Exact Sampling, Regeneration and Minorization Conditions.

Blanchet, J., and Meng, X.

June, 2005.

Abstract

Suppose that $X = (X_n : n \geq 0)$ is a positive recurrent Harris chain. We investigate the connection between various types of drift and minorization conditions on X and exact simulation. We present an exact simulation algorithm that can be implemented if three elements are available: 1) We can identify the regeneration epochs of X, 2) We can find a positive lower bound for the probability of regeneration in one step, and 3) We can find an upper bound for moments of order larger than 1 for the regeneration time of X. We also discuss explicit minorization and drift conditions that provide the bounds required in item 3).

1 Introduction

Suppose that $X = (X_n : n \geq 0)$ is a positive recurrent Harris chain living in a polish space $(\mathcal{X}, \mathcal{B})$. More precisely, define $T_C = \min\{n \geq 0 : X_n \in C\}$. We shall first assume that X satisfies the following two conditions:

1. $P_x (T_C < \infty) = 1$ for $x \notin C$ and

2. There is $\lambda > 0$ and a probability measure $v(\cdot)$ such that, for some $k_0 \in \mathbb{N}$

$$P^{k_0}(x, A) \geq \lambda v(A),$$

for all $A \in \mathcal{B}$ and $x \in C$.

We concentrate on the case $k_0 = 1$. Throughout this paper we shall use the notation $E_\mu(\cdot)$ for the expectation operator induced by the chain X given that X_0 has initial distribution $\mu(\cdot)$ that is

$$E_\mu(\cdot) = \int E(\cdot | X_0 = x)\, \mu(dx).$$

In this paper we present a simulation algorithm that allows to generate exact samples from the stationary distribution of X. Exact (or perfect) sampling has been the subject of much research during the last 10 years after the introduction of the so-called CFTP (coupling-from-the-past) protocol by Propp and Wilson (1996). It is known (see Foss and Tweedie (1998) that CFTP can be used only if the chain of interest is geometrically ergodic (which essentially means that convergence to stationarity occurs at an exponential rate, see Meyn and Tweedie (1993)). Other exact simulation algorithms such as dominated CFTP, introduced by in Kendall (1998) take advantage of CFTP-type ideas combined with suitable dominating processes (see also Corcoran and Tweedie (2001) and Huber (2004)).

The algorithm that we present has the following key features.

i) It takes advantage of the regenerative structure of the chain induced by conditions 1 and 2.

ii) It can be applied even when the chain is not geometrically ergodic.

iii) The structure of the algorithm is based on suitable minorization or drift conditions for the moments of regeneration times, thereby connecting exact sampling with drift conditions usually developed to analyze the rate of convergence to stationarity.

The simulation algorithm is that we present here is based on a general procedure suggested by Asmussen, Glynn and Thorisson (1992). In their paper, the authors proposed a method to generate exact samples from the stationary distribution of a regenerative process. The idea is based on the following observation. If $Z = (Z_n : n \geq 0)$ is a non-delayed regenerative process with stationary distribution $\pi(\cdot)$ and τ is the time to the first regeneration. Then

$$\begin{aligned} \pi(A) &= E\left(\sum_{n=0}^{\tau-1} 1(Z_n \in A)\right) / E\tau \\ &= E\left(\sum_{n=0}^{\infty} 1(Z_n \in A; \tau > n)\right) / E\tau = E\left(P\left(Z_T \in A \middle| \tau > T\right)\right), \quad (1) \end{aligned}$$

where $P(T = n) = P(\tau > n) / E\tau$ and T is independent of the chain Z. The probabilistic representation (1) was used by Asmussen et al (1992) to develop an exact simulation algorithm to sample from $\pi(\cdot)$ assuming that one is able to sample from T. Asmussen et al (1992) discuss some important classes of regenerative processes for which it is feasible to generate samples from T. One such class of examples is given by so-called Doeblin chains, which are Harris chains for which conditions 1. and 2. hold with $C = X$. We shall explain in Section 2 (see Algorithm 1) how to use (1) for exact simulation.

In a recent paper, Hobert and Roberts (2004) also took advantage of (1) to develop a MCMC that can be implemented in the context of Doeblin chains (the general procedure is very similar to that described by Asmussen et al (1992), which we also describe here for completeness in Algorithm 1 of Section 2). Hobert and Roberts (2004) also showed how, in the Doeblin case, the exact simulation procedure described in Algorithm 1 coincides with the algorithm Gamma coupling proposed by Murdoch and Green (1998). The use of (1) for exact simulation as explained in Algorithm 1 is relatively straightforward in the context of Doeblin chains because in this case $P(T = k) = \lambda(1-\lambda)^k$ (i.e. T is geometrically distributed) and therefore it is very easy to sample from the distribution of T. However, for more general Harris chains, sampling from T seems to be difficult because in principle it involves being able to compute the distribution of the regeneration time τ (which is also geometric in the Doeblin case). In this paper we discuss a technique that overcomes the need for computing the distribution of τ.

In recent work, Kendall (2004) explored the connection between minorization and drift conditions for geometrically ergodic chains and perfect sampling. This paper relates to Kendall (2004) in that we also provide connections between minorization and drift conditions and their implications in perfect sampling, but our investigations go beyond the case of geometrically ergodic chains.

In Section 2 we explain the general principle of our algorithm and in Section 3 we provide the necessary drift and minorization conditions required to implement our algorithm.

2 A Regeneration Algorithm

Suppose that $X = (X_n : n \geq 0)$ is a positive recurrent Harris chain living in a polish space $(\mathcal{X}, \mathcal{B})$. More precisely, define $T_C = \min\{n \geq 0 : X_n \in C\}$. Then, X satisfies the following two conditions:

1. $P_x (T_C < \infty) = 1$ for $x \notin C$ and

2. there is $\lambda > 0$ and a probability measure $v(\cdot)$ such that

$$P(x, A) \geq \lambda v(A),$$

for all $A \in \mathcal{B}$ and $x \in C$ (this should be called $v-$irreducibility).

Observe that conditions **1.** and **2.** above allow to express the evolution of X as a regenerative process (see Nummelin (1978)). Indeed, note that if $x \in C$, condition **2.** indicates that

$$P(x, A) = \lambda v(A) + (1 - \lambda) Q(x, A),$$

where $Q(\cdot)$ is the so-called "residual Markov transition kernel". Then, one can construct the process X by allowing a transition from $X_n = x$ to X_{n+1} according to $P(x, \cdot)$ if $x \notin C$, whereas, if $x \in C$, we generate a Bernoulli I_n such that $P(I_n = 1) = \lambda$ and we transition to X_{n+1} $v(\cdot)$ if $I_n = 1$. Otherwise (if $I_n = 0$) the transition occurs according to $Q(x, \cdot)$. Hence, a regeneration occurs every time the process X visits the set C (which occurs infinitely often) and we obtain a head in the corresponding flip coin.

In our mathematical development it will be useful to consider the process

$$W_n \triangleq (X_n, I_n), \tag{2}$$

where I_n is obtained as described in the previous paragraph if $X_n \in C$ and $I_n = 0$ if $X_n \notin C$.

It is *important* to observe that one does not have to simulate the process using the previous regenerative construction. It is possible to simulate the process according the kernel $P(x,dy)$ and later identify the regeneration epochs by observing that when $x \in C$

$$\begin{aligned} P\left(I_n = 1 | \, X_{n+1} = y + dy, X_n = x \right) &= \frac{P\left(I_n = 1, X_{n+1} = y + dy | \, X_n = x \right)}{P\left(X_{n+1} = y + dy | \, X_n = x \right)} \\ &= \frac{\lambda v\left(y + dy \right)}{P\left(x, y + dy \right)}. \end{aligned}$$

Typically, both $v(\cdot)$ and $P(x,\cdot)$ have densities $\upsilon(\cdot)$ and $p(x,\cdot)$ with respect to the Lebesgue measure. In this case, we have that

$$P\left(I_n = 1 | \, X_{n+1} = y, X_n = x \right) = \lambda \upsilon\left(y \right) / p\left(x, y \right).$$

Regeneration provides a convenient vehicle for the steady-state analysis of the chain X.

In order to take advantage of the regeneration ideas, it is convenient to introduce some notation. Let $(\theta_k : k \geq -1)$ be the sequence of regeneration times for the chain X (let us introduce the convention that $\theta_{-1} = 0$). There are many ways of embedding regeneration times in the chain X, one way was described in the previous paragraphs, we can think that the value of the chain at time θ_k, X_{θ_k}, has distribution $v(\cdot)$, that is X_{θ_k} has been generated after observing $X_{\theta_k - 1} \in C$ and $I_{\theta_k - 1} = 1$. Set $\theta_k = \theta_{k-1} + \tau_k$, and note that the random elements $((Y_k, \tau_k) : k \geq 0)$ (where $Y_k = (X(\theta_{k-1} + n) : 0 \leq n \leq \tau_k - 1)$) are iid. We now can apply well studied regeneration ideas to the chain X. Using regenerative theory we can write

$$\begin{aligned} P\left(X_\infty \in A \right) &= \frac{E\left(\sum_{k=0}^{\tau-1} 1\left(X_k \in A \right) \right)}{E\tau} \\ &= \frac{E\left(\sum_{k=0}^{\infty} 1\left(\tau > k; X_k \in A \right) \right)}{E\tau} \\ &= \frac{\sum_{k=0}^{\infty} P\left(X_k \in A | \, \tau > k \right) P\left(\tau > k \right)}{E\tau} = Eh\left(T \right) \end{aligned} \tag{3}$$

where

$$h\left(k \right) \triangleq P\left(X_k \in A | \, \tau > k \right),$$

and

$$P\left(T = k \right) = P\left(\tau > k \right) / E\tau \quad \text{for } k \geq 0. \tag{4}$$

The distribution of T is the so-called "equilibrium distribution" of the discrete renewal process with inter-arrivals $(\tau_k : k \geq 1)$. Representation (3) readily yields the following simulation algorithm for that allows to generate exact samples from the stationary distribution of the chain X.

Algorithm 1

Underlying we have regeneration times $(\theta_k : k \geq -1)$ (suppose that $\theta_{-1} = 0$) and that $\theta_k = \theta_{k-1} + \tau_k$. Note that the random elements $((Y_k, \tau_k) : k \geq 0)$ (where $Y_k \triangleq (X(\theta_{k-1} + n) : 0 \leq n \leq \tau_k - 1)$) are iid.

STEP 1 Simulate a random variable T with distribution (4) independent of the chain X.

STEP 2 Let $K^* = \min(k : \tau_k > T)$ and return $X(\tau_{K^*} + T)$.

The problem with Algorithm 1 is that it does not specify how to generate T in STEP 1. In the so-called Doeblin case, which occurs if condition 2. holds for $C = E$, it can be easily seen that T is Geometrically distributed with mean $(1-\lambda)/\lambda$ and therefore Algorithm 1 can be implemented. A natural path to pursue in the general case is to take advantage of acceptance/rejection ideas in order to sample from the distribution of T. For instance, suppose that is possible to find a bound of the form

$$P(\tau > k) \leq c_1 g(k) \tag{5}$$

for some computable constant $c_1 > 0$ and a function $g(k)$ such that the sum

$$\sum_{k=0}^{\infty} g(k) = c_2 < \infty$$

can be evaluated explicitly. The problem of finding bounds such as (5) will be discussed in the next section. We then would like to implement the following algorithm.

Algorithm for the Equilibrium Distribution (ED)

STEP 1 Sample T_0 according to the distribution proportional to $g(k)$, set $T_0 = t$.

STEP 2 Generate a Bernoulli I_0 with probability $P(\tau > t)/(c_1 g(t))$ and accept $T_0 = t$ if $I = 1$, otherwise go to STEP 1 until acceptance.

Although the previous algorithm (**Algorithm ED)** generates copies of T, it seems that in its implementation it is required to evaluate $P(\tau > k)$ explicitly, which is typically extremely complicated except in some cases such as the Doeblin case discussed before. However, as we mentioned, the work by Keane and O'Brien (1994) and Nacu and Peres (2004) allow to do this because, in the context of **Algorithm 1**, we can easily generate Bernoulli's with success probability $P(\tau > t)$. We then have proved the following theorem.

Theorem 1 *Suppose that the chain X can be simulated and that regeneration times $(\theta_k : k \geq -1)$ (with $\theta_{-1} = 0$) can be identified and that (5) can be obtained. Then, **Algorithm 1** samples from the stationary distribution of the chain X in finite time.*

The previous result does not provide an idea of the running time of **Algorithm 1**. The next proposition provides a negative result in terms of the complexity of **Algorithm 1**.

Proposition 2 *Let K^* be defined as in STEP 2 of* ***Algorithm 1****. Then,*

$$EK^* = \infty.$$

Proof. Given that STEP 1 in **Algorithm 1** yields $T = t$, K^* is geometric with mean $P(\tau \leq t)/P(\tau > t)$. More precisely, for $k \geq 0$

$$P(K^* = k|\, T = t) = P(\tau > t)\, P(\tau \leq t)^k .$$

Therefore,

$$\begin{aligned} EK^* &= \sum_{t=0}^{\infty} E(K^*|\, T = t)\, P(T = t) \\ &= \sum_{t=0}^{\infty} P(\tau \leq t)\, P(T = t)/P(\tau > t) \\ &= \sum_{k=0}^{\infty} P(\tau \leq t)/E\tau = \infty. \end{aligned}$$

■

As a consequence, **Algorithm 1** should not be used in a multiple-replication scheme. More precisely, if we are interested in estimating $Ef(X_\infty)$ via simulation, we should not attempt to generate multiple independent copies of X_∞. Instead, we should use **Algorithm 1** just to delete the bias in a single replication of the chain and combine this with an effective variance reduction technique.

3 Drift Conditions

As we indicated earlier, a crucial element of **Algorithm 1** is being able to find a bound such as (5). In this section we shall study several drift conditions that can be used to obtain successful bound such as (5). We shall indicate later how can we easily obtain bounds for moments of τ in terms of bounds for moments of the first passage time, T_C, to the small set C. Thus, first we shall focus on developing sufficient conditions that can be exploited to obtain bounds for the moments of T_C. In order to understand the development of such bounds let us consider the case of obtaining bounds for the first moment of τ. The next result will be useful for this purpose.

Proposition 3 *Let $h_1 : E \rightarrow \mathbb{R}_+$ be such that*

$$E(h_1(X_1)|\, X_0 = x) \leq h_1(x) - 1 + b_1 I_C(x), \tag{6}$$

for some $b_1 \in (0, \infty)$. Then,

$$E(T_C|\, X_0 = x) \leq h_1(x),$$

for $x \notin C$.

Proof. Apply optional sampling to the non-negative supermartingale

$$M_n = h_1\left(X_{n\wedge T_C}\right) + \left(n \wedge T_C\right),$$

where $n \wedge T_C = \min\left(n, T_C\right)$. (See Proposition 4 for a completely analogous argument that contains all the details.) We then obtain (using that $g\left(\cdot\right) \geq 0$)

$$E\left(n \wedge T_C\right) \leq EM_n \leq h_1\left(x\right).$$

The proposition follows by letting $n \nearrow \infty$ using the Monotone Convergence Theorem. ■

Let us define $\widetilde{T}_C = \inf\{n \geq 1 : X_n \in C\}$. The time $\widetilde{T}_C$ plays an important role in the analysis of the regeneration times. In order to see this let us define $\widetilde{T}_C^{(j)} = \inf\{n > \widetilde{T}_C^{(j-1)} : X_n \in C\}$. Set $\widetilde{\theta}_j = \widetilde{T}_C^{(j)} - \widetilde{T}_C^{(j-1)}$ for $j \geq 2$ and put $\widetilde{\theta}_1 = T_C$. Then,

$$P\left(\tau \leq n\right) = P_v\left(1 + \sum_{j=1}^{G} \widetilde{\theta} j \leq n\right), \tag{7}$$

where $G = \inf\{n \geq 1 : I_{\widetilde{\theta}_n} = 1\}$ is the number of trials required to obtain a regeneration, which occurs if we visit the set C (recall the definition of I_n in (2)). Therefore, $P\left(G = k\right) = \lambda\left(1-\lambda\right)^{k-1}$ for $k \geq 1$, that is, G is geometrically distributed with parameter λ. Consequently, it is not hard to see that

$$E\tau \leq 1 + E_v\left(T_C\right) + 1/\lambda \sup_{x\in C} E\left(\widetilde{T}_C \middle| X_0 = x\right). \tag{8}$$

This bound can typically be controlled if C is a compact set, we shall explain how can we estimate the right hand side of (8). Note that

$$E\left(\widetilde{T}_C \middle| X_0 = x\right) = 1 + E_x\left(1\left(X_1 \in C\right) E_x\left(T_C \middle| X_1\right)\right) \tag{9}$$

$$+E_x\left(1\left(X_1 \notin C\right) E_x\left(T_C \middle| X_1\right)\right). \tag{10}$$

Using Proposition 3 we observe that (10) can be bounded via

$$\begin{aligned} E_x\left(1\left(X_1 \notin C\right) E_x\left(T_C \middle| X_1\right)\right) &\leq E_x\left(1\left(X_1 \notin C\right) h_1\left(X_1\right)\right) \\ &\leq E\left(h_1\left(X_1\right) \middle| X_0 = x\right). \end{aligned}$$

On the other hand,

$$E_x\left(1\left(X_1 \in C\right) E_x\left(T_C \middle| X_1\right)\right) = 0.$$

Therefore, combining these estimates with (9) and (10) we obtain

$$\sup_{x\in C} E\left(\widetilde{T}_C \middle| X_0 = x\right) \leq 1 + \sup_{x\in C} E\left(h_1\left(X_1\right) \middle| X_0 = x\right). \tag{11}$$

Likewise,

$$E_v T_C = E_v \left(E_{X_0} \left(T_C \right); X_0 \notin C \right) \leq E_v \left(h_1 \left(X_0 \right); X_0 \notin C \right).$$

Hence, using (8) we arrive at

$$E\tau \leq 1 + E_v \left(h_1 \left(X_0 \right); X_0 \notin C \right) + 1/\lambda \sup_{x \in C} E \left(h_1 \left(X_1 \right) \middle| X_0 = x \right). \tag{12}$$

It is convenient to let C take the form $C = \{x : h_1(x) \leq c\}$ for some $c > 0$. In this case we obtain, using (12)

$$\sup_{x \in C} E \left(\widetilde{T}_C \middle| X_0 = x \right) \leq c + b.$$

Bounds such as (12) can also be developed for higher order moments of τ. As we have seen the main ingredient for these bounds is given by suitable drift conditions such as (6) and a careful analysis of the moments of T_C, just as we discussed in the previous paragraphs. We then shall focus on the analysis of higher order moments of T_C via suitable drift conditions such as (6) and later we will take advantage of this analysis to provide bounds on higher order moments of τ.

Assuming that we are able to find $h_1(\cdot)$ from Proposition 3 it is straightforward to obtain conditions in order to bound $E\left(T_C^2 \middle| X_0 = x \right)$. These are given in the next proposition

Proposition 4 *Suppose that there exists a function $h_2 : E \to \mathbb{R}_+$ and a constant $b_2 \in (0, \infty)$ such that*

$$E \left(h_2 \left(X_1 \right) \middle| X_0 = x \right) \leq h_2(x) - h_1(x) + b_2 I_C(x). \tag{13}$$

Then,

$$E \left(T_C^2 \middle| X_0 = x \right) \leq h_2(x),$$

for $x \notin C$.

Proof. Define

$$M_n = h_2 \left(X_{n \wedge T_C} \right) + \sum_{k=0}^{n \wedge T_C - 1} h_1 \left(X_k \right).$$

We claim that $(M_n : n \geq 0)$ is a non-negative supermartingale. The fact that M_n is non-negative is immediate. Now, let $(\mathcal{F}_n : n \geq 0)$ be the filtration generated by the process $X = (X_n : n \geq 0)$, then

$$\begin{aligned} E \left(M_{n+1} \middle| \mathcal{F}_n \right) &= E \left(M_{n+1} \middle| \mathcal{F}_n \right) + E \left(M_{n+1} \middle| \mathcal{F}_n \right) \\ &= 1 \left(T_C \leq n \right) \left(h_2 \left(X_{T_C} \right) + \sum_{k=0}^{T_C - 1} h_1 \left(X_k \right) \right) \\ &\quad + 1 \left(T_C > n \right) E \left(h_2 \left(X_{n+1} \right) + \sum_{k=0}^{n} h_1 \left(X_k \right) \middle| \mathcal{F}_n \right). \end{aligned} \tag{14}$$

Using the Markov property and (13) we obtain

$$1(T_C > n)\, E\left(h_2(X_{n+1}) \middle| \mathcal{F}_n \right) \le 1(T_C > n)\left(h_2(X_n) - h_1(X_n) \right).$$

Therefore,

$$\begin{aligned} & 1(T_C > n)\, E\left(h_2(X_{n+1}) + \sum_{k=0}^{n} h_1(X_k) \middle| \mathcal{F}_n \right) \\ \le \quad & 1(T_C > n)\left(h_2(X_n) + \sum_{k=0}^{n-1} h_1(X_k) \right) \\ = \quad & 1(T_C > n)\left(h_2(X_{n\wedge T_C}) + \sum_{k=0}^{n\wedge T_C - 1} h_1(X_k) \right). \end{aligned}$$

Therefore, combining this estimate with (13) we conclude

$$\begin{aligned} E\left(M_{n+1} \middle| \mathcal{F}_n \right) \quad \le \quad & 1(T_C \le n)\left(h_2(X_{T_C}) + \sum_{k=0}^{T_C - 1} h_1(X_k) \right) \\ & + 1(T_C > n)\left(h_2(X_{n\wedge T_C}) + \sum_{k=0}^{n\wedge T_C - 1} h_1(X_k) \right) \\ = \quad & M_n, \end{aligned}$$

which implies the supermartingale property of M. Consequently, using the fact that both $h_2(\cdot)$ and $h_1(\cdot)$ are non-negative and applying the monotone convergence theorem we obtain that

$$E \sum_{k=0}^{T_C - 1} h_1(X_k) \le E M_n \le M_0 = h_2(x). \tag{15}$$

Now, a straightforward application of Fubini's theorem yields

$$E \sum_{k=0}^{T_C - 1} h_1(X_k) = \sum_{k=0}^{\infty} E\left(1(T_C > k)\, h_1(X_k) \right). \tag{16}$$

By assumption, since $h_1(\cdot)$ satisfies the conditions in Proposition 3, we obtain

$$1(T_C > k)\, h_1(X_k) \ge 1(T_C > k)\, E_{X_k}(\tau),$$

where $E_x(\cdot)$ is the expectation operator induced by the Markov chain X starting at x. We then obtain

$$\begin{aligned} \sum_{k=0}^{\infty} E\left(1(T_C > k)\, h_1(X_k) \right) \quad \ge \quad & \sum_{k=0}^{\infty} E\left(1(T_C > k)\, E_{X_k}(T_C) \right) \\ = \quad & \sum_{k=0}^{\infty} E\left(1(T_C > k)\, E\left(T_C \middle| \mathcal{F}_k \right) \right) \\ = \quad & \sum_{k=0}^{\infty} E\left(T_C 1(T_C > k) \right) = E\left(T_C^2 \right). \end{aligned} \tag{17}$$

Inequality (17) combined with (15) and (16) yield the proof of the proposition. ■

More generally, the following theorem allows to obtain bounds on higher order moments of the first passage time T_C.

Theorem 5 *For $1 \leq j \leq k$, suppose that $h_j : E \to \mathbb{R}_+$ satisfies*

$$E\left(h_j\left(X_1\right) \middle| X_0 = x\right) \leq h_j\left(x\right) - h_{j-1}\left(x\right) + b_j I_C\left(x\right),$$

where $b_j \in (0, \infty)$ and $h_0(\cdot) = 1$. Then,

$$E\left(T_C^j \middle| X_0 = x\right) \leq h_j\left(x\right),$$

for $x \notin C$.

Proof. Similar to that of Proposition 4. ■

The previous estimates, can be easily adapted to provide bounds for the moments of $\widetilde{T}_C$. Specifically, if $(h_j(\cdot) : 0 \leq j \leq k)$ satisfy the conditions of Theorem 5, then

$$\begin{aligned}
E_x\left(\widetilde{T}_C^k\right) &= E_x\left(E_x\left(\widetilde{T}_C^k \middle| X_1\right); X_1 \in C\right) + E_x\left(E_x\left(\widetilde{T}_C^k \middle| X_1\right); X_1 \notin C\right) \\
&= 1 + \sum_{j=1}^{k} \binom{k}{j} E_x\left(E_x\left(T_C^j \middle| X_1\right); X_1 \notin C\right) \\
&\leq 1 + \sum_{j=1}^{k} \binom{k}{j} E_x\left(h_j\left(X_1\right); X_1 \notin C\right) \\
&\leq 1 + \sum_{j=1}^{k} \binom{k}{j} E_x\left(h_j\left(X_1\right)\right) \\
&\leq 1 + \sum_{j=1}^{k} \binom{k}{j} \left(h_j\left(x\right) - h_{j-1}\left(x\right) + b_j 1_C\left(x\right)\right).
\end{aligned}$$

The previous estimates together with identity (7) provide the means to obtain bounds for the moments of τ using the bounds developed for moments of T_C as the next series of results indicate.

Lemma 6 *Suppose that $(h_j(\cdot) : 0 \leq j \leq k)$ satisfy the conditions of Theorem 5, then*

$$E_v T_C^k \leq E_v\left(h_k\left(X_0\right); X_0 \notin C\right) \leq E_v\left(h_k\left(X_0\right)\right).$$

Proof. We proceed by direct computation,

$$E_v\left(T_C^k\right) = E_v\left(E\left(T_C^k \middle| X_0\right); X_0 \notin C\right) \leq E_v\left(h_k\left(X_0\right); X_0 \notin C\right).$$

■

Using the previous computation it is not conceptually hard to obtain bounds for higher order moments of τ. We shall illustrate the idea next for the second moment of τ.

Theorem 7 *Suppose that* $h_1(\cdot)$ *and* $h_2(\cdot)$ *satisfy the conditions of Proposition 4 and that* $\sup_{x\in C} h_j(x) \le c_j < \infty$*. Define* $a = c_1 + c_2 + 2b_1 + b_2 - 1$*. Then,*

$$\begin{aligned} E\tau^2 &\le 1 + E_v T_C^2 + a/\lambda + 2(1+\lambda) a E_v (h_2(X_0))/\lambda \\ &\quad + 2a^2/((1-\lambda)\lambda) + 2E\tau. \end{aligned}$$

Proof. Recall the definitions of $(I_n : n \ge 0)$ and $\left(\widetilde{\theta}_n : n \ge 1\right)$given in (2) and (7) respectively. Note that if X_0 is distributed according to $v(\cdot)$, then the following equality in distribution holds

$$\tau = 1 + \sum_{j=1}^{\infty} \widetilde{\theta}_j \prod_{k=1}^{j-1} \left(1 - I_{\widetilde{\theta}_k}\right).$$

Hence,

$$\begin{aligned} \tau^2 &= 1 + \sum_{j=1}^{\infty} \widetilde{\theta}_j^2 \prod_{k=1}^{j-1} (1 - I_k) + \\ &\quad 2\sum_{j=1}^{\infty} \sum_{m=j+1}^{\infty} \widetilde{\theta}_j^2 \widetilde{\theta}_m^2 \prod_{k=1}^{j-1} \left(1 - I_{\widetilde{\theta}_k}\right) \prod_{k=1}^{m-1} \left(1 - I_{\widetilde{\theta}_k}\right) \\ &\quad + 2\sum_{j=1}^{\infty} \widetilde{\theta}_j \prod_{k=1}^{j-1} \left(1 - I_{\widetilde{\theta}_k}\right) \end{aligned}$$

Now, for $j \ge 2$, we have

$$\begin{aligned} & E_v\left(\widetilde{\theta}_j^2 \prod_{k=1}^{j-1} \left(1 - I_{\widetilde{\theta}_k}\right)\right) \\ \le\ & E_v\left(\prod_{k=1}^{j-2} \left(1 - I_{\widetilde{\theta}_k}\right) E_v\left(\widetilde{\theta}_j^2 \,\middle|\, W_0, W_{\widetilde{\theta}_1}, \ldots W_{\widetilde{\theta}_{j-2}}, X_{\widetilde{\theta}_{j-1}}\right)\right). \end{aligned}$$

(Note we have dropped $I_{\widetilde{\theta}_{j-1}}$ because it is not independent of $\widetilde{\theta}_j^2$.) However, observe that (6) and (13) imply

$$\begin{aligned} & E_v\left(\widetilde{\theta}_j^2 \,\middle|\, W_0, W_{\widetilde{\theta}_1}, \ldots W_{\widetilde{\theta}_{j-2}}, X_{\widetilde{\theta}_{j-1}}\right) = E_{X_{\widetilde{\theta}_{j-1}}}\left(\widetilde{T}_C^2\right) \\ \le\ & 1 + 2\left(h_1\left(X_{\widetilde{\theta}_{j-1}}\right) - 1 + b_1 1_C\left(X_{\widetilde{\theta}_{j-1}}\right)\right) \\ & + \left(h_2\left(X_{\widetilde{\theta}_{j-1}}\right) - h_1\left(X_{\widetilde{\theta}_{j-1}}\right) + b_2 1_C\left(X_{\widetilde{\theta}_{j-1}}\right)\right) \\ \le\ & c_1 + c_2 + 2b_1 + b_2 - 1 = a. \end{aligned}$$

Hence,

$$E_v \sum_{j=1}^{\infty} \widetilde{\theta}_j^2 \prod_{k=1}^{j-1} (1 - I_k) = E_v T_C^2 + a/\lambda. \tag{18}$$

Similarly, if $m \geq 2$,

$$\begin{aligned}
& E_v \left(\widetilde{\theta}_1^2 \widetilde{\theta}_m^2 \prod_{k=1}^{m-1} \left(1 - I_{\widetilde{\theta}_k}\right) \right) \\
\leq \quad & a E_v \left(\widetilde{\theta}_1^2 \prod_{k=1}^{m-2} \left(1 - I_{\widetilde{\theta}_k}\right) \right) \leq a \left(1-\lambda\right)^{(m-3)\vee 0} E_v \left(\widetilde{\theta}_1^2\right) \\
= \quad & a \left(1-\lambda\right)^{(m-3)\vee 0} E_v \left(h_2 \left(X_0\right)\right).
\end{aligned}$$

This gives,

$$2 \sum_{m=2}^{\infty} E \widetilde{\theta}_1^2 \widetilde{\theta}_m^2 \prod_{k=1}^{m-1} \left(1 - I_{\widetilde{\theta}_k}\right) \leq 2 \left(1+\lambda\right) a E_v \left(h_2 \left(X_0\right)\right)/\lambda. \tag{19}$$

Finally, if $m > j \geq 2$, then

$$\begin{aligned}
& E_v \left(\widetilde{\theta}_j^2 \widetilde{\theta}_m^2 \prod_{k=1}^{j-1} \left(1 - I_{\widetilde{\theta}_k}\right) \prod_{k=1}^{m-1} \left(1 - I_{\widetilde{\theta}_k}\right) \right) \\
= \quad & E_v \left(\widetilde{\theta}_j^2 \prod_{k=1}^{m-2} \left(1 - I_{\widetilde{\theta}_k}\right) E_v \left(\widetilde{\theta}_m^2 \middle| W_0, W_{\widetilde{\theta}_1}, \ldots W_{\widetilde{\theta}_{m-2}}, X_{\widetilde{\theta}_{m-1}} \right) \right) \\
\leq \quad & a E_v \left(\widetilde{\theta}_j^2 \prod_{k=1}^{m-2} \left(1 - I_{\widetilde{\theta}_k}\right) \right) \\
\leq \quad & a \left(1-\lambda\right)^{(m-2-j)\vee 0} E_v \left(\widetilde{\theta}_j^2 \prod_{k=j}^{j-1} \left(1 - I_{\widetilde{\theta}_k}\right) \right) \\
\leq \quad & a^2 \left(1-\lambda\right)^{(m-2-j)\vee 0} \left(1-\lambda\right)^{(j-2)}.
\end{aligned}$$

This implies that

$$\begin{aligned}
& 2 \sum_{j=2}^{\infty} \sum_{m=j+1}^{\infty} E_v \left(\widetilde{\theta}_j^2 \widetilde{\theta}_m^2 \prod_{k=1}^{j-1} \left(1 - I_{\widetilde{\theta}_k}\right) \prod_{k=1}^{m-1} \left(1 - I_{\widetilde{\theta}_k}\right) \right) \\
\leq \quad & 2a^2 \left(1-\lambda\right)^{-4} \sum_{j=2}^{\infty} \sum_{m=j+1}^{\infty} \left(1-\lambda\right)^m = 2a^2 / \left(\left(1-\lambda\right)\lambda\right).
\end{aligned} \tag{20}$$

Combining (18), (19) and (20) the theorem follows. ■

Theorem 7 can be extended to cover higher order moments. The details are not conceptually complicated but somewhat burdensome. Also, note that

Theorem 7 can be used in (5), even if the chain is not geometrically ergodic in order to implement Algorithm 1. Of course, if the chain is geometrically ergodic it is possible to obtain bounds for $E\exp(\delta\tau)$ for some $\delta > 0$. This implies that the acceptance/rejection procedure in Algorithm ED can be implemented by drawing proposals from a "light-tailed" distribution, thereby speeding up the running time of Algorithm ED.

Just as in the case of the analysis of the moments of τ, in order to obtain a bound for $E\exp(\delta\tau)$, we will proceed by considering $E_x\exp(\delta T_C)$.

Theorem 8 *Suppose that there exists a function $h : \mathbb{R} \longrightarrow [1,\infty)$ (for some $\varepsilon > 0$) and a constant $b < \infty$ such that,*

$$\frac{h(x)}{h_0(x)} \geq e^{\delta} \quad \text{if} \quad x \notin C, \tag{21}$$

where

$$h_0(x) = E^Q\left(h(X_1) \mid X_0 = x\right).$$

Then, for $x \notin C$

$$h(x) \geq E_x \exp(\delta T_C).$$

Proof. Define the process

$$M_n = h(X_{n\wedge T_C})\exp(n \wedge T_C) = h(X_{n\wedge T_C}) \prod_{k=0}^{n\wedge T_C - 1} \exp(\delta).$$

We claim that M_n is a supermartingale (with respect to the standard stopped filtration generated by the process X, which we shall denote by $\mathcal{F}_n$). In order to see this note that

$$\begin{aligned} E\left(M_{n+1} \mid \mathcal{F}_n\right) &= E\left(M_{n+1} \mid \mathcal{F}_n\right)1(T_C \leq n) \\ &\quad + E\left(M_{n+1} \mid \mathcal{F}_n\right)1(T_C > n) \\ &= M_\tau 1(T_C \leq n) + E\left(M_{n+1} \mid \mathcal{F}_n\right)1(T_C > n). \end{aligned}$$

Now, on $\{\tau > n\}$ we have that

$$\begin{aligned} & E\left(M_{n+1} \mid \mathcal{F}_n\right)1(T_C > n) \\ = \; & h_0(X_n)1(T_C > n)\prod_{k=0}^{n}\exp(\delta) \\ \leq \; & \frac{h(X_n)}{h_0(X_n)}h_0(X_n)1(T_C > n)\prod_{k=0}^{n-1}\exp(\delta) \\ = \; & M_n 1(T_C > n). \end{aligned}$$

Therefore, we conclude that M is a non-negative supermartingale and therefore, if $S_0 = y$ we obtain

$$h(y) \geq EM_n$$

moreover, using Fatou's lemma we obtain (since $P_x(T_C < \infty) = 1$ and $h(\cdot) \geq 1$)

$$h(x) \geq E_x h(X_{T_C}) \exp(\delta T_C) \geq E_x \exp(\delta T_C).$$

■

The previous theorem provides the means to obtain a bound for $E\exp(\delta\tau)$.

Theorem 9 *Let $h(\cdot)$ and $\delta > 0$ satisfy the conditions of Theorem 8. In addition, let*

$$\beta = \sup_{x\in C} \left(P_x(X_1 \in C) + E_x(h(X_1); X_1 \notin C)\right)$$

and assume that

$$\alpha \triangleq \exp(\delta)\left(\beta - \lambda P_v(X_0 \in A) - \lambda E_v(h(X_0); X_0 \notin C)\right) < 1.$$

Then,

$$E\exp(\delta\tau) \leq \lambda \exp(\delta)\left(P_v(X_0 \in C) + E_v(\exp(\delta) h(X_1); X_1 \notin C)\right)/\beta.$$

Proof. Using (7) we obtain

$$E\exp(\delta\tau) = \exp(\delta) E_v\left(E_v\left(\exp\left(\delta\sum_{k=1}^{G}\widetilde{\theta}_k\right)\middle| G\right)\right).$$

Now, if $k \geq 2$, we have that

$$\begin{aligned} & P\left(\widetilde{\theta}_k \in A\middle| X_{\widetilde{\theta}_{k-1}} = x, I_{\widetilde{\theta}_{k-1}} = 0\right) \\ = & \int P_y(T_C \in A)\, Q(x,dy), \end{aligned} \tag{22}$$

where $Q(x,dy) = (P(x,dy) - \lambda v(dy))/(1-\lambda)$. Hence, (since $h(\cdot) \geq 1$)

$$\begin{aligned} & E\left(\exp\left(\delta\widetilde{\theta}_k\right)\middle| X_{\widetilde{\theta}_{k-1}} = x, I_{\widetilde{\theta}_{k-1}} = 0\right) \\ = & \exp(\delta)\left(P_x^Q(X_1 \in C) + E_x^Q(h(X_1); X_1 \notin C)\right) \\ \leq & \exp(\delta)\,\alpha/(1-\lambda). \end{aligned}$$

Consequently,

$$\begin{aligned} & \sum_{n=1}^{\infty} E_v\left(\exp\left(\delta\sum_{k=1}^{G}\widetilde{\theta}_k\right)\middle| G = n\right)\lambda(1-\lambda)^{n-1} \\ \leq & E_v(\exp(\delta T_C))\,\lambda\sum_{n=1}^{\infty}\alpha^{n-1} = E_v(\exp(\delta T_C))\,\lambda/\alpha. \end{aligned}$$

The conclusion of the theorem follows immediately by noting that (since $h(\cdot) \geq 1$)

$$\begin{aligned} & E_v\left(\exp\left(\delta T_C\right)\right) \\ = \; & P_v\left(X_0 \in C\right) + E_v\left(\exp\left(\delta\right) E\left(\left.\exp\left(\delta T_C\right)\right| X_1\right); X_1 \notin C\right) \\ \leq \; & P_v\left(X_0 \in C\right) + E_v\left(\exp\left(\delta\right) h\left(X_1\right); X_1 \notin C\right) \leq \exp\left(\delta\right) E_v h\left(X_1\right). \end{aligned}$$

■

Remark The previous theorem is very related to Theorem 8 of Rosenthal (1995).

The results of this section provide all the necessary elements required to apply Algorithm 1 in a great variety of application settings.

References

[1] Asmussen, Glynn and Thorisson (1992). Stationarity Detection in the Initial Transient Problem. *ACM Transactions on Modeling and Computer Simulation.* vol. **2**, 130-157.

[2] Corcoran, J. and Tweedie, R. (2001) Perfect Sampling of Ergodic Harris Chains. *Ann. Appl. Probab.* 11 (2) 438 - 451.

[3] Foss, S. and Tweedie, R. (1998) Perfect Simulation and Backward Coupling, *Stochastic Models*. Vol **14**. 187–203.

[4] Hobert, J. and Robert, C. (2004) A mixture representation of π with applications in Markov chain Monte Carlo and perfect sampling, *Annals of Applied Probability*. **14**, no. 3, 1295-1305.

[5] Hobert, J., Jones, G., and Robert, C. (2005) Using a Markov chain to construct a tractable approximation of an intractable probability distribution, *Scandinavian Journal of Statistics*. To appear.

[6] Huber, M. (2004) Perfect sampling using bounding chains, *Ann. Appl. Probab.* **14**, no. 2, 734-753.

[7] Kendall, W. (1998) Perfect simulation for the area-interaction point process. In L. Accardi and C. C. Heyde, editors, *Probability Towards 2000*, pages 218–234. Springer.

[8] Keane, M and O'Brien, G. (1994) A Bernoulli Factory. *ACM Transactions on Modeling and Computer Simulation*. Vol. 4, 213-219.

[9] Meyn, S. and Tweedie, R. (1993) *Markov Chains and Stochastic Stability.* Springer, London.

[10] Murdoch, D. J. and Green, P. J. (1998). Exact sampling from a continuous state space. *Scandinavian Journal of Statistics*, 25:483–502.

[11] Nacu, S. and Peres, Y. (2004) Fast Simulation of New Coins From Old. (preprint).

[12] Nummelin, E. (1978) A Splitting Technique for Harris Recurrent Markov Chains. *Z. Wahrsch. Verw. Gebiete.* 43 (4) 309-318.

[13] Propp, J. G. and Wilson, D. B. (1996). Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Structures and Algorithms*, 9:223–252.

[14] Roshenthal, J. (1995) Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo. *J. Amer. Statist. Assoc.*, **90**:558-566.