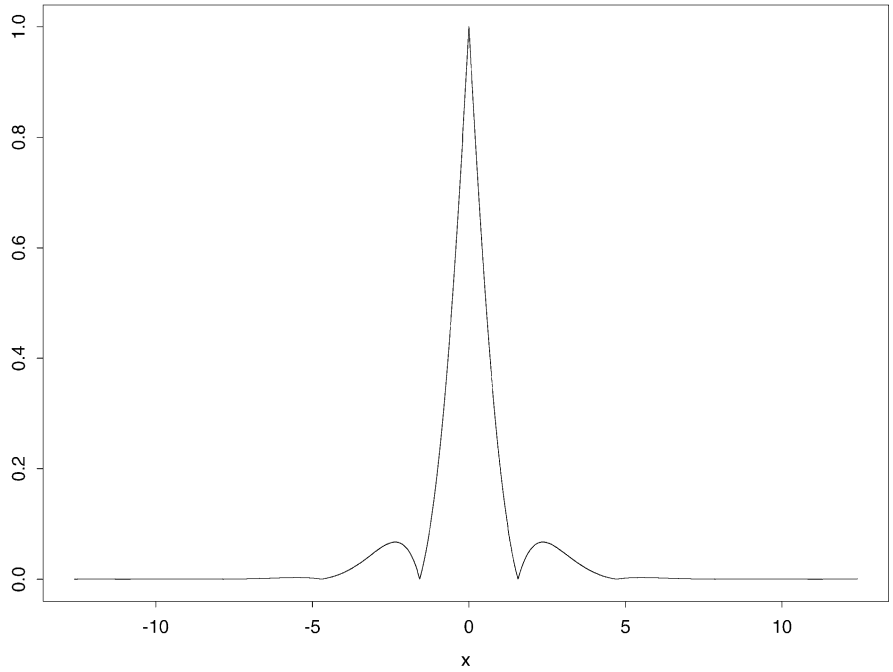


Fig. 2. Unnormalized target density.

3.2. Example: a continuous density

We next simulate from a continuous density for which we cannot find the normalizing constant,

$$\pi(x) \propto h(x) = |\cos(x)|e^{-|x|}, \quad -\infty < x < \infty,$$

using a normal candidate density with mean 0 and variance 10. This candidate was chosen to be heavier in the tails than π so that the minimal state could be easily found. Fig. 2 shows the graph of the unnormalized target density.

Note that we do not need to explicitly reorder the states when running the algorithm, nor do we need to calculate the value of the maximum. We merely need to find this minimal state, in this case, 0.

We ran the IMH algorithm by starting, at each negative integer time point, a path from 0 and a path drawn from q. Fig. 3 and shows the resulting histogram of simulated values drawn from π and a histogram of the corresponding values of the backward coupling time T which is known to have a geometric distribution with parameter β_0. In this example, $\beta_0 = q(l)/\pi(l)$ is unknown since the normalizing constant for π is unknown. Since β_0 is the parameter for the geometric distribution shown in the inset of Fig. 3, we can estimate it to be 0.183. Hence, since $q(l) = q(0) = 1/\sqrt{20\pi}$, we

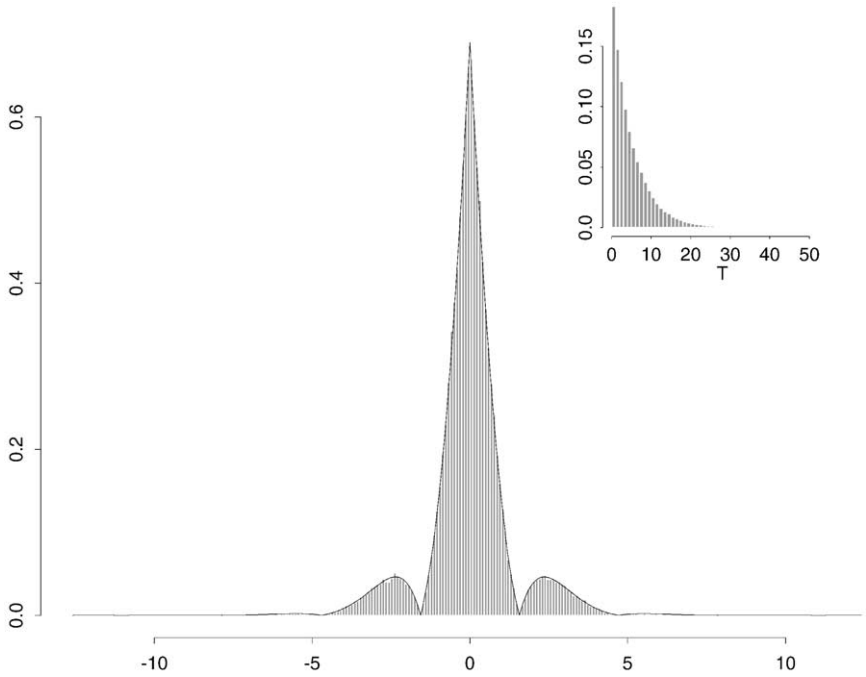


Fig. 3. 100,000 draws from $\pi(x) \propto |\cos(x)|\, e^{-|x|}$. (Inset: Histogram of the corresponding values of the backward coupling time T.)

estimate the normalizing constant to be $1/0.183\sqrt{20\pi} \approx 0.6893$. Over the histogram for π in Fig. 3 we show this estimated density.

The mean backward coupling time in 100,000 draws was 5.53 with a minimum backward coupling time of 1 and a maximum of 55.

3.3. Example: a bivariate normal

The backward coupling IMH algorithm may prove to be particularly efficient for higher dimensional spaces when one can find a minimal starting point in the sense of (9). In this section, we simulate a bivariate normal density with

$$\vec{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & 1/\sqrt{2} \\ 1/\sqrt{2} & 1 \end{pmatrix}.$$

In this case, $\pi(x, y) \propto h(x, y) = \exp\{-x^2 + \sqrt{2}xy - y^2\}$. Again the candidate transition density is chosen to be heavier in the tails than π: we use

$$q(x, y) = (\tfrac{1}{2}e^{x}\mathbb{1}_{\{x<0\}} + \tfrac{1}{2}e^{-x}\mathbb{1}_{\{x\geqslant 0\}})(\tfrac{1}{2}e^{y}\mathbb{1}_{\{y<0\}} + \tfrac{1}{2}e^{-y}\mathbb{1}_{\{y\geqslant 0\}}).$$

This is simple to use since we need only simulate independent exponentials to find the candidate values.

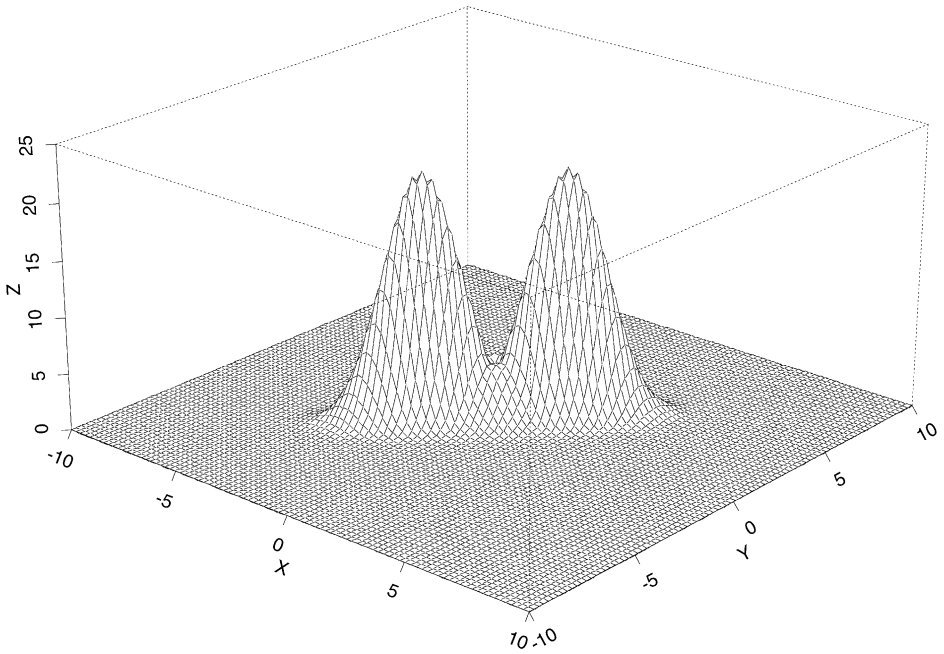


Fig. 4. Finding the modes of h/Q: normal example.

In this case h/Q attains a global maximum twice; at $(2+\sqrt{2}/2, 2+\sqrt{2}/2)$ and at $(-2-\sqrt{2}/2, -2-\sqrt{2}/2)$. Fig. 4 shows this bimodal ratio. The fact that there are two "minimal" points is immaterial to the theory: we simulated using $l=(2+\sqrt{2}/2, 2+\sqrt{2}/2)$.

Figs. 5(a) and (b) show a three-dimensional histogram of values for 100,000 draws from π and the true density, respectively. Figs. 6(a) and (b) show the same histogram and true density from a different perspective. It is clear that the perfect sampling is indeed covering the entire space, regardless of the arbitrary choice of l.

Fig. 7 shows the histogram of backward coupling times. In this example, $\beta_0 = q(l)/\pi(l) \approx 0.20$. The mean backward coupling time in 100,000 draws was 4.9 with a minimum backward coupling time of 1 and a maximum of 48.

The proportion of values drawn from π were also checked against the true mass in various regions in the (x, y)-plane. In particular, for

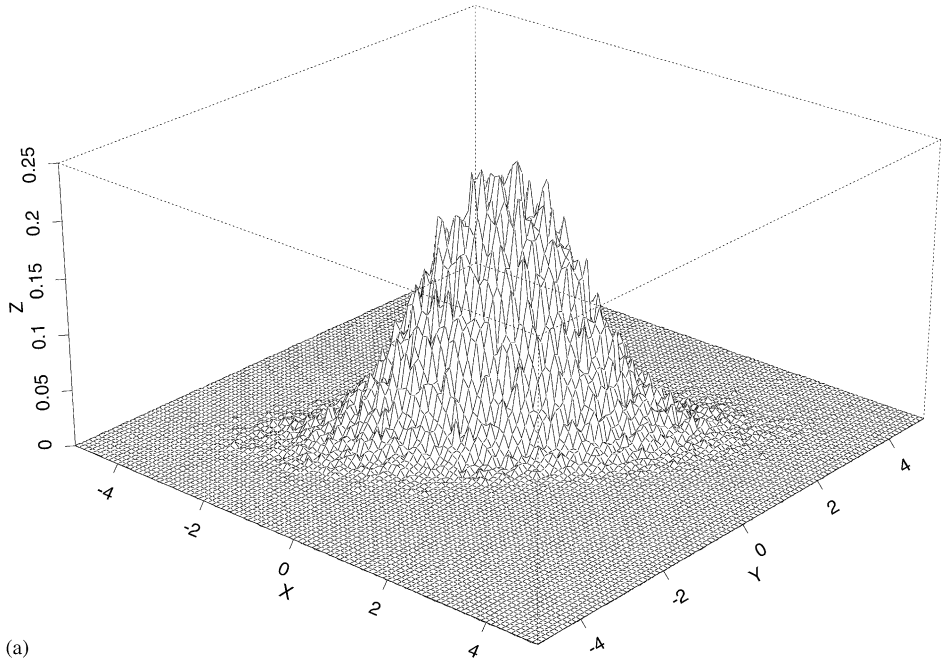
$$Q_{\mathrm{I}} = \{(x, y): x > 0, y > 0\}$$

$$Q_{\mathrm{II}} = \{(x, y): x < 0, y > 0\}$$

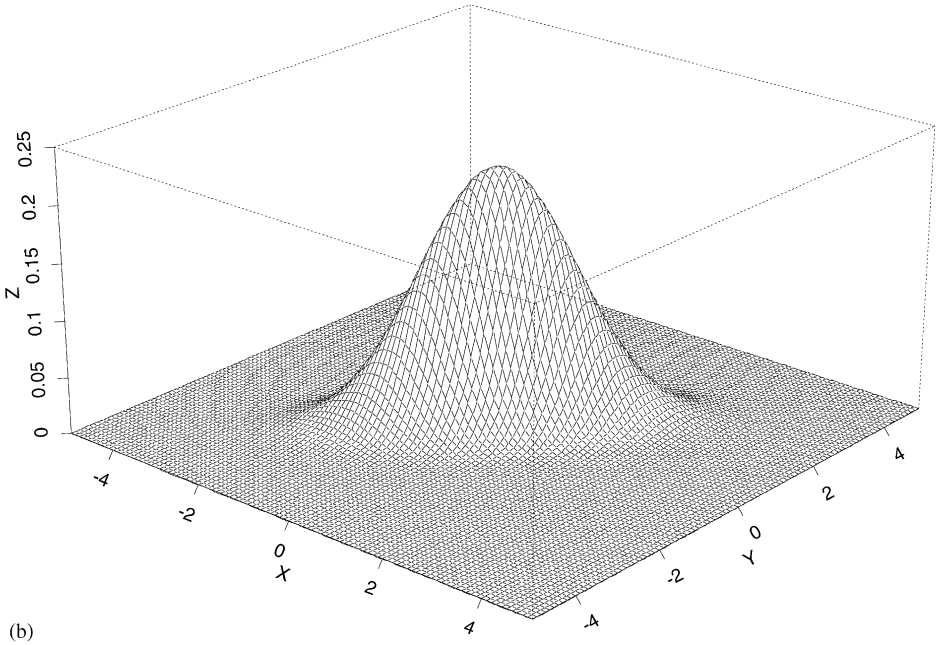
$$Q_{\mathrm{III}} = \{(x, y): x < 0, y < 0\}$$

$$Q_{\mathrm{IV}} = \{(x, y): x > 0, y < 0\}$$

we have the following comparisons.

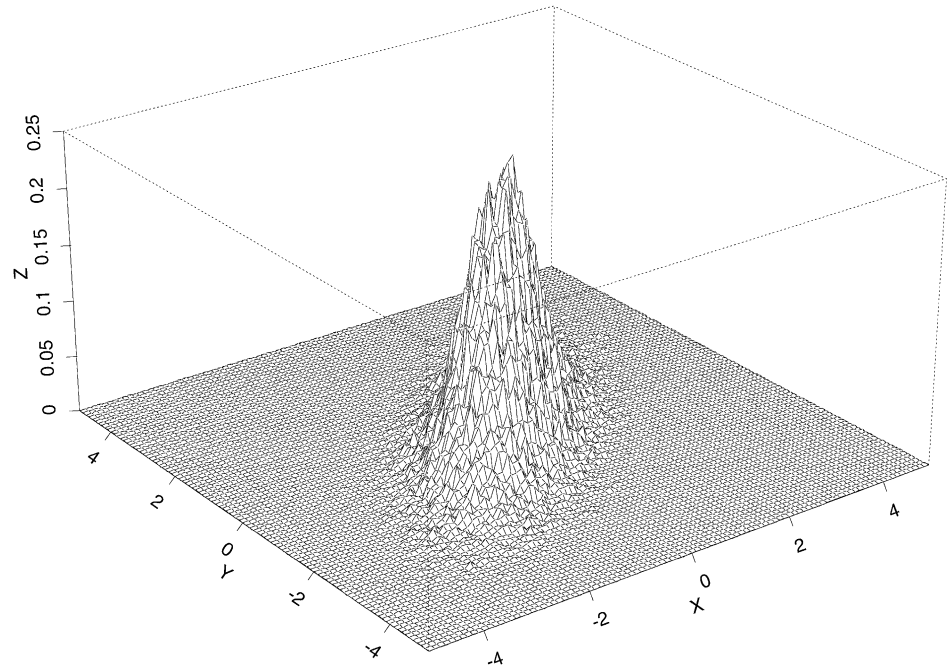


(a)

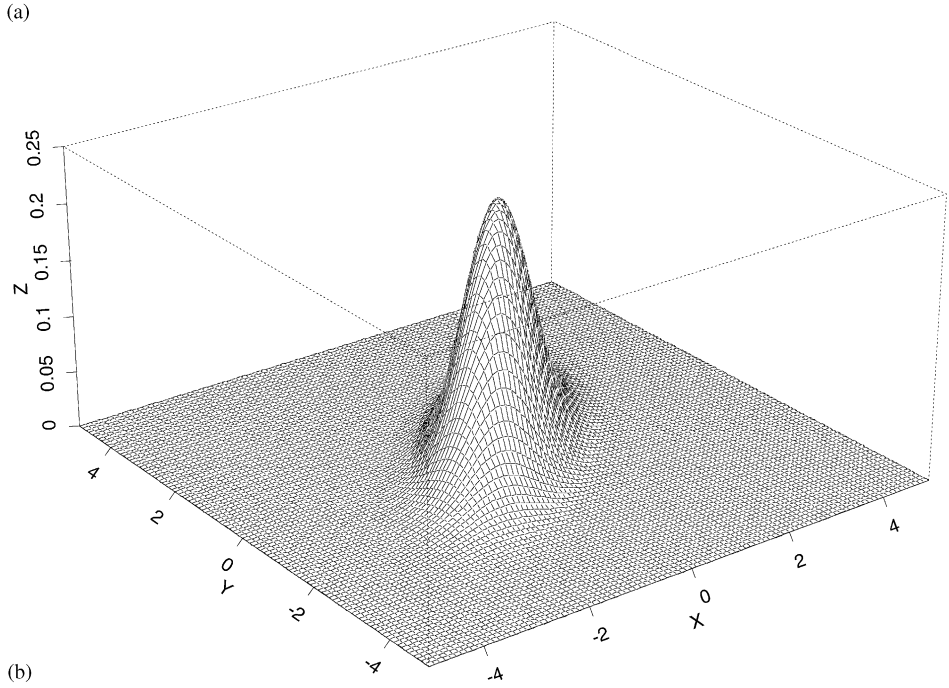


(b)

Fig. 5. (a) 100,000 draws from $\pi(x, y) \propto \exp\{-x^2 + \sqrt{2}xy - y^2\}$. (b) $\pi(x, y) = 1/\sqrt{2}\pi \exp\{-x^2 + \sqrt{2}xy - y^2\}$.



(a)



(b)

Fig. 6. (a) 100,000 draws from $\pi(x, y) \propto \exp\{-x^2 + \sqrt{2}xy - y^2\}$. (b) $\pi(x, y) = 1/\sqrt{2}\pi \exp\{-x^2 + \sqrt{2}xy - y^2\}$.

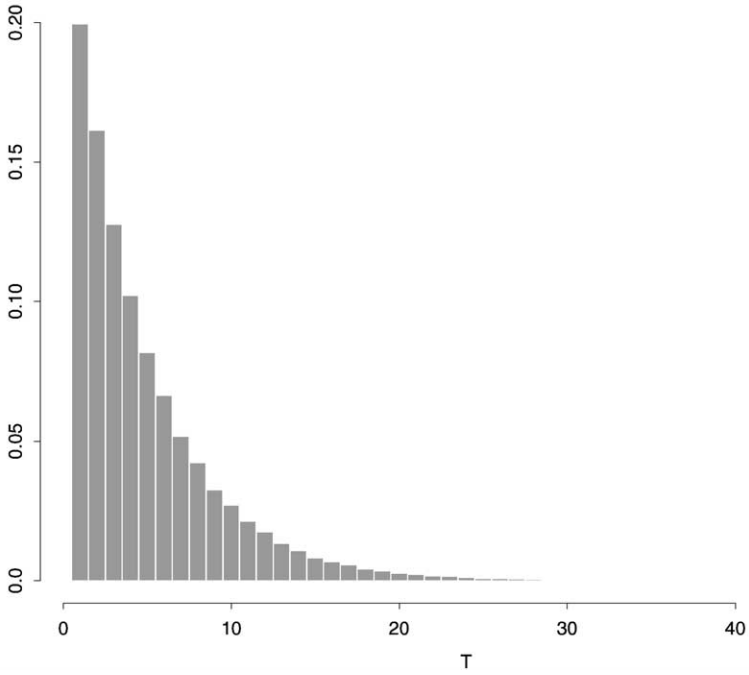


Fig. 7. Histogram of simulated T: normal example.

	$\pi(Q_{\mathrm{I}})$	$\pi(Q_{\mathrm{II}})$	$\pi(Q_{\mathrm{III}})$	$\pi(Q_{\mathrm{IV}})$
Simulated	0.37349	0.12623	0.37481	0.12547
True	0.37500	0.12500	0.37500	0.12500

The simulated values are within two standard errors of the true values for 100,000 draws.

Appendix A. Stochastic recursive sequences and backward coupling

To make the paper self-contained, we give a more formal description of the path structure required to construct successful backward couplings.

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 p. 18 of Borovkov and Foss (1992). The chain has transition law $P(x,A)$ and invariant measure π.

Since we are concerned with sample path behaviour, we write $\mathbf{X} = \{X_n\}_{n=0}^{\infty}$ for the version of the chain starting at one given x_0 (which may be fixed or random), and omit x_0 from the notation.

The stochastic recursive sequence (SRS) coupling construction framework forms the basis of the Propp–Wilson algorithm in Propp and Wilson (1996). The form below is based on work by Borovkov and Foss (Borovkov, 1984; Foss, 1983; Borovkov and Foss, 1994), although they did not develop it for algorithmic use. The SRS construction enables us to use deterministic sample path arguments which are particularly suited to a simulation environment, and more details are in (Foss and Tweedie, 1998).

The SRS approach is based on the fact that one can construct a probability space $(\Omega, \mathscr{F}, \mathsf{P})$, an independent and identically distributed sequence $\{\xi_n\}_{n=-\infty}^{\infty}$ of Uniform $[0,1]$ random variables, and a measurable function $f : \mathsf{X} \times [0,1] \to \mathsf{X}$ such that $\mathbf{X}$ satisfies the recursion

$$X_0 = x_0, \quad X_{n+1} = f(X_n, \xi_n), \quad n \geqslant 0 \tag{11}$$

and has transition probabilities $P(x, \cdot)$. Note that in (12) we only use ξ_n, $n \geqslant 0$; the doubly-infinite construction becomes relevant in constructing backward coupling times below.

There are infinitely many such constructions to choose from, depending on, for example, the joint distributions selected for $f(x, \xi_1)$ and $f(y, \xi_1)$ for different x and y. The particular construction used is largely a matter of convenience. However, one special situation worth noting is when $f(x, \xi_1)$ is monotone in the first variable: such chains are called *stochastically monotone* and backward coupling algorithms are particularly easy to implement in this case, as we have seen.

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 $\psi : \Omega \to \mathsf{X}$, 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 $\theta^m \mathbf{X} = \{\theta^m X_n\}$, $n \geqslant 0$. This has the following interpretation: the "original" Markov chain $\mathbf{X}$ starts from x_0 at time 0 and takes the value X_n at time n; the "shifted" Markov chain $\theta^m \mathbf{X}$ starts at time m from the value $\theta^m x_0$, and takes the value $\theta^m X_n$ at time $m+n$, $n = 1, 2, \ldots$. Note that since θ^m is measure-preserving, $\mathbf{X}$ and $\theta^m \mathbf{X}$ coincide in law.

Using shift operators we define as in Foss and Tweedie (1998) the *minimal backward coupling time* $v(\mathbf{X})$ by

$$v(\mathbf{X}) = \min\{m \geqslant 0 : \theta^{-n_1} X_{n_1} = \theta^{-n_2} X_{n_2} \ \forall n_1, n_2 \geqslant m\} \leqslant \infty. \tag{12}$$

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

$$\{v \leqslant m\} \Rightarrow \{\theta^{-n_1} X_{n_1} = \theta^{-n_2} X_{n_2} \ \forall n_1, n_2 \geqslant m\} \tag{13}$$

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

Note that the backward coupling time is defined for the chains starting from x_0 at all times in the past. Clearly, if v is a backward coupling time and v' is any integer-valued random variable such that $v' \geqslant v$ almost surely then v' is also a backward coupling time.

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 (Foss and Tweedie, 1998) the following:

Theorem A. *Let v be a successful backward coupling time. Put $\tilde{X}^0 = \theta^{-v} X_v$ and define $\tilde{X}^n = \theta^n \tilde{X}^0$ for $n \in \mathbb{Z}$. Then the sequence $\tilde{\mathbf{X}} = \{\tilde{X}^n\}_{n=-\infty}^{\infty}$ forms a stationary Markov chain with transition probabilities $P(x, \cdot)$, and for any n, $\tilde{\mathbf{X}}$ satisfies the recursion*

$$\tilde{X}^{n+1} = f(\tilde{X}^n, \xi_n) \tag{14}$$

for each $n \in \mathbb{Z}$ almost surely.

Thus, in order to draw from π it suffices to find a backward coupling time v and then, starting from v, accept the value $\theta^{-v} X_v$ which is the value returned at time zero: Theorem A 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 each version $\mathbf{X}^{(x)}$ starting from a fixed $x \in \mathsf{X}$. 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 Propp and Wilson (1996), this common value is a perfect draw from π: and such a T (called a "vertical" backward coupling time in Foss and Tweedie (1998) is indeed a backward coupling time as described above, since all chains starting at x_0 from times earlier than $-T$ must pass through at least one value at time $-T$, and thus return the common value at time zero.

References

Borovkov, A., 1984. Asymptotic Methods in Queueing Theory. Wiley, Chichester, New York, Toronto.

Borovkov, A., Foss, S., 1992. Stochastically recursive sequences and their generalizations. Siberian Adv. Math. 2, 16–81.

Borovkov, A., Foss, S., 1994. Two ergodicity criteria for stochastically recursive sequences. Acta Appl. Math. 34, 125–134.

Cai, H., 1997. Monotone Metropolis-Hastings Markov chains in exact simulation. (Preprint at http://www.stats.bris.ac.uk/MCMC/).

Chen, P., Tweedie, R., 1998. Convergence of extended independent Hastings-Metropolis algorithms, in preparation.

Corcoran, J., Tweedie, R., 2001. Perfect sampling of Harris recurrent Markov chains. Ann. Appl. Probab. 11 (2).

Fill, J. 1998. An interruptible algorithm for perfect sampling via Markov chains. Ann. Appl. Probab. 8 (1), 131–162.

Foss, S., 1983. On ergodicity conditions in multi-server queues. Siberian Math. J. 24, 168–175.

Foss, S., Tweedie, R., 1998. Perfect simulation and backward coupling. Stochastic Models 14, 187–203.

Foss, S., Tweedie, R., Corcoran, J., 1998. Simulating the invariant measures of Markov chains using horizontal backward coupling at regeneration times. Probab. Eng. Inform. Sci. 12, 303–320.

Häggström, O., Nelander, K., 1998. Exact sampling from anti-monotone systems. Statist. Neerlandica 52, 360–380.

Häggström, O., van Liesholt, M., Møller, J., 1999. Characterisation results and Markov chain Monte Carlo algorithms including exact simulation for some spatial point processes. Bernoulli 5 (4), 641–658.

Kendall, W., 1998. Perfect simulation for the area-interaction point process. In: Heyde, C., Accardi, L. (Eds.), Probability Towards 2000. Springer, Berlin, pp. 218–234.

Liu, J., 1992. Eigenanalysis for a Metropolis sampling scheme with comparisons to rejection sampling and importance resampling. Research Report R-427, Department of Statistics, Harvard University.

Mengersen, K., Tweedie, R., 1996. Rates of convergence of the Hastings and Metropolis algorithms. Ann. Statist. 24, 101–121.

Meyn, S.P., Tweedie, R.L., 1993. Markov Chains and Stochastic Stability. Springer, London.

Møller, J., 1999. Perfect simulation of conditionally specified models. J. Roy. Statist. Soc. Ser. B 61(1), 251–264.

Murdoch, D., Green, P., 1998. Exact sampling from a continuous state space. Scand. J. Statist. 25(3), 483–502.

Propp, J., Wilson, D., 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. Random Struct. Algorithms 9, 223–252.

Smith, R., Tierney, L., 1997. Exact transition probabilities for the independence Metropolis sampler (Preprint at http://www.stats.bris.ac.uk/MCMC/).

Tierney, L., 1994. Markov chains for exploring posterior distributions (with discussion). Ann. Statist. 22, 1701–1762.