Electronic Journal of Statistics
Vol. 8 (2014) 2664–2692
ISSN: 1935-7524
DOI: 10.1214/14-EJS969

Monte Carlo methods for improper target distributions

Krishna B. Athreya

Department of Mathematics and Statistics, Iowa State University, Ames, Iowa, 50011
e-mail: kbathreya@gmail.com

and

Vivekananda Roy

Department of Statistics, Iowa State University, Ames, Iowa 50011
e-mail: vroy@iastate.edu

Abstract: Monte Carlo methods (based on iid sampling or Markov chains) for estimating integrals with respect to a proper target distribution (that is, a probability distribution) are well known in the statistics literature. If the target distribution π happens to be improper then it is shown here that the standard time average estimator based on Markov chains with π as its stationary distribution will converge to zero with probability 1, and hence is not appropriate. In this paper, we present some limit theorems for regenerative sequences and use these to develop some algorithms to produce strongly consistent estimators (called *regeneration* and *ratio* estimators) that work whether π is proper or improper. These methods may be referred to as *regenerative sequence Monte Carlo (RSMC) methods*. The regenerative sequences include Markov chains as a special case. We also present an algorithm that uses the domination of the given target π by a probability distribution π_0. Examples are given to illustrate the use and limitations of our algorithms.

AMS 2000 subject classifications: Primary 65C05, 62M99; secondary 60G50.
Keywords and phrases: Importance sampling, improper posterior, Markov chains, MCMC, null recurrence, ratio limit theorem, regenerative sequence.

Received February 2014.

1. Introduction

Let π be a distribution (that is, a measure) on a set S and let f be a real valued function on S, integrable with respect to π, that is, $\int_S |f| d\pi < \infty$. Let $\lambda := \int_S f d\pi$. The problem addressed in this paper is to find appropriate statistical procedures to estimate λ. That is, the problem is to find a way to generate data $(X_1, X_2, \ldots, X_n)$ and find a statistic $\hat{\lambda}_n$ such that for large n, $\hat{\lambda}_n$ is close to λ in a suitable sense. When π is known to be a probability distribution, that is when $\pi(S) = 1$, a well established statistical procedure is the classical (iid) Monte Carlo method. This method requires one to generate iid observations

$(X_1, X_2, \ldots, X_n)$ with distribution π and estimate λ by $\hat{\lambda}_n \equiv \sum_{i=1}^n f(X_i)/n$. By the law of large numbers for iid random variables, $\hat{\lambda}_n \xrightarrow{\text{a.s.}} \lambda$ as $n \to \infty$. If in addition $\int_S f^2 d\pi < \infty$ then the accuracy of the estimate $\hat{\lambda}_n$, that is the behavior of the quantity $(\hat{\lambda}_n - \lambda)$ is (by the classical CLT) of the order $n^{-1/2}$.

Over the past twenty years an alternative statistical procedure known as Markov chain Monte Carlo (MCMC) has come into extensive use in statistical literature. In particular, if generating iid observations from π is not easy then classical Monte Carlo cannot be used to estimate λ. The MCMC method is to generate a Markov chain $\{X_n\}_{n\geq 0}$ with π as its invariant distribution and that is suitably irreducible. Then one appeal to the law of large numbers for such chains that says that for any initial value of X_0, the "time average" $\sum_{j=1}^n f(X_j)/n$ converges almost surely to the "space average" $\lambda = \int_S f d\pi$. Indeed, the pioneering paper by [15] did introduce this method where the target distribution π, although a probability distribution, turned out to be the so called Gibbs measure on the configuration space in statistical mechanics and it was difficult to simulate from π. In this fundamental paper, they constructed a Markov chain $\{X_n\}_{n\geq 0}$ that is appropriately irreducible and has π as its stationary distribution. This idea was adapted to some statistical problems by [10]. This whole apparatus was rediscovered in the late 80's and early 90's by a number of statisticians and since then the subject has exploded [see e.g. 21]. Both the above two methods, that is, the iid Monte Carlo and the MCMC depend crucially on the assumption that π is a probability distribution.

A natural question suggested by the above discussion is the following. Suppose π is not proper (that is, $\pi(S) = \infty$) but $f : S \to \mathbb{R}$ is such that $\int_S |f| d\pi < \infty$. How do we estimate λ by statistical tools as distinct from classical numerical analytic tools? Clearly, the iid Monte Carlo to generate data from π is not feasible since $\pi(S) = \infty$. However, it may be possible to generate a Markov chain $\{X_n\}_{n\geq 0}$ that is appropriately irreducible and admits the (improper) distribution π as its stationary distribution. In such situations, it can be shown (see Corollary 2) that the usual estimator $\sum_{j=1}^n f(X_j)/n$ will indeed not work, that is, will not converge to λ, but rather go to zero with probability 1. In particular, in Bayesian statistics if an improper prior is assigned to the parameters then the resulting posterior distribution π (for the given data) is not guaranteed to be proper. If posterior propriety is not checked and π happens to be improper then Corollary 2 implies that the usual MCMC estimator $\sum_{j=1}^n f(X_j)/n$ based on a Markov chain $\{X_n\}_{n\geq 0}$ with π as its stationary (improper) distribution will simply not work in the sense that it will go to zero with probability 1 and not to λ. The use of MCMC for improper posterior distribution does not seem to have received much attention in the literature. The exceptions include [22, 12], and [13].

In this paper we show that given a distribution π on some space S that may or may *not* be proper, it may be possible to generate data $\{X_n\}_{n\geq 0}$ such that more general time averages (called *regeneration* and *ratio* estimators in Theorem 1) than the standard Monte Carlo estimator, $\sum_{j=1}^n f(X_j)/n$, will converge to λ. This is based on the idea of regenerative sequences (Definition 1) which in-

clude iid samples and Markov chains (null or positive recurrent) as special cases (see section 2). For Harris recurrent Markov chains, the possibility of using the ratio estimator proposed here was previously noted by [22]. Based on these theoretical results we present several algorithms for statistical estimation of λ (see sections 2 and 3). Section 3 discusses the use of a special class of regenerative sequences, namely regenerative sequences (not necessarily Markov chains) based on null recurrent random walks for estimating countable sums and integrals when $S = \mathbb{R}^d$ for *any* $d < \infty$. Some examples are given in Section 4. Some of these examples illustrate the use of *null recurrent* Gibbs chains for consistent estimation of quantities like λ. This use of null recurrent Markov chains does not seem to have been noted before in the statistical literature. Section 4 also presents a generalization of the importance sampling methods and its potential application in Bayesian sensitivity analysis.

Another method discussed in this paper uses the idea of dominating the given target π by a probability distribution π_0 and then using the standard Monte Carlo methodology (iid or MCMC) with π_0 as its stationary distribution. This is discussed in Section 5. Some concluding remarks appear in Section 6. Finally, proofs of theorems and corollaries are given in Appendix B and some technical results appear in Appendix A and Appendix C.

2. Convergence results for regenerative sequences

In this section we introduce the concept of regenerative sequences and establish some convergence results for these. This in turn provides some useful statistical tools for estimating parameters. These tools are more general than the Monte Carlo methods based on iid observations or Markov chains. In particular, they are applicable to the case when the target distribution is not proper. The following definition of a regenerative sequence is given in [2].

Definition 1. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(S, \mathcal{S})$ be a measurable space. A sequence of random variables $\{X_n\}_{n\geq 0}$ defined on $(\Omega, \mathcal{F}, P)$ with values in $(S, \mathcal{S})$ is called *regenerative* if there exists a sequence of integer (random) times $T_0 = 0 < T_1 < T_2 < \cdots$ such that the excursions $\{X_n : T_j \leq n < T_{j+1}, \tau_{j+1}\}_{j\geq 0}$ are iid where $\tau_j = T_j - T_{j-1}$ for $j = 1, 2, \ldots$, that is,

$$P(\tau_{j+1} = k_j, X_{T_j+q} \in A_{q,j}, 0 \leq q < k_j, j = 0, \ldots, r) = \prod_{j=0}^{r} P(\tau_1 = k_j, X_{T_1+q} \in A_{q,j}, 0 \leq q < k_j),$$

for all $k_0, k_1, \ldots, k_r \in \mathbb{N}$, the set of positive integers, $A_{q,j} \in \mathcal{S}, 0 \leq q < k_j, j = 0, 1, \ldots, r$, and $r \geq 1$. The random times $\{T_n\}_{n\geq 0}$ are called *regeneration times* and $\{\tau_j\}_{j\geq 1}$ are the *excursion times*.

Example 1. Let $\{X_n\}_{n\geq 0}$ be an irreducible, recurrent Markov chain with countable state space $S = \{s_0, s_1, s_2, \ldots\}$. Let $X_0 = s_0$, $T_0 = 0$, $T_{r+1} = \inf\{n : n \geq T_r + 1, X_n = s_0\}$, $r = 0, 1, 2, \ldots$. Then $\{X_n\}_{n\geq 0}$ is regenerative with regeneration times $\{T_n\}_{n\geq 0}$.

Example 2. Let $\{X_n\}_{n\geq 0}$ be a Markov chain with general state space $(S, \mathcal{S})$. Assume that $\mathcal{S}$ is countably generated and $\{X_n\}_{n\geq 0}$ is Harris recurrent. Then it has been independently shown by [3] and [18] that for some $n_0 \geq 1$, the Markov chain $\{Z_n\}_{n\geq 0}$ has an atom that it visits infinitely often, where $Z_n \equiv X_{nn_0}$ for $n = 0, 1, \ldots$. By the Markov property, the excursions between consecutive visits to this atom are iid making $\{Z_n\}_{n\geq 0}$ a regenerative sequence (see e.g. Meyn and Tweedie [16, Section 17.3.1] or Athreya and Lahiri [2, Theorem 14.2.8]).

Example 3. Let $\{Y_n\}_{n\geq 0}$ be a Harris recurrent Markov chain as in Example 2. Given $\{Y_n = y_n\}_{n\geq 0}$, let $\{A_n\}_{n\geq 0}$ be independent positive integer valued random variables. Set

$$X_n = \begin{cases} y_0 & 0 \leq n < A_0 \\ y_1 & A_0 \leq n < A_0 + A_1 \\ \vdots & \end{cases}.$$

Then $\{X_n\}_{n\geq 0}$ is called a semi Markov chain with embedded Markov chain $\{Y_n = y_n\}_{n\geq 0}$ and sojourn times $\{A_n\}_{n\geq 0}$. Since $\{Y_n\}_{n\geq 0}$ is regenerative, it follows that $\{X_n\}_{n\geq 0}$ is regenerative.

The following theorem presents some useful limit results for regenerative sequences that can be used for constructing consistent estimators of integrals with respect to improper measures.

Theorem 1. *Let $\{X_n\}_{n\geq 0}$ be a regenerative sequence with regeneration times $\{T_n\}_{n\geq 0}$. Let*

$$\check{\pi}(A) = E\left(\sum_{j=0}^{T_1-1} I_A(X_j)\right) \text{ for } A \in \mathcal{S}. \tag{2.1}$$

(The measure $\check{\pi}$ is known as the canonical (or, occupation) measure for $\{X_n\}_{n\geq 0}$.) Let $N_n = k$ if $T_k \leq n < T_{k+1}, k, n = 0, 1, 2, \ldots$, that is, N_n is the number of regenerations by time n. Suppose $f, g : S \to \mathbb{R}$ are two measurable functions such that $\int_S |f| d\check{\pi} < \infty$, $\int_S |g| d\check{\pi} < \infty$, and $\int_S g d\check{\pi} \neq 0$. Then, as $n \to \infty$,

(i) (regeneration estimator)

$$\hat{\lambda}_n = \frac{\sum_{j=0}^n f(X_j)}{N_n} \xrightarrow{a.s.} \int_S f d\check{\pi}, \tag{2.2}$$

(ii) (ratio estimator) and

$$\hat{R}_n = \frac{\sum_{j=0}^n f(X_j)}{\sum_{j=0}^n g(X_j)} \xrightarrow{a.s.} \frac{\int_S f d\check{\pi}}{\int_S g d\check{\pi}}. \tag{2.3}$$

A proof of Theorem 1 is given in Appendix B. Theorem 1(i) for continuous time regenerative sequences is available in [1]. Theorem 1(ii) for Harris recurrent Markov chains is available in Meyn and Tweedie [16, Theorem 17.3.2]. In Definition 1, if $\{X_n : T_j \leq n < T_{j+1}, \tau_{j+1}\}$'s are assumed to be iid *only* for

$j \geq 1$, then $\{X_n\}_{n\geq 0}$ is called a *delayed* regenerative process. Straightforward modification of the proof of Theorem 1 shows that it holds for any delayed regenerative process.

The next theorem describes the asymptotic behavior of the regeneration times, T_n and the number of regenerations, N_n. The result on the asymptotic distribution of T_n presented in the theorem below is already proved in Feller [8, Chapter XIII, Section 6].

Theorem 2. *Let $\{T_j\}_{j\geq 0}$ and $\{\tau_j\}_{j\geq 1}$ be as defined in Definition 1. Suppose in addition $P(\tau_1 > x) \sim x^{-\alpha}L(x)$ as $x \to \infty$, where $0 < \alpha < 1$ and $L(x)$ is slowly varying at ∞ (i.e., $L(cx)/L(x) \to 1$ as $x \to \infty$, $\forall\, 0 < c < \infty$).*

(i) Then

$$\frac{T_n}{a_n} \xrightarrow{d} V_\alpha, \quad \text{as } n \to \infty, \tag{2.4}$$

where $\{a_n\}$ satisfies $na_n^{-\alpha}L(a_n) \to 1$ and V_α is a positive random variable with a stable distribution with index α such that

$$E(\exp(-sV_\alpha)) = \exp(-s^\alpha \Gamma(1-\alpha)), \quad 0 \leq s < \infty, \tag{2.5}$$

where for $p > 0$, $\Gamma(p) = \int_0^\infty u^{p-1}\exp(-u)du$ is the Gamma function.

(ii) Also

$$\frac{N_n}{b_n} \xrightarrow{d} Y_\alpha, \quad \text{as } n \to \infty, \tag{2.6}$$

where $b_n = n^\alpha/L(n)$ and

$$P(Y_\alpha \leq x) = P(V_\alpha \geq x^{-\frac{1}{\alpha}}) \ \text{ for } 0 < x < \infty,$$

that is $Y_\alpha \stackrel{d}{=} V_\alpha^{-\alpha}$ and V_α is as defined in (2.5).

As mentioned before, the proof of Theorem 2(i) is available in Feller [8, Chapter XIII, Section 6]. The proof of Theorem 2(ii) for random walk Markov chains on $\mathbb{R}$ using method of moments is presented in [4, page 229]. We provide a proof of Theorem 2(ii) in Appendix B.

Recall from the Introduction that our goal is to obtain a consistent estimate of $\lambda = \int_S f d\pi$ for a given (possibly infinite) measure π and an integrable function $f : S \to \mathbb{R}$. Theorem 1 yields the following generic recipe for estimating λ.

---

1. Construct a regenerative sequence $\{X_n\}_{n\geq 0}$ with π as its canonical measure.
2. As defined in Theorem 1, let N_n be the number of regenerations by time n. Estimate λ by $\hat{\lambda}_n = \sum_{j=1}^n f(X_j)/N_n$. By Theorem 1(i) $\hat{\lambda}_n \xrightarrow{\text{a.s.}} \lambda$.
3. If we can find a function g such that $E_\pi g := \int_S g d\pi (\neq 0)$ is known then we can estimate λ by $(E_\pi g)\hat{R}_n$ consistently, where $\hat{R}_n$ is defined in (2.3).

In order to use the above recipe, one needs to find a regenerative sequence with π as its canonical measure. Later in this section, we see that a Markov

chain $\{X_n\}_{n\geq 0}$ that has a recurrent *atom* and has π as its invariant measure can be used for this purpose.

The following corollary uses (2.6) to provide an alternative estimator of λ.

Corollary 1. *From (2.2) and (2.6) it follows that under the hypotheses of Theorem 1 and Theorem 2,* $\sum_{j=0}^{n} f(X_j)/b_n \xrightarrow{d} (\int_S f d\check{\pi})\, Y_\alpha$. *If* $E(Y_\alpha)$ *is known and is finite then we can estimate* $\int_S f d\check{\pi}$ *in the following way. Run* k *independent copies of* $\{X_n\}$ *and let* $D_n^r = \sum_{j=0}^{n} f(X_j^r)/b_n$ *where* $\{X_j^r\}_{j=0}^{n}$ *are the realized values of the chain in the* r*th run,* $r = 1, 2, \ldots, k$. *Since for large* n, $D_n^r, r = 1, 2, \ldots, k$ *are independent draws from an approximation to* $(\int_S f d\check{\pi})\, Y_\alpha$, *we can estimate* $\int_S f d\check{\pi}$ *by* $\hat{\lambda}_{n,k} \equiv \sum_{r=1}^{k} D_n^r/(kE(Y_\alpha))$. *It can be shown that there exists a sub-sub sequence* $k_{j_l} \uparrow \infty$ *and* $n_{k_{j_l}} \uparrow \infty$ *such that* $\hat{\lambda}_{n_{k_{j_l}}, k_{j_l}} \xrightarrow{a.s.} \int_S f d\check{\pi}$ *as* $l \to \infty$.

A proof of Corollary 1 is given in Appendix B.

For the Markov chains in Example 1, it can be shown that $P(X_n = s_i$ for some $1 \leq n < \infty | X_0 = s_j) = 1$ for all $s_i, s_j \in S$, that is, for such chains any $s_i \in S$ is a *recurrent point* and the Markov chain regenerates every time it reaches s_i. If $\{X_n\}_{n\geq 0}$ is a Markov chain with general state space $(S, \mathcal{S})$ and is ϕ-irreducible with a nontrivial σ-finite reference measure ϕ, then by the regeneration method of [3] and [18], the Markov chain $\{X_n\}_{n\geq 0}$ is equivalent to an appropriate Markov chain $\check{X}_n$ on an enlarged space $\check{S}$ so that $\check{X}_n$ admits an *atom* Δ in $\check{S}$ [16, Theorem 5.1.3]. Now in addition if we assume that $\{X_n\}_{n\geq 0}$ is Harris recurrent as in Example 2, then $P(\check{X}_n = \Delta$ for some $1 \leq n < \infty | \check{X}_0 = x) = 1$ for all $x \in \check{S}$, i.e., Δ is a recurrent point for $\check{X}_n$. In Appendix A we present some results on the existence and the uniqueness of invariant measure for Markov chains. Using these results and Theorem 1, the following remark discusses how integrals with respect to a given target distribution can be estimated using Markov chains.

Remark 1. Let π be a given (proper or improper) target distribution and $\{X_n\}_{n\geq 0}$ be a Markov chain as in Examples 1 or 2 with invariant distribution π. Then from Corollary 3 in Appendix A we know that π is the unique (up to a multiplicative constant) invariant measure for $\{X_n\}_{n\geq 0}$. Also from Theorem 1 we know that for any two integrable functions $f, g : S \to \mathbb{R}$ with $\int_S g d\pi \neq 0$, no matter what the distribution of the starting value X_0, as $n \to \infty$, the ratio $\hat{R}_n = \sum_{j=0}^{n} f(X_j)/\sum_{j=0}^{n} g(X_j)$ converges almost surely to $\int_S f d\pi / \int_S g d\pi$. If we choose g such that $E_\pi g = \int_S g d\pi$ is easy to evaluate, then $\lambda = \int_S f d\pi$ can be consistently estimated by $\hat{R}_n(E_\pi g)$. In particular, if $g(\cdot) = I_K(\cdot)$, where K is a (possibly cylinder) subset of S with known $\pi(K) < \infty$ then $\hat{R}_n = \sum_{j=0}^{n} f(X_j)/N_n(K)$, where $N_n(K) = \sum_{j=0}^{n} I_K(X_j)$, the number of times the Markov chain visits K by time n. Further, from Corollary 3 in Appendix A we note that one may use the regenerative estimator $\hat{\lambda}_n$ as in (2.2) if $\pi(\{\Delta\})$ is known.

From Remark 1 we see that when π is not proper, the regeneration estimator (2.2) and the ratio estimator (2.3) based on a Markov chain provide consistent

estimates of λ. On the contrary, in Corollary 2 below we show that when π is improper, the usual time average MCMC estimator $\sum_{j=1}^{n} f(X_j)/n$ is not consistent for λ. Indeed, we show that it converges to zero with probability 1. A result similar to Corollary 2 but under a different set of hypotheses was established by [13].

Theorem 3. *Let $\{X_n\}_{n\geq 0}$ be a regenerative sequence with occupation measure $\check{\pi}$ as in (2.1). Let $\check{\pi}(S) = E(T_1) = \infty$. Then for any $f : S \to \mathbb{R}$ such that $\int_S |f| d\check{\pi} < \infty$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(X_i) = 0 \quad \textit{with probability } 1.$$

Corollary 2. *Let $\{X_n\}_{n\geq 0}$ be a Harris recurrent Markov chain with improper invariant measure π, that is, $\pi(S) = \infty$. Let $f : S \to \mathbb{R}$ be such that $\int_S |f| d\pi < \infty$. Then the standard time average estimator $\sum_{i=1}^{n} f(X_i)/n$ converges to 0 with probability 1.*

The proof of Theorem 3 is given in Appendix B. Corollary 2 follows from Theorem 3 since as pointed out in Example 2, every Harris recurrent Markov chain is regenerative. In the improper target case, N_n, the number of regenerations, grows slower than n as noted in (2.6). On the other hand, from (2.2) we know that the time sums $\sum_{i=1}^{n} f(X_i)$ grow exactly at the rate N_n. This may explain why the usual time average MCMC estimator is not consistent for λ when π is improper.

3. Algorithms based on random walks

In this section we focus on the use of a special class of regenerative sequences, namely regenerative sequences (not necessarily Markov chains) based on null recurrent random walks on $\mathbb{Z}$ and its variants for estimating countable sums and integrals. We begin with the case when S is countable.

Let S be any countable set and $f : S \to \mathbb{R}$ be such that $\sum_{s\in S} |f(s)| < \infty$. Our goal is to obtain a consistent estimator of $\lambda = \sum_{s\in S} f(s)$. Without loss of generality (see Remark 3) we take S to be $\mathbb{Z}$, the set of integers. Let $\{X_n\}_{n\geq 0}$ be a simple symmetric random walk (SSRW) on $\mathbb{Z}$ starting at X_0. That is, $X_{n+1} = X_n + \delta_{n+1}, n \geq 0$ where $\{\delta_n\}_{n\geq 1}$ are iid with distribution $P(\delta_1 = +1) = 1/2 = P(\delta_1 = -1)$ and independent of X_0. This Markov chain $\{X_n\}_{n\geq 0}$ is null recurrent with the counting measure $\{\pi(i) \equiv 1 : i \in \mathbb{Z}\}$ being its unique (up to multiplicative constant) invariant measure [see e.g. 16, Section 8.4.3]. From Example 1 of Section 2, we know that $\{X_n\}_{n\geq 0}$ is regenerative and it satisfies the hypothesis of Corollary 3 in Appendix A with $\Delta = 0$. Since $\pi(\Delta) = 1$, Corollary 3 in Appendix A allows us to identify the occupation measure $\tilde{\pi}$ with the counting measure and hence from (2.2) it follows that for any initial X_0,

$$\hat{\lambda}_n = \frac{\sum_{j=0}^{n} f(X_j)}{N_n} \xrightarrow{\text{a.s.}} \sum_{i\in\mathbb{Z}} f(i) \equiv \lambda, \tag{3.1}$$

where $N_n = \sum_{j=0}^{n} I(X_j = 0)$ is the number of visits to zero by $\{X_j\}_{j=0}^{n}$. So $\hat{\lambda}_n$ is a consistent estimator of λ. Let $X_0 = 0$, $T_0 = 0$, $T_{r+1} = \inf\{n : n \geq T_r + 1, X_n = 0\}$, $r = 0, 1, 2, \ldots$ and $\tau_i = T_i - T_{i-1}, i \geq 1$. Then $\{\tau_i\}_{i\geq 1}$ are iid positive random variables with $P(\tau_1 > n) \sim \sqrt{2/\pi} n^{-1/2}$ as $n \to \infty$ [see e.g. 7, page 203]. So from Theorem 2 (that is, from (2.6)) using $b_n = n^{1/2}\sqrt{\pi/2}$, we have

$$\frac{N_n}{\sqrt{n}} \xrightarrow{\text{d}} \sqrt{\frac{\pi}{2}} Y_{1/2},$$

where $Y_{1/2} = V_{1/2}^{-1/2}$ and V_α is as defined in (2.5). Since the density function of $V_{1/2}$ is available [see e.g. 8, page 173], simple calculations show that $Y_{1/2} \stackrel{\text{d}}{=} |Z|$ where $Z \sim \text{N}(0, 1)$. (We can also use (2.5) to find the distribution of $Y_{1/2}$.) So $\sum_{j=0}^{n} f(X_j)/\sqrt{n} \xrightarrow{\text{d}} \sqrt{\frac{\pi}{2}}\lambda|Z|$ and since $E(|Z|) = \sqrt{2/\pi}$ is known we can use Corollary 1 to obtain an alternative estimator of λ.

Remark 2. It can be shown [see 23] that the number of distinct values of X_j for $0 \leq j \leq n$ grows at the rate of $\sqrt{n}$. Thus, after n steps of running the SSRW Markov chain, the computation of $\sum_{j=0}^{n} f(X_j)$ involves evaluation of f only at approximately $\sqrt{n}$ sites in $\mathbb{Z}$.

The above discussion implies the following algorithm, that can be used to estimate of $\lambda = \sum_{s\in S} f(s)\pi(s)$ where π is a given (possibly infinite) measure on a countable set S and $f : S \to \mathbb{R}$ is a function satisfying $\sum_{s\in S} |f(s)|\pi(s) < \infty$. Again, without loss of generality (see Remark 3) we assume that $S = \mathbb{Z}$.

Algorithm I:

1. Run the simple symmetric random walk (SSRW), $\{X_n\}_{n\geq 0}$ on $\mathbb{Z}$.
2. Let $\hat{\lambda}_n = \sum_{j=0}^{n} f(X_j)\pi(X_j)/N_n$ where N_n is the number of visits by $\{X_j\}_{j=0}^{n}$ to 0 by time n. Then $\hat{\lambda}_n \xrightarrow{\text{a.s.}} \lambda$, that is $\hat{\lambda}_n$ is a consistent estimator of λ.

Remark 3. Let S be any countable set and as before let $\mathbb{N}$ be the set of positive integers. Then there exists a $1-1$, onto map $\zeta : \mathbb{N} \to S$. Consider the map $\xi : \mathbb{Z} \to \mathbb{N}$, defined as follows

$$\xi(j) = \begin{cases} 2j & \text{if } j = 1, 2, \ldots \\ -2j + 1 & \text{if } j = 0, -1, -2, \ldots \end{cases}.$$

Then ξ is $1-1$, onto map from $\mathbb{Z}$ to $\mathbb{N}$ and so $\zeta \circ \xi$ is a $1-1$, onto map from $\mathbb{Z}$ to S. Using Algorithm I, we can estimate $\lambda = \sum_{s\in S} f(s)\pi(s)$ by $\sum_{j=1}^{n}(f \circ \zeta \circ \xi(X_j))(\pi \circ \zeta \circ \xi(X_j))/N_n$. We now present some examples of the function ζ.

(i) Note that, if $S = \mathbb{Z}$, we can take ζ to be simply ξ^{-1}.

(ii) When $S = \mathbb{Z}^2$, define $A = \{2^i 3^j : i \geq 1, j \geq 1\}$. There exists a $1-1$, onto map $\rho_1 : \mathbb{N} \to A$, for example take $\rho_1(1)$ to be the smallest element of $A = 2 \times 3 = 6$, $\rho_1(2)$ to be the second smallest element $= 2^2 \times 3 = 12$, and so on. Define another map $\rho_2 : A \to \mathbb{N}^2$ by $\rho_2(2^i 3^j) = (i, j)$. Lastly consider $\rho_3 : \mathbb{N}^2 \to \mathbb{Z}^2$ where $\rho_3(i, j) = (\varphi(i), \varphi(j))$ and $\varphi = \xi^{-1}$, that is

$$\varphi(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even} \\ -\frac{n-1}{2} & \text{if } n \text{ is odd} \end{cases}.$$

So $\zeta \equiv \rho_3 \circ \rho_2 \circ \rho_1 : \mathbb{N} \to S = \mathbb{Z}^2$ is $1-1$, onto.

(iii) We can similarly define ζ when $S = \mathbb{Z}^d$ for $d \geq 3$.

Now we consider the case $S = \mathbb{R}$. Let $f : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable and let $\lambda = \int_{\mathbb{R}} f(x)dx$. Our goal is to obtain a consistent estimator of λ. Let $\{\eta_i\}_{i\geq 1}$ be iid Uniform $(-1/2, 1/2)$ random variables. Let $X_{n+1} = X_n + \eta_{n+1}, n \geq 0, X_0 = x$ for some $x \in \mathbb{R}$. Then it can be shown (see e.g. Durrett [7, page 195], Meyn and Tweedie [16, Proposition 9.4.5]) that $\{X_n\}_{n\geq 0}$ is a Harris recurrent Markov chain with state space $\mathbb{R}$ and the Lebesgue measure as its invariant measure (unique up to a multiplicative constant). From Example 2 of Section 2, we know that $\{X_n\}_{n\geq 0}$ is regenerative and Corollary 3 in Appendix A identifies the occupation measure $\tilde{\pi}$ with (up to a multiplicative constant) the Lebesgue measure. Since the Lebesgue measure of $(-1/2, 1/2)$ is 1, it follows from (2.3) of Theorem 1 that for any initial X_0,

$$\hat{R}_n = \frac{\sum_{j=0}^n f(X_j)}{N_n(K)} \xrightarrow{\text{a.s.}} \int_{\mathbb{R}} f(x)dx,$$

where $N_n(K) = \sum_{j=0}^n I_K(X_j)$ is the number of visits to $K = (-1/2, 1/2)$ by $\{X_j\}_{j=0}^n$. So $\hat{R}_n$ is a consistent estimator of λ.

Since $E(\eta_i) = 0$, $\text{Var}(\eta_i) = 1/12$, and the length of the interval $(-1/2, 1/2)$ is 1, straightforward calculations using Theorem 10.30 in [4] show that

$$\frac{N_n(K)}{\sqrt{n}} \xrightarrow{\text{d}} \sqrt{12}|Z|, \quad Z \sim \text{N}(0, 1).$$

So as in the S countable case, an alternative estimator of λ can be formed following Corollary 1.

Next we consider the case $S = \mathbb{R}^2$. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be Lebesgue integrable and let $\lambda = \int_{\mathbb{R}^2} f(x)dx$. Let $X_{n+1} = X_n + \eta_{n+1}, n \geq 0, X_0 = (0, 0)$ and $\{\eta_i\}_{i\geq 1} \equiv \{(\eta_i^1, \eta_i^2)\}_{i\geq 1}$'s are iid random variables where η_i^1, η_i^2 are iid Uniform $(-1/2, 1/2)$. From Durrett's (2010) Theorem 4.2.8 we know that starting anywhere $X_0 = x$, $\{X_n\}_{n\geq 0}$ visits any circle infinitely often with probability one. Also since η_i has an absolutely continuous distribution, the density of η_i is bounded below by ϵ, for any $0 < \epsilon < 1$ [see e.g. 7, Example 6.8.2]. Thus a regeneration scheme as in [3] exists making $\{X_n\}_{n\geq 0}$ regenerative. Since the Lebesgue measure on $\mathbb{R}^2$ is an invariant measure for $\{X_n\}_{n\geq 0}$ by Corollary 3 in Appendix A and (2.3) of Theorem 1, a consistent estimator of λ is given by

$$\hat{R}_n = \frac{\sum_{j=0}^n f(X_j)}{N_n(K)} \xrightarrow{\text{a.s.}} \int_{\mathbb{R}^2} f(x)dx,$$

where $N_n(K) = \sum_{j=0}^{n} I_K(X_j)$ is the number of visits to $K = [-1/2, 1/2] \times [-1/2, 1/2]$ by $\{X_j\}_{j=0}^{n}$. Since $X_0 = (0,0)$, note that

$$\begin{aligned} P(X_n \in K) &= \Big[P\Big(\sum_{i=1}^{n} \eta_i^1 \in [-1/2, 1/2]\Big)\Big]^2 \\ &= \Big[P\Big(\frac{\sum_{i=1}^{n} \eta_i^1}{\sqrt{n}} \in \Big[\frac{-1}{2\sqrt{n}}, \frac{1}{2\sqrt{n}}\Big]\Big)\Big]^2. \end{aligned}$$

By the local central limit theorem, as $n \to \infty$,

$$\sqrt{n} P\Big(\frac{\sum_{i=1}^{n} \eta_i^1}{\sqrt{n}} \in \Big[\frac{-1}{2\sqrt{n}}, \frac{1}{2\sqrt{n}}\Big]\Big) \to c,\ 0 < c < \infty.$$

Since $E(\eta_i^1) = 0$, Var $(\eta_i^1) = 1/12$, we have $c = \sqrt{12}/\sqrt{2\pi}$. Thus $nP(X_n \in K) \to c^2$, as $n \to \infty$ and hence

$$E[N_n(K)] = 1 + \sum_{j=1}^{n} P(X_j \in K) \sim c^2 \sum_{j=1}^{n} \frac{1}{j} \sim c^2 \log n.$$

(Here, as $n \to \infty$, $c_n \sim d_n$ means that $\lim_{n\to\infty} c_n/d_n = 1$.) This yields the following proposition.

Proposition 1. *Let $N_n(K)$ be as defined above. Then*

$$E\Big(\frac{N_n(K)}{\log n}\Big) \to \frac{6}{\pi}.$$

Next using discrete renewal theory and Karamata's Tauberian theorem, arguing as in [4, Theorem 10.30], one can show that for each $r = 1, 2, \ldots$

$$\lim_n E\Big(\frac{N_n(K)}{\log n}\Big)^r = \mu_r < \infty \ \text{ exists}$$

and $\{\mu_r\}_{r\geq 1}$ is a moment sequence of an exponential distribution and hence satisfies Carleman's condition for determining the distribution. This implies the following theorem holds.

Theorem 4. *Let $N_n(K)$ be as defined above. Then*

$$\frac{N_n(K)}{\log n} \xrightarrow{d} \frac{6}{\pi} Y.$$

where Y is an Exponential(1) random variable.

Some recent results (analogous to Theorem 4) for random walks on $\mathbb{Z}^2$ can be found in [9] and [19]. As in the case of $S = \mathbb{R}$, we can use Theorem 4 to obtain an alternative estimator of λ (Corollary 1).

Let $S = \mathbb{R}$ and π be absolutely continuous with density h. From the above discussion we see that the following algorithm can be used to estimate of $\lambda = \int_{\mathbb{R}} f(x)\pi(dx)$ where $f : \mathbb{R} \to \mathbb{R}$ is π-integrable.

Algorithm II:

1. Run the mean 0 random walk, $\{X_n\}_{n\geq 0}$ on $\mathbb{R}$ with increment distribution Uniform $(-1/2, 1/2)$.
2. Let $\hat{R}_n = \frac{\sum_{j=1}^n f(X_j)h(X_j)}{N_n(K)}$ where $N_n(K)$ is the number of visits by $\{X_j\}_{j=0}^n$ to the unit interval $K = (-1/2, 1/2)$. Then $\hat{R}_n \xrightarrow{\text{a.s.}} \lambda$, that is, $\hat{R}_n$ is a consistent estimator of λ

From the above discussion we know that Algorithm II can be suitably modified for the case $S = \mathbb{R}^2$.

Remark 4. When S is countable, the method proposed here works with the Markov chain SSRW replaced by any mean 0, random walk on $\mathbb{Z}$ such that $X_n/n \to 0$ in probability. By Chung-Fuchs theorem this random walk is recurrent [7, page 195]. Similarly when $S = \mathbb{R}$ or $\mathbb{R}^2$, any mean 0, finite variance random walk with absolutely continuous distribution can be used. On the other hand, a symmetric random walk on $\mathbb{R}$ may not be recurrent. For example, consider a random walk Markov chain $\{X_n\}_{n\geq 0}$ on $\mathbb{R}$ where the increment random variable η has a symmetric stable distribution with index α, $0 < \alpha < 1$, that is, the characteristic function of η is $\psi(t) = \exp(-|t|^\alpha), 0 < \alpha < 1$. It can be shown that $\{X_n\}_{n\geq 0}$ is not recurrent [see 6, page 253] in the sense that for any nondegenerate interval, I in $\mathbb{R}$, $\sum_{n=0}^\infty P(X_n \in I) < \infty$ and hence the number of visits by $\{X_n\}_{n\geq 0}$ to I is finite with probability 1.

Lastly we present an algorithm that is based on the SSRW on $\mathbb{Z}$ and a randomization tool to estimate integrals with respect to an absolutely continuous measure π on *any* $\mathbb{R}^d$, $d < \infty$. Let $f : \mathbb{R}^d \to \mathbb{R}$ be Borel measurable and π be an absolutely continuous distribution on $\mathbb{R}^d$ with Radon-Nikodym derivative $p(\cdot)$. Assume that $\int_{\mathbb{R}^d} |f(x)|p(x)dx < \infty$. The following theorem provides a consistent estimator of $\lambda = \int_{\mathbb{R}^d} f(x)p(x)dx$.

Theorem 5. *Let $\{X_n\}_{n\geq 0}$ be a SSRW on $\mathbb{Z}$. Let $\{U_{ij} : i = 0, 1, \ldots; j = 1, 2, \ldots, d\}$ be a sequence of iid Uniform $(-1/2, 1/2)$ random variables and independent of $\{X_n\}_{n\geq 0}$. Assume $\kappa : \mathbb{Z} \to \mathbb{Z}^d$ be $1-1$, onto. Define $Y_n := \kappa(X_n) + U^n$, $n = 0, 1, , \ldots$ where $U^n = (U_{n1}, U_{n2}, \ldots, U_{nd})$. Then*

$$\hat{\lambda}_n := \frac{\sum_{j=0}^n f(Y_j)p(Y_j)}{N_n} \xrightarrow{a.s.} \int_{\mathbb{R}^d} f(x)p(x)dx, \tag{3.2}$$

where $N_n = \sum_{j=0}^n I(X_j = 0)$ is the number of visits to zero by $\{X_j\}_{j=0}^n$.

It may be noted that the sequence $\{Y_n\}_{n\geq 0}$ defined above is not a Markov chain but is indeed regenerative as in Theorem 1 with $\{T_n\}$ sequence corresponding to the returns of SSRW $\{X_n\}_{n\geq 0}$ to zero. The proof of Theorem 5 is given in Appendix B. Note that, the function κ can be chosen as $\kappa = \zeta \circ \xi$ where ζ and ξ are as given in Remark 3. Theorem 5 suggests the following variation of Algorithm I.

Algorithm III:

1. Generate the sequence $\{Y_n\}_{n\geq 0}$ as in Theorem 5.
2. Estimate $\lambda = \int_{\mathbb{R}^d} f(x)p(x)dx$ by $\hat{\lambda}_n$ defined in (3.2).

4. Examples

In this section we demonstrate the use of the statistical algorithms developed in the previous sections with some examples. Some of our simulations also illustrate that the convergence of these algorithms could be rather slow.

4.1. Simulation examples

We first consider estimating $\lambda = \sum_{m=1}^{\infty} 1/m^2$ using SSRW (Algorithm I) with $X_0 = 0$. The left panel in Figure 1 shows the estimates for 6 values ($\log_{10} n = 3, 4, \ldots, 8$, where $\log_{10}$ denotes logarithm base 10). The plot includes the median estimates and the 90% empirical confidence interval estimates based on 30 independent repetitions. The median and 90% interval estimates for $n = 10^8$ are 1.654, and (1.560, 1.701) respectively. Note that, the true value λ is $\pi^2/6 = 1.645$. The time to run the SSRW chain for 10^8 steps using R [20] in an old Intel Q9550 2.83GHz machine running Windows 7 is about 3 seconds. Next we used the same chain (SSRW with $X_0 = 0$) to estimate $\tilde{\lambda} = \sum_{m=1001}^{\infty} 1/(m-1000)^2$ and found that the first 10^5 iterations yielded an estimate of 0 (red lines in the right panel in Figure 1). However, estimation of $\tilde{\lambda}$ can be improved if we start the SSRW at $X_0 = 1000$, and define regenerations as the returns to 1000 (green lines in the right panel in Figure 1).

Next we consider estimating $\lambda = \int_1^{\infty} 1/x^2 dx$ (see left panel in Figure 2) using Algorithms II (red line) and III (green line) for the same six n values mentioned above. We observe that the interval estimates based on Algorithm II are much

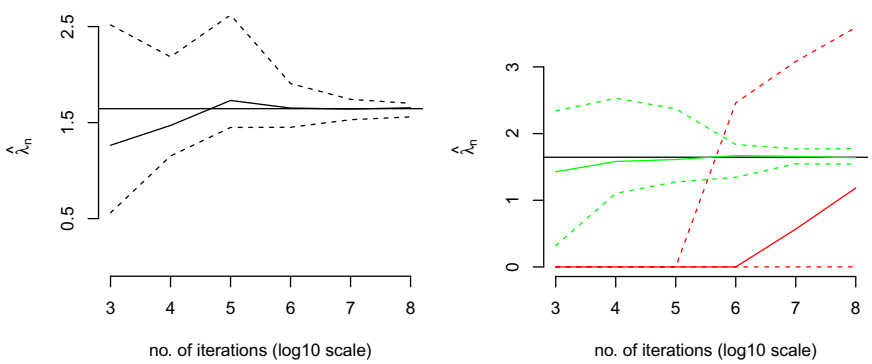

FIG 1. *Point and 90% interval estimates of* $\sum_{m=1}^{\infty} 1/m^2$ *(left panel) and* $\tilde{\lambda} = \sum_{m=1001}^{\infty} 1/(m-1000)^2$ *(right panel).*

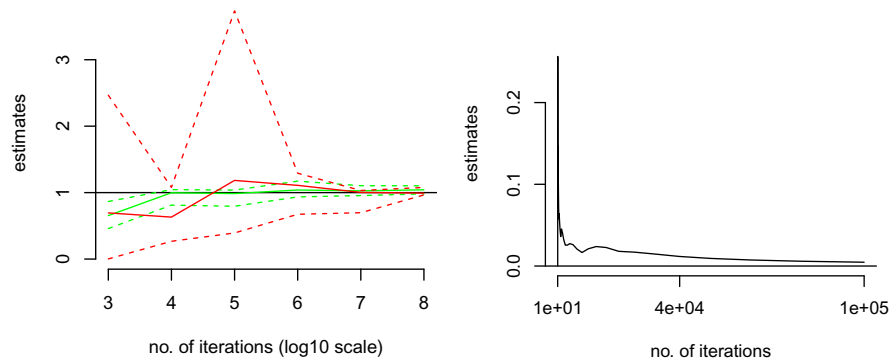

FIG 2. *The left panel shows the point and 90% interval estimates of* $\lambda = \int_1^\infty (1/x^2)dx$. *The right panel shows the behavior of the standard MCMC estimator based on the Gibbs chain.*

wider than those for Algorithm III particularly for smaller values of n. In the examples of this section, for applying Algorithm III we use $\kappa = \zeta \circ \xi$ where ζ and ξ are as given in Remark 3.

Lastly, we consider two examples with $S \subset \mathbb{R}^2$. The first one is the following. Let

$$f(x,y) = \exp[-(x^2 - 2xy + y^2)/2], \quad (x,y) \in \mathbb{R}^2. \tag{4.1}$$

Define the measure π by $\pi(A) = \int\int_A f(x,y)dx\,dy, A \in \mathcal{B}(\mathbb{R}^2)$. Note that π is a σ-finite, infinite measure on $\mathbb{R}^2$. Note that

$$f_{X|Y}(x|y) := \frac{f(x,y)}{\int_{\mathbb{R}} f(x',y)dx'} = \phi(x-y),$$

is a probability density on $\mathbb{R}$, where $\phi(u) = \exp(-u^2/2)/\sqrt{2\pi}, u \in \mathbb{R}$ is the standard normal density. Our goal is to estimate $\pi(A)$ where $A = [-1,1] \times [-1,1]$. Notice that the integral, $\pi(A)$ is not available in closed form. Let $\{(X_n, Y_n)\}_{n\geq 0}$ be the Gibbs sampler Markov chain that uses $f_{X|Y}(\cdot|y)$ and $f_{Y|X}(\cdot|x)$ alternately. Then $\{(X_n, Y_n)\}_{n\geq 0}$ has π as its invariant measure. From Theorem 3 we know that the usual MCMC estimator, $\sum_{j=0}^n I_A(X_j, Y_j)/n \xrightarrow{\text{a.s.}} 0$ since π is an improper measure (see the right panel in Figure 2). It can be verified that $\{(X_n, Y_n)\}_{n\geq 0}$ is regenerative (a proof of this is given in Appendix C). So by Theorem 1, for any initial (X_0, Y_0), and for any $-\infty < a < b < \infty$,

$$\hat{R}_n = \frac{\sum_{j=0}^n I_A(X_j, Y_j)}{\sum_{j=0}^n I_{[a,b]}(Y_j)} \xrightarrow{\text{a.s.}} \frac{\pi(A)}{\sqrt{2\pi}(b-a)},$$

as $\pi(\mathbb{R} \times [a,b]) = \sqrt{2\pi}(b-a)$. Thus $\sqrt{2\pi}(b-a)\hat{R}_n$ is a consistent estimator of $\pi(A)$. The solid lines in the left panel in Figure 3 shows the median (based on 30 independent repetitions) estimates of $\pi((-1,1) \times (-1,1))$ using Algorithm II with uniform random walk in $\mathbb{R}^2$ (red line), Algorithm III (green line) and the above ratio estimator based on the Gibbs sampler (with $-a = b = 1000$) (blue line) for the same six n values mentioned above. The median estimates

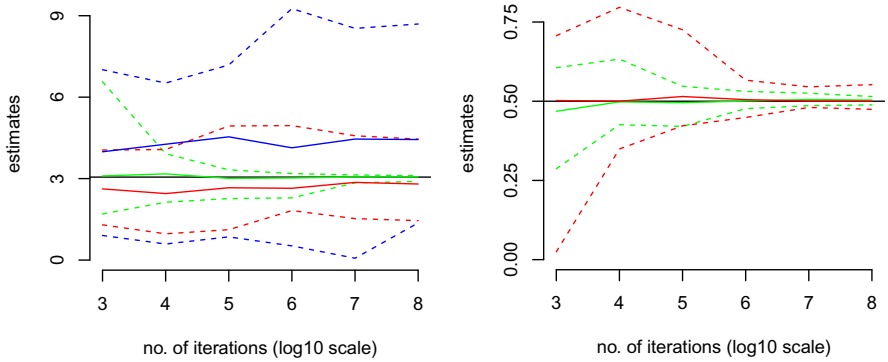

FIG 3. *Estimates using Gibbs sampler and random walks.*

for $n = 10^8$ corresponding to the Gibbs sampler, Algorithm II (with uniform random walk in $\mathbb{R}^2$) and Algorithm III are 4.438, 2.801, and 3.060 respectively. Numerical integration yields 3.056 as an estimate of $\pi(A)$. Like in the previous example, here also Algorithm III performs better than Algorithm II.

The next example is as follows. Let

$$f(x, y) = \exp(-xy) \;\; x, y \in \mathbb{R}_+, \tag{4.2}$$

where $\mathbb{R}_+ = (0, \infty)$. Let

$$f_{X|Y}(x|y) := \frac{f(x, y)}{\int_{\mathbb{R}_+} f(x', y)dx'} = y \exp(-xy); \; 0 < x, y < \infty.$$

Then for each y, $f_{X|Y}(\cdot|y)$ is a probability density on $\mathbb{R}_+$. Consider the Gibbs sampler $\{(X_n, Y_n)\}_{n\geq 0}$ that uses the two conditional densities $f_{X|Y}(\cdot|y)$ and $f_{Y|X}(\cdot|x)$ alternately. [5] considered this example and found that the usual estimator $\sum_{j=0}^{n} f_{X|Y}(\check{x}|Y_j)/n$ for the marginal density $f_X(\check{x}) = \int_{\mathbb{R}_+} f(\check{x}, y)dy = \int_{\mathbb{R}_+} f_{X|Y}(\check{x}|y) f_Y(y)dy = 1/\check{x}$ breaks down. Indeed from Theorem 3 we know that this estimator converges with probability 1 to zero as the above Gibbs sampler chain is regenerative (see Proposition 3 in Appendix C) with an improper invariant measure. ([11] showed that the Gibbs sampler $\{(X_n, Y_n)\}_{n\geq 0}$ is a null recurrent Markov chain, in the sense that $\sum_{n=0}^{\infty} E[I_A(X_n, Y_n)] = \infty$ for any set A with positive Lebesgue measure on $\mathbb{R}^2$.) [5] suggested as a remedy that one restricts the conditional densities to compact intervals. Theorem 1 provides alternatives that do not require this restriction.

By Theorem 1 we have the following consistent estimator of $f_X(\check{x})$. Let $0 < a < b < \infty$. Then for all $\check{x} \in \mathbb{R}_+$,

$$\begin{aligned} \log(b/a)\hat{R}_n &= \log(b/a) \frac{\sum_{j=0}^{n} f_{X|Y}(\check{x}|Y_j)}{\sum_{j=0}^{n} I_{[a,b]}(Y_j)} \\ &= \log(b/a) \frac{\sum_{j=0}^{n} Y_j \exp(-\check{x}Y_j)}{\sum_{j=0}^{n} I_{[a,b]}(Y_j)} \xrightarrow{\text{a.s.}} 1/\check{x} = f_X(\check{x}), \end{aligned}$$

as $\int_S f_{X|Y}(\check{x}|y)f(x,y)dxdy = \int_{\mathbb{R}^+} f_{X|Y}(\check{x}|y)f_Y(y)dy = f_X(\check{x})$ and $\int_S I_K(x,y)f(x,y)dxdy = \log(b/a)$ with $K = \mathbb{R}_+ \times [a,b]$.

Although $\log(b/a)\hat{R}_n$ is a consistent estimator of $f_X(\check{x})$ based on the Gibbs sampler, our simulations show that in this case the Gibbs chain is apt to take values too large or too small (positive) that are outside the range of machine precision. Here we use Algorithms II and III to estimate $f_X(2) = 1/2$. The right panel in Figure 3 shows the estimates (median and 90% interval based on 30 repetitions) using Algorithms II (red line) and III (green line) for the same six n values mentioned above. The median estimates corresponding to red, and green lines for $n = 10^8$ are 0.503, and 0.501 respectively.

Remark 5. Note that in order to use the ratio estimator (2.3) based on the Gibbs sampler $\{(X_n, Y_n)\}_{n\geq 0}$ in the above two examples, by Theorem 1 it suffices to establish that these chains are regenerative. This avoids having to establish the Harris recurrence of $\{(X_n, Y_n)\}_{n\geq 0}$, which is required to apply the Theorem 17.3.2 in Meyn and Tweedie [16].

4.2. Generalization of importance sampling

Importance sampling is a Monte Carlo technique in which a sample from one density is used to estimate an expectation with respect to another. Suppose $\acute{\pi}$ is a probability density on S (with respect a measure ν) and we want to estimate $E_{\acute{\pi}}f = \int_S f\acute{\pi}d\nu$ where $f : S \to \mathbb{R}$ is such that $\int_S |f|\acute{\pi}d\nu < \infty$. Often, such a $\acute{\pi}$ is known only up to a normalizing constant, that is let $\acute{\pi}(x) = cp(x)$, where $c > 0$ is a finite constant, not easy to evaluate and $p(x)$ is a known function. Suppose π is another density (*not* necessarily proper) with respect to the measure ν such that $\{x : \acute{\pi}(x) > 0\} \subset \{x : \pi(x) > 0\}$. Now note that

$$\begin{aligned} E_{\acute{\pi}}f = \int_S f\acute{\pi}d\nu &= \int_S f(x)\frac{\acute{\pi}(x)}{\pi(x)}\pi(x)\nu(dx) \Big/ \int_S \frac{\acute{\pi}(x)}{\pi(x)}\pi(x)\nu(dx) \\ &= \int_S f(x)\frac{p(x)}{\pi(x)}\pi(x)\nu(dx) \Big/ \int_S \frac{p(x)}{\pi(x)}\pi(x)\nu(dx). \end{aligned}$$

Note that $\int_S f(x)[p(x)/\pi(x)]\pi(x)\nu(dx) = E_{\acute{\pi}}f/c$ is well defined and is finite, and $0 < \int_S [p(x)/\pi(x)]\pi(x)\nu(dx) = 1/c < \infty$. So if we can generate a Harris recurrent Markov chain (regenerative sequence) $\{X_n\}_{n\geq 0}$ with invariant (occupation) density π, then by Theorem 1, we have for any initial X_0,

$$\hat{R}_n = \frac{\sum_{j=0}^n f(X_j)\frac{p(X_j)}{\pi(X_j)}}{\sum_{j=0}^n \frac{p(X_j)}{\pi(X_j)}} \xrightarrow{\text{a.s.}} \frac{\int_S f(x)\frac{p(x)}{\pi(x)}\pi(x)\nu(dx)}{\int_S \frac{p(x)}{\pi(x)}\pi(x)\nu(dx)} = E_{\acute{\pi}}f. \tag{4.3}$$

Thus $\hat{R}_n$ is a consistent estimator of $E_{\acute{\pi}}f$. Note that, here the ratio limit theorem (Theorem 1 (ii)) is naturally applicable since the above importance sampling estimator of $E_{\acute{\pi}}f$ has the form of a ratio. That is, in order to use the above importance sampling estimator, we do *not* need to find any function g such

that $\int_S g d\pi (\neq 0)$ is known. Note that the standard importance sampling using MCMC *requires* π to be a probability density and $\{X_n\}_{n\geq 0}$ to be positive recurrent [10]. Here we are able to drop both these conditions.

4.2.1. An example involving Bayesian sensitivity analysis

The importance sampling estimator (4.3) can be useful in Bayesian analysis when one wishes to see the effect of a change in the prior distribution on a posterior expectation. Let $\{\pi_h(\theta), h \in \mathcal{H}\}$ be a family of prior densities, where h is called a hyperparameter. The resulting posterior densities are denoted by $\pi_h(\theta|\mathbf{y}) \propto \ell(\theta|\mathbf{y})\pi_h(\theta)$, where $\mathbf{y}$ denotes the observed data and $\ell(\theta|\mathbf{y})$ is the likelihood function. Let $f(\theta)$ be a quantity of interest where $f(\cdot)$ is a function. In sensitivity analysis one wants to assess the changes in the posterior expectation $E_h(f(\theta)|\mathbf{y})$ as h varies. If two different values of h result in the almost same posterior expectations, then the use of either of those two values will not be controversial. On the other hand, if $E_h(f(\theta)|\mathbf{y})$ changes significantly as h varies, one may want to make a plot of the posterior expectations to see what values of h cause big changes. If posterior expectations are not available in closed form, then the importance sampling estimator (4.3) based on a *single* chain may be used to estimate $E_h(f(\theta)|\mathbf{y})$ for all $h \in \mathcal{H}$. Note that, here the parameter θ corresponds to x in (4.3) and $\pi_h(\theta|\mathbf{y})$ plays the role of $\acute{\pi}$. In general, the importance density π in (4.3) is preferred to have heavier tails than $\pi_h(\theta|y)$ for *all* $h \in \mathcal{H}$, as otherwise the ratio estimate will tend to be unstable because the weights $p_h(\theta)/\pi(\theta)$ can be arbitrarily large, where $p_h(\theta) = \ell(\theta|\mathbf{y})\pi_h(\theta)$. On the other hand, it may be difficult to construct a *proper* importance density π which has heavier tails than $\pi_h(\theta|y)$ for *all* $h \in \mathcal{H}$. Since the importance sampling estimator (4.3) can be used even when π is improper, in the following example, we take π to be the improper uniform density on $\mathbb{R}$.

Let $\mathbf{y} \equiv (y_1, y_2, \ldots, y_m)$ be a random sample from $\mathrm{N}(\mu, 1)$. Suppose, the conjugate normal prior $\mathrm{N}(\mu_0, \sigma_0^2)$ is assumed for μ. Here $h = (\mu_0, \sigma_0)$. Let $\{\mu^j\}_{j\geq 0}$ be the random walk on $\mathbb{R}$ with Uniform $(-1/2, 1/2)$ increment distribution. As mentioned in Section 3, $\{\mu^j\}_{j\geq 0}$ is a Harris recurrent Markov chain with state space $\mathbb{R}$ and the Lebesgue measure as its invariant measure. The estimator (4.3) for $E_h(f(\mu)|\mathbf{y})$ in this case becomes

$$\hat{R}_n = \frac{\sum_{j=0}^n f(\mu^j)\exp(-m(\bar{\mathbf{y}} - \mu^j)^2/2)\phi(\mu^j; \mu_0, \sigma_0^2)}{\sum_{j=0}^n \exp(-m(\bar{\mathbf{y}} - \mu^j)^2/2)\phi(\mu^j; \mu_0, \sigma_0^2)}, \tag{4.4}$$

where $\bar{\mathbf{y}} = \sum_{i=1}^m y_i/m$, and $\phi(x; a, b)$ is the density of the normal distribution with mean a, variance b, evaluated at the point x. A single random walk chain $\{\mu^j\}_{j=0}^n$ started at $\mu^0 = 0$ with $n = 10^7$ is used to estimate $E_h(f(\mu)|\mathbf{y})$ with $f(\mu) \equiv \mu$ for different values of (μ_0, σ_0). The plots in Figure 4 give graphs of the estimate (4.4) as μ_0 and σ_0 vary. For making these plots, we took $m = 5$, and $\bar{\mathbf{y}} = 0$. We fixed $\sigma_0 = 0.5$ for the graph in the left panel in Figure 4 and the plot is based on the estimate (4.4) for 41 values of μ_0. (We took μ_0 ranging from -10 to 10 by increments of 0.5.) The right panel in Figure 4 is

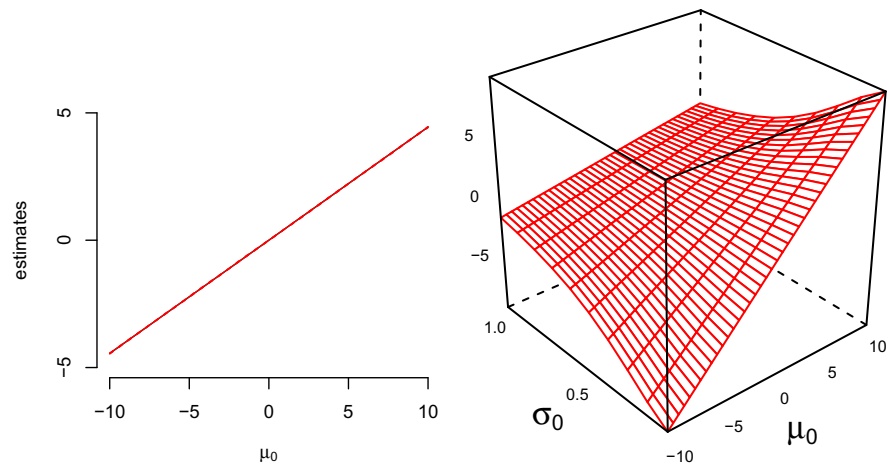

FIG 4. *Posterior expectations for different values of hyperparameters.*

based on the estimate (4.4) for 410 values of (μ_0, σ_0) with μ_0 ranging from -10 to 10 by increments of 0.5, and σ_0 ranging from 0.1 to 1 by increments of 0.1. Both the plots are median estimates of posterior expectations $E_h(\mu|\mathbf{y})$ based on 30 independent repetitions of random walk chains of length $n = 10^7$ with $\mu^0 = 0$. The pointwise empirical 90% confidence intervals are too narrow to distinguish these quantiles from the median estimates in these plots. The median estimates also match (up to two decimal places) with the true values $E_{(\mu_0,\sigma_0)}(\mu|\mathbf{y}) = m\sigma_0^2\bar{\mathbf{y}}/(m\sigma_0^2 + 1) + \mu_0/(m\sigma_0^2 + 1)$.

5. Reduction to proper target distribution

In the previous sections we have presented several algorithms for estimating integrals with respect to a measure π that may be proper or improper. In this section, we propose one more algorithm that involves reduction to a proper target distribution. We begin with the following result.

Lemma 1. *Let π be a σ-finite measure on a measurable space $(S, \mathcal{S})$.*

(i) Let $\{A_n\}_{n\geq 1}$ be a finite or countable sequence of $\mathcal{S}$- sets such that A_n's are disjoint, $\pi(A_n) < \infty$ for all $n \geq 1$, and $S = \cup_{n=1}^{\infty} A_n$. Let $J = \{j : \pi(A_j) > 0\}$ and let $\{p_j : j \in J\}$ be such that $p_j > 0$ for all $j \in J$ and $\sum_{j\in J} p_j = 1$. Then π is dominated by the probability measure π_0 defined by

$$\pi_0(A) = \sum_{j\in J} p_j \frac{\pi(A \cap A_j)}{\pi(A_j)}, \quad A \in \mathcal{S}. \tag{5.1}$$

That is, $A \in \mathcal{S}, \pi_0(A) = 0$ implies $\pi(A) = 0$.

(ii) Further, the Radon-Nikodym derivative $h = d\pi/d\pi_0$ is given by

$$h(\cdot) = \sum_{j\in J} \frac{\pi(A_j)}{p_j} I_{A_j}(\cdot). \tag{5.2}$$

Remark 6. Since there can be many choices of $\{A_n\}_{n\geq 1}$, and since we can use any $\{p_j : j \in J\}$ such that $p_j > 0$ for all $j \in J$ and $\sum_{j\in J} p_j = 1$, the above lemma shows that given a σ-finite measure π, there are uncountably many probability measures π_0 dominating π.

Let π be the given σ-finite (probably infinite, that is, $\pi(S) = \infty$) measure on $(S, \mathcal{S})$ and suppose we can find a probability measure π_0 on S such that π is dominated by π_0. (For example, if π is absolutely continuous on $\mathbb{R}^d$, we can take π_0 to be the probability measure corresponding to the d-variate Student's t distribution.) Let $h \equiv d\pi/d\pi_0$ be the Radon-Nikodym derivative of π with respect to π_0 and suppose we know h only upto a normalizing constant, that is, $h = ch^*$, where the finite constant $c > 0$ is unknown and h^* is known. In that case, if we have an integrable (with respect to π) function $g : S \to \mathbb{R}$ such that $\int_S g d\pi \neq 0$ is known, we have a consistent estimator of $\lambda = \int_S f d\pi$ via a *new* ratio estimator

$$\tilde{R}_n = \frac{\sum_{j=0}^n f(X_j)h^*(X_j)}{\sum_{j=0}^n g(X_j)h^*(X_j)} \xrightarrow{\text{a.s.}} \frac{\int_S fhd\pi_0}{\int_S ghd\pi_0} = \frac{\int_S fd\pi}{\int_S gd\pi}, \tag{5.3}$$

where $\{X_j\}_{j=0}^n$ is either iid sample from π_0 or realizations of a *positive* Harris recurrent Markov chain with invariant measure π_0 started at any initial $X_0 = x_0$.

Lemma 1 and the above discussion is formalized in the following algorithm for estimating $\lambda = \int_S f d\pi$ where π is a σ-finite measure on S.

Algorithm IV:

1. Find a probability measure π_0 dominating π (there are infinitely many such probability measures). Let h be the density of π with respect to π_0.
2. Draw iid sample $\{X_j\}_{j=0}^n$ from π_0 or run a positive Harris recurrent Markov chain $\{X_j\}_{j=0}^n$ (starting at any $X_0 = x_0$) with stationary distribution π_0. Estimate λ by the usual estimator $\sum_{j=0}^n f(X_j)h(X_j)/(n+1)$ if h is completely known, otherwise if h is known up to a multiplicative constant, estimate λ via the ratio estimator (5.3).

One of the difficulties with Algorithm IV is that if π_0 has thinner tails than π, then the function h (or h^*) can take rather large values for some X_j's making the estimator highly varying.

Example 5.1 Let $S = \mathbb{R}$ and let $f : \mathbb{R} \to \mathbb{R}$ be Lebesgue integrable. Our goal is to obtain a consistent estimator of $\lambda = \int_{\mathbb{R}} f(x)\pi(dx)$ using Algorithm IV where $\pi(dx) = dx$ is the Lebesgue measure on $\mathbb{R}$.

Define $A_j = [j, j+1)$ for all $j \in \mathbb{Z}$ and let $p_j > 0,\ j \in \mathbb{Z}$, be such that $\sum_{j\in\mathbb{Z}} p_j = 1$. Since $\pi([j, j+1)) = 1$, from (5.1) we have

$$\pi_0(A) = \sum_{j\in J} p_j \pi(A \cap [j, j+1)), \quad A \in \mathcal{B}(\mathbb{R}).$$

Suppose $U \sim$ Uniform $(0, 1)$ and let N be a $\mathbb{Z}$ valued random variable with $P(N = j) = p_j,\ j \in \mathbb{Z}$. Let U and N be independent and define $X = N + U$. Then it is easy to see that X has an absolutely continuous distribution with probability density function

$$f_X(x) = p_j \ \text{ if } j \le x < j+1, j \in \mathbb{Z},$$

that is $X \sim \pi_0$. From (5.2), the Radon-Nikodym derivative $h = d\pi/d\pi_0 = 1/p_j$ if $j \le x < j+1, j \in \mathbb{Z}$. Let $X_1, X_2, \ldots, X_n$ be iid random sample with common distribution π_0. Then a consistent estimator of λ is $\sum_{j=1}^n f(X_j)h(X_j)/n = (1/n)\sum_{j=1}^n (f(X_j)/p_{\lfloor X_j \rfloor})$.

6. Concluding remarks

In this paper we have considered the problem of estimating $\lambda \equiv \int_S f d\pi$ where $f : S \to \mathbb{R}$, $\int_S |f| d\pi < \infty$, but $\pi(S)$ could be ∞. We have shown that the regular (iid or Markov chain) Monte Carlo methods fail in the case when $\pi(S) = \infty$. Here we have presented several statistical algorithms for estimating λ (when $\pi(S)$ is finite or infinite) based on the notion of regenerative sequences. It may be noted that the general recipe in Section 2 and Algorithm III use regenerative sequences that may not be Markov chains, Algorithms I and II use null recurrent Markov chains. All these methods may be referred to as regenerative sequence Monte Carlo (RSMC) methods. Algorithm IV uses positive recurrent Markov chains.

Although the Algorithms I–IV do provide consistent estimators, in this paper we have not addressed the important question of the second order accuracy, that is, the rate of decay of $(\hat{\lambda}_n - \lambda)$ where $\hat{\lambda}_n$ is the estimator of λ. This is a matter for further research and has been addressed in a forthcoming paper by the authors. In MCMC with proper target (that is, π is a probability measure) [17] discuss the estimation of the standard error of $\hat{\lambda}_n$ using the regeneration of the underlying Markov chain. In the present context of improper target distribution something similar based on standard error computed from the observed excursions for the underlying regenerative sequences could be tried. The results in [14] may be applicable to some of the Markov chains discussed here, although verifying Karlsen and Tjøstheim's (2001) conditions for functional limit theorems does not seem to be easy. In the case of reduction to a proper target distribution π_0 (that is, Algorithm IV) it is possible to establish the rate of decay of $(\hat{\lambda}_n - \lambda)$ under strong conditions on f as well as the mixing rates for the chain used (with stationary distribution π_0).

Appendices

Appendix A: Results on the existence and the uniqueness of invariant measure for Markov chains

Theorem 6. *Let $\{X_n\}_{n\geq 0}$ be a Markov chain with state space $(S, \mathcal{S})$. Let $\Delta \in S$ be a singleton such that $\{\Delta\} \in \mathcal{S}$. Let T be the hitting time (also known as the first passage time) for Δ, that is, $T = min\{n : 1 \leq n < \infty, X_n = \Delta\}$ and $T = \infty$ if $X_n \neq \Delta$ for all $n < \infty$. Let*

$$\tilde{\pi}(A) = E\left(\sum_{j=0}^{T-1} I_A(X_j)|X_0 = \Delta\right) \text{ for } A \in \mathcal{S}. \tag{A.1}$$

(i) (Existence of a subinvariant measure) Let $P_\Delta(\cdot)$ denote $P(\cdot|X_0 = \Delta)$, then we have

$$\tilde{\pi}(A) = I_A(\Delta)P_\Delta(T = \infty) + \int_S P(x, A)\tilde{\pi}(dx) \text{ for all } A \in \mathcal{S}, \tag{A.2}$$

and hence $\tilde{\pi}(\cdot)$ is a subinvariant measure with respect to P, that is,

$$\tilde{\pi}(A) \geq \int_S P(x, A)\tilde{\pi}(dx) \text{ for all } A \in \mathcal{S}. \tag{A.3}$$

(ii) (Uniqueness of the subinvariant measure) If $\breve{\pi}$ is a measure on $(S, \mathcal{S})$ that is subinvariant with respect to P, then

$$\breve{\pi}(A) \geq \breve{\pi}(\{\Delta\})\pi^*(A) \text{ for all } A \in \mathcal{S}, \tag{A.4}$$

where $\pi^*(A) := \tilde{\pi}(A) - I_A(\Delta)P_\Delta(T = \infty) = \int_S P(x, A)\tilde{\pi}(dx) = (\tilde{\pi}P)(A)$.

Corollary 3. *Let $\tilde{\pi}(\cdot)$ be as defined in (A.1). Then*

$$\tilde{\pi}(A) = \int_S P(x, A)\tilde{\pi}(dx) \text{ for all } A \in \mathcal{S}, \tag{A.5}$$

that is, $\tilde{\pi}(\cdot)$ is an invariant measure with respect to P if and only if Δ is recurrent, that is, $P_\Delta(T < \infty) = 1$. Further, Δ is recurrent if and only if $\tilde{\pi} = \pi^* = \tilde{\pi}P$. Also, if Δ is accessible from everywhere in S, that is, if for all $x \in S$, there exists $n(x) \geq 1$ such that $P^{n(x)}(x, \{\Delta\}) > 0$, then for every subinvariant measure $\breve{\pi}$ with $\breve{\pi}(\{\Delta\}) < \infty$, we have $\breve{\pi}(A) = \breve{\pi}(\{\Delta\})\tilde{\pi}(A)$ for all $A \in \mathcal{S}$.

A version of Theorem 6 can be found in Meyn and Tweedie [16, section 10.2]. For the sake of completeness we give a proof of Theorem 6 and Corollary 3 in Appendix B.

Appendix B: Proof of results

Proof of Theorem 1. Let $h : S \to \mathbb{R}_+$ be a nonnegative function. Recall that, $N_n = k$ if $T_k \le n < T_{k+1}, k, n = 0, 1, 2, \ldots$. Since $h \ge 0$,

$$\sum_{j=0}^{T_{N_n}} h(X_j) \le \sum_{j=0}^{n} h(X_j) \le \sum_{j=0}^{T_{N_n+1}-1} h(X_j).$$

Let $\xi_r = \sum_{j=T_r}^{T_{r+1}-1} h(X_j)$ for $r = 0, 1, 2, \ldots$. So we have

$$\frac{1}{N_n} \sum_{r=0}^{N_n-1} \xi_r \le \frac{1}{N_n} \sum_{j=0}^{n} h(X_j) \le \frac{1}{N_n} \sum_{r=0}^{N_n} \xi_r, \tag{B.1}$$

for all $n \ge 1$. Since $\{X_n\}_{n\ge 0}$ is a regenerative sequence, $\{\xi_i\}_{i\ge 0}$ are iid nonnegative random variables with $E(\xi_0) = \int_S h d\check{\pi}$, where $\check{\pi}$ is as defined in (2.1). Also, we have $N_n \to \infty$ as $n \to \infty$. By the strong law of large numbers, $\sum_{r=0}^{N_n} \xi_r / N_n \to E(\xi_0)$ as $n \to \infty$. Using (B.1) we then have

$$\int_S h d\check{\pi} \le \liminf_{n\to\infty} \frac{1}{N_n} \sum_{j=0}^{n} h(X_j) \le \limsup_{n\to\infty} \frac{1}{N_n} \sum_{j=0}^{n} h(X_j) \le \int_S h d\check{\pi}$$

and hence

$$\lim_{n\to\infty} \frac{1}{N_n} \sum_{j=0}^{n} h(X_j) = \int_S h d\check{\pi}. \tag{B.2}$$

Let f^+ and f^- are the positive and negative parts of the function f. Applying the above argument to f^+ and f^- separately yields (i) of Theorem 1. Next the proof of (ii) of the theorem follows by noting that

$$\frac{\sum_{j=0}^{n} f(X_j)}{\sum_{j=0}^{n} g(X_j)} = \frac{\sum_{j=0}^{n} f^+(X_j) - \sum_{j=0}^{n} f^-(X_j)}{\sum_{j=0}^{n} g^+(X_j) - \sum_{j=0}^{n} g^-(X_j)},$$

and applying (B.2) to each of the four terms in the above ratio. □

Proof of Theorem 2. To prove Theorem 2(ii), note that,

$$P(N_n \le k) = P(T_{k+1} > n) \text{ for all integers } k \ge 0 \text{ and } n \ge 0.$$

This implies that for $0 < x < \infty$,

$$\begin{aligned} P(N_n \le b_n x) = P(N_n \le \lfloor b_n x \rfloor) &= P\Big(T_{\lfloor b_n x\rfloor+1} > n\Big) \\ &= P\left(\frac{T_{\lfloor b_n x\rfloor+1}}{a_{\lfloor b_n x\rfloor+1}} > \frac{n}{a_{\lfloor b_n x\rfloor+1}}\right), \end{aligned}$$

where $\{a_n\}$ is as in (2.4) and $b_n > 0$ will be chosen shortly. Suppose $b_n \uparrow \infty$ is such that for each $0 < x < \infty$,

$$\frac{n}{a_{\lfloor b_n x\rfloor+1}} \to x^{-\frac{1}{\alpha}}. \tag{B.3}$$

Then since V_α has a continuous distribution, by (2.4) it follows that as $n \to \infty$

$$P\left(\frac{T_{\lfloor b_n x\rfloor+1}}{a_{\lfloor b_n x\rfloor+1}} > \frac{n}{a_{\lfloor b_n x\rfloor+1}}\right) \to P(V_\alpha > x^{-\frac{1}{\alpha}}).$$

(This uses the fact that if W_n is a sequence of random variables such that $W_n \xrightarrow{\mathrm{d}} W$ with $P(W = w_0) = 0$ and $w_n \to w_0$ then $P(W_n > w_n) \to P(W > w_0)$.) Now we find b_n so that (B.3) holds. Let $\Psi(x) \equiv P(\tau_1 > x)$. Then $\Psi(x)$ is monotone in x. Thus $\Psi^{-1}(\cdot)$ is well defined on $(0,1)$. Also, by hypothesis of Theorem 2, $\Psi(x) \sim x^{-\alpha}L(x)$ as $x \uparrow \infty$. We know that $na_n^{-\alpha}L(a_n) \to 1$, that is $n\Psi(a_n) \to 1$ as $n \to \infty$. This in turn implies that as $n \to \infty$, $\Psi(a_n) \sim 1/n$, or $a_n \sim \Psi^{-1}(1/n)$, that is $a_{\lfloor b_n x\rfloor+1} \sim \Psi^{-1}(1/\{\lfloor b_n x\rfloor + 1\})$. Thus (B.3) will hold if as $n \to \infty$,

$$\frac{n}{\Psi^{-1}(1/(\lfloor b_n x\rfloor + 1))} \to x^{-\frac{1}{\alpha}},$$

that is, $\Psi^{-1}(1/\lfloor b_n x\rfloor + 1) \sim n/x^{-\frac{1}{\alpha}}$, or $1/(\lfloor b_n x\rfloor + 1) \sim \Psi(nx^{\frac{1}{\alpha}}) \sim (nx^{\frac{1}{\alpha}})^{-\alpha}L(nx^{\frac{1}{\alpha}})$. This implies that as $n \to \infty$,

$$\frac{1}{b_n x} \sim \frac{n^{-\alpha}}{x}L(nx^{\frac{1}{\alpha}}) = \frac{n^{-\alpha}L(n)}{x}\frac{L(nx^{\frac{1}{\alpha}})}{L(n)}.$$

Since $L(\cdot)$ is slowly varying at ∞, the above equation will hold if $b_n = n^\alpha/L(n)$, thus establishing (B.3). Hence (2.6) is proved. □

Proof of Corollary 1. Let $\check{\lambda} \equiv \int_S f d\check{\pi}$. Note that $D_n^r \xrightarrow{\mathrm{d}} W_r$ as $n \to \infty$ where $W_r \overset{d}{=} \check{\lambda}Y_\alpha$, and W_r's are independent for all $r = 1, 2, \ldots, k$. Since $D_n^r, r = 1, \ldots, k$ are also independent, $(D_1^n, \ldots, D_k^n) \xrightarrow{\mathrm{d}} (W_1, \ldots, W_k)$, which, by continuous mapping theorem, implies that $\hat{\lambda}_{n,k} \equiv \sum_{r=1}^k D_n^r/(kE(Y_\alpha)) \xrightarrow{\mathrm{d}} \sum_{r=1}^k W_r/(kE(Y_\alpha)) \equiv \overline{W}_k$ as $n \to \infty$. By Skorohod's theorem, for each k, there exists $\tilde{\lambda}_{n,k} \overset{d}{=} \hat{\lambda}_{n,k}$ and $\tilde{W}_k \overset{d}{=} \overline{W}_k$, such that $\tilde{\lambda}_{n,k} \xrightarrow{\text{a.s.}} \tilde{W}_k$ as $n \to \infty$. Since by the strong law of large numbers, $\overline{W}_k \xrightarrow{\text{a.s.}} \check{\lambda}$ as $k \to \infty$, and $\tilde{W}_k \overset{d}{=} \overline{W}_k$, we have $\tilde{W}_k \xrightarrow{\mathrm{d}} \check{\lambda}$ as $k \to \infty$. Since $\check{\lambda}$ is a constant and convergence in probability implies almost surely convergence along a subsequence, there exists, k_j such that $\tilde{W}_{k_j} \xrightarrow{\text{a.s.}} \check{\lambda}$. So $\tilde{\lambda}_{n_{k_j},k_j} \xrightarrow{\text{a.s.}} \check{\lambda}$. However, $\tilde{\lambda}_{n,k} \overset{d}{=} \hat{\lambda}_{n,k}$ implying $\hat{\lambda}_{n_{k_j},k_j} \xrightarrow{\mathrm{d}} \check{\lambda}$ as $j \to \infty$. The corollary follows since there exists further subsequences $k_{j_l} \uparrow \infty$ and $n_{k_{j_l}} \uparrow \infty$ such that $\hat{\lambda}_{n_{k_{j_l}},k_{j_l}} \xrightarrow{\text{a.s.}} \check{\lambda}$ as $l \to \infty$. □

Proof of Theorem 3. Note that

$$\frac{1}{n}\sum_{i=1}^n f(X_i) = \frac{N_n}{n}\frac{1}{N_n}\sum_{i=1}^n f(X_i), \tag{B.4}$$

and by Theorem 1 we know that $\sum_{i=1}^n f(X_i)/N_n \xrightarrow{\text{a.s.}} \int_S f d\check{\pi}$. Let T_j and τ_j be as in Definition 1. Now

$$\frac{T_{N_n}}{N_n} = \frac{\sum_{j=1}^{N_n}\tau_j}{N_n} \le \frac{n}{N_n} \le \frac{\sum_{j=1}^{N_n+1}\tau_j}{N_n} = \frac{T_{N_n+1}}{N_n}$$

Since τ_i's are iid, by SLLN, $T_n/n \to E(\tau_1) = \check{\pi}(S) = \infty$. So $\lim_{n\to\infty} n/N_n$ exists and $\lim_{n\to\infty} n/N_n = \infty$ with probability 1, which implies that $N_n/n \xrightarrow{\text{a.s.}} 0$ and hence the proof follows from (B.4). □

Proof of Theorem 5. Let $T_0 = 0, T_{r+1} = \inf\{n : n \geq T_r + 1, X_n = 0\}$ for $r \geq 0$ be the successive return times of the SSRW $\{X_n\}_{n\geq 0}$ to the state 0. Let $h : \mathbb{R}^d \to \mathbb{R}_+$ be a nonnegative function. Since $\{Y_n\}_{n\geq 0}$ is regenerative with regeneration times T_r's, from (B.2) we have

$$\lim_{n\to\infty} \frac{1}{N_n} \sum_{j=0}^{n} h(Y_j) = E\Big(\sum_{j=0}^{T_1-1} h(Y_j)\Big). \tag{B.5}$$

Note that

$$\begin{aligned} E\Big(\sum_{j=0}^{T_1-1} h(Y_j)\Big) &= \sum_{j=0}^{\infty} E(h(Y_j)I(j < T_1)) \\ &= \sum_{j=0}^{\infty} E(E(h(\kappa(X_j) + U^j)I(j < T_1)|\{X_i\}_{i=0}^{j})) \\ &= \sum_{j=0}^{\infty} E(E(h(\kappa(X_j) + U^j)|\{X_i\}_{i=0}^{j})I(j < T_1)) \\ &= \sum_{j=0}^{\infty} E(\psi(X_j)I(j < T_1)), \end{aligned}$$

where $\psi : \mathbb{Z} \to \mathbb{R}$ is defined as $\psi(x) := \int_K h(\kappa(x)+u)du$, with $K = [-1/2, 1/2]^d \equiv [-1/2, 1/2] \times [-1/2, 1/2] \times \cdots d$ times. Since

$$\sum_{j=0}^{\infty} E(\psi(X_j)I(j < T_1)) = E\Big(\sum_{j=0}^{T_1-1} \psi(X_j)\Big) = \sum_{j\in\mathbb{Z}} \psi(j),$$

it follows that

$$E\Big(\sum_{j=0}^{T_1-1} h(Y_j)\Big) = \sum_{j\in\mathbb{Z}} \int_K h(\kappa(j) + u)du = \int_{\mathbb{R}^d} h(x)dx.$$

From (B.5) it follows that $\lim_{n\to\infty} \sum_{j=0}^{n} h(Y_j)/N_n = \int_{\mathbb{R}^d} h(x)dx$. Applying the above arguments to f^+p and f^-p separately yields (3.2). □

Proof of Lemma 1. The measure π_0 defined in (5.1) is a probability measure since

$$\pi_0(S) = \sum_{j\in J} p_j \frac{\pi(S \cap A_j)}{\pi(A_j)} = \sum_{j\in J} p_j = 1$$

Now the 1st part of the lemma follows from the fact that

$$\pi_0(A) = 0 \Rightarrow \pi(A \cap A_j) = 0 \text{ for all } j \in J \Rightarrow \pi(A) = 0.$$

In order to prove Lemma 1 (ii), note that

$$\pi(A) = \sum_{j \in J} \pi(A \cap A_j) = \sum_{j \in J} p_j \frac{\pi(A \cap A_j)}{\pi(A_j)} \frac{\pi(A_j)}{p_j}. \tag{B.6}$$

Since A_n's are disjoint, for any $j_0 \in J$,

$$\pi_0(A \cap A_{j_0}) = \sum_{j \in J} p_j \frac{\pi(A \cap A_{j_0} \cap A_j)}{\pi(A_j)} = p_{j_0} \frac{\pi(A \cap A_{j_0})}{\pi(A_{j_0})}.$$

So from (B.6) we have for all $A \in \mathcal{S}$

$$\begin{aligned}\pi(A) &= \sum_{j \in J} \frac{\pi(A_j)}{p_j} \pi_0(A \cap A_j) = \sum_{j \in J} \frac{\pi(A_j)}{p_j} \int_A I_{A_j}(\cdot) d\pi_0 \\ &= \int_A \sum_{j \in J} \frac{\pi(A_j)}{p_j} I_{A_j}(\cdot) d\pi_0,\end{aligned}$$

and hence the Radon-Nikodym derivative,

$$h = \frac{d\pi}{d\pi_0} = \sum_{j \in J} \frac{\pi(A_j)}{p_j} I_{A_j}(\cdot). \qquad \square$$

Proof of Theorem 6. Let $\tilde{\pi}(\cdot)$ be as defined in (A.1). Then for any $A \in \mathcal{S}$,

$$\begin{aligned}\tilde{\pi}(A) &= E\Big(\sum_{j=0}^{T-1} I_A(X_j) | X_0 = \Delta\Big) \\ &= I_A(\Delta) + E\Big(\sum_{j=1}^{\infty} I_A(X_j) I(T-1 \geq j) | X_0 = \Delta\Big).\end{aligned}$$

Now letting $E_\Delta(\cdot)$ denote $E(\cdot | X_0 = \Delta)$, we have

$$\begin{aligned}& E_\Delta\Big(\sum_{j=1}^{\infty} I_A(X_j) I(T-1 \geq j)\Big) \\ = \quad & E_\Delta\Big(\Big[\sum_{j=1}^{\infty} I_A(X_j) I(T > j-1) - \sum_{j=1}^{\infty} I_A(X_j) I(T = j)\Big]\Big).\end{aligned}$$

Thus

$$\tilde{\pi}(A) \quad = \quad I_A(\Delta) + \sum_{j=1}^{\infty} E_\Delta\Big(I_A(X_j) I(T > j-1)\Big) - \sum_{j=1}^{\infty} I_A(\Delta) P_\Delta(T = j)$$

$$
\begin{aligned}
&= I_A(\Delta) + \sum_{j=1}^{\infty} E_\Delta\Big(I_A(X_j)I(T > j-1)\Big) - I_A(\Delta)P_\Delta(T < \infty) \\
&= I_A(\Delta)P_\Delta(T = \infty) + \sum_{j=1}^{\infty} E_\Delta\Big(I_A(X_j)I(T > j-1)\Big). \qquad \text{(B.7)}
\end{aligned}
$$

Now for $j \geq 1$, $\{T > j-1\} = \{T \leq j-1\}^c$ belongs to $\sigma(X_0, X_1, \ldots, X_{j-1})$, the σ-algebra generated by $X_0, X_1, \ldots, X_{j-1}$. By the Markov property of $\{X_n\}_{n\geq 0}$, for $j \geq 1$,

$$
\begin{aligned}
E_\Delta\Big(I_A(X_j)I(T > j-1)\Big) &= E_\Delta\Big(I(T > j-1)E[I_A(X_j)|X_{j-1}]\Big) \\
&= E_\Delta\Big(I(T > j-1)P(X_{j-1}, A)\Big).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\sum_{j=1}^{\infty} E_\Delta\Big(I_A(X_j)I(T > j-1)\Big) &= E_\Delta\Big(\sum_{j=1}^{\infty} P(X_{j-1}, A)I(T > j-1)\Big) \\
&= E_\Delta\Big(\sum_{j=0}^{\infty} P(X_j, A)I(T > j)\Big) \\
&= \int_S P(x, A)\tilde{\pi}(dx),
\end{aligned}
$$

where the last equality follows from the definition of $\tilde{\pi}(\cdot)$ as in (A.1). So from (B.7) we get $\tilde{\pi}(A) = I_A(\Delta)P_\Delta(T = \infty) + \int_S P(x, A)\tilde{\pi}(dx)$, establishing (A.2), that is, proving Theorem 6 (i).

Next, let $\breve{\pi}$ be a σ-finite measure on $(S, \mathcal{S})$ that satisfies (A.3). We need to show that $\breve{\pi}(A) \geq \breve{\pi}(\{\Delta\})\pi^*(A)$ for all $A \in \mathcal{S}$. From (A.3) we have

$$
\breve{\pi}(A) \geq \breve{\pi}(\{\Delta\})P(\Delta, A) + \int_{S-\{\Delta\}} P(x, A)\breve{\pi}(dx).
$$

But,

$$
\begin{aligned}
\int_{S-\{\Delta\}} P(x, A)\breve{\pi}(dx) &\geq \int_{S-\{\Delta\}} P(x, A) \int_S P(y, dx)\breve{\pi}(dy) \\
&= \int_{S-\{\Delta\}} \int_{S-\{\Delta\}} P(y, dx)P(x, A)\breve{\pi}(dy) \\
&\quad + \breve{\pi}(\{\Delta\}) \int_{S-\Delta} P(x, A)P(\Delta, dx) \\
&= \breve{\pi}(\{\Delta\})P_\Delta(X_2 \in A, X_1 \in S - \{\Delta\}) \\
&\quad + \int_{S-\{\Delta\}} \int_{S-\{\Delta\}} P(y, dx)P(x, A)\breve{\pi}(dy).
\end{aligned}
$$

Iterating this we find that (A.3) implies that for $n \geq 2$,

$$
\breve{\pi}(A) \geq \breve{\pi}(\{\Delta\})(P_\Delta(X_1 \in A) + P_\Delta(X_2 \in A, T > 1) + \cdots + P_\Delta(X_n \in A, T > n-1)).
$$

So for $n \geq 2$,

$$\begin{aligned}
\breve{\pi}(A) &\geq \breve{\pi}(\{\Delta\})\Big(\sum_{j=1}^{n} P_\Delta(X_j \in A, T > j-1)\Big) \\
&= \breve{\pi}(\{\Delta\})\Big(\sum_{j=1}^{n} P_\Delta(X_j \in A, T > j) + \sum_{j=1}^{n} P_\Delta(T = j) I_A(\Delta)\Big) \\
&= \breve{\pi}(\{\Delta\})\Big(I_A(\Delta) P_\Delta(T \leq n) + \sum_{j=1}^{n} P_\Delta(X_j \in A, T > j)\Big).
\end{aligned}$$

Letting $n \to \infty$, we see that the right side above reduces to

$$\begin{aligned}
&\breve{\pi}(\{\Delta\})\Big(I_A(\Delta) P_\Delta(T < \infty) + \sum_{j=1}^{\infty} P_\Delta(X_j \in A, T > j)\Big) \\
= \;& \breve{\pi}(\{\Delta\})\Big(I_A(\Delta) + \sum_{j=1}^{\infty} P_\Delta(X_j \in A, T > j) - I_A(\Delta) P_\Delta(T = \infty)\Big) \\
= \;& \breve{\pi}(\{\Delta\}\Big(\sum_{j=0}^{\infty} P_\Delta(X_j \in A, T > j) - I_A(\Delta) P_\Delta(T = \infty)\Big) \\
= \;& \breve{\pi}(\{\Delta\}\Big(\tilde{\pi}(A) - I_A(\Delta) P_\Delta(T = \infty)\Big) = \breve{\pi}(\{\Delta\}\pi^*(A).
\end{aligned}$$

Thus $\breve{\pi}(A) \geq \breve{\pi}(\{\Delta\})\pi^*(A)$ for all $A \in \mathcal{S}$ establishing (A.4). □

Proof of Corollary 3. From (A.2) it follows that (A.5) holds if and only if Δ is recurrent.

Since Δ is accessible, $S = \cup_n D_n$ where $D_n = \{x : P^n(x, \{\Delta\}) > 0\}$. So to complete the proof of Corollary 3, it is sufficient to show that

$$\breve{\pi}(A) = \breve{\pi}(\{\Delta\})\tilde{\pi}(A) \;\text{ for all }\; A \subset D_n, n \geq 1.$$

If the above is false, then there exists $A_0 \in D_n$ for some $n \geq 1$ such that $\breve{\pi}(A_0) > \breve{\pi}(\{\Delta\})\tilde{\pi}(A_0)$ as by (A.4), $\breve{\pi}(A) \geq \breve{\pi}(\{\Delta\})\tilde{\pi}(A)$ for all $A \in \mathcal{S}$. Then

$$\begin{aligned}
\breve{\pi}(\{\Delta\}) &\geq \int_S P^n(x, \{\Delta\})\breve{\pi}(dx) \\
&= \int_{A_0} P^n(x, \{\Delta\})\breve{\pi}(dx) + \int_{S-A_0} P^n(x, \{\Delta\})\breve{\pi}(dx) \\
&> \breve{\pi}(\{\Delta\}) \int_{A_0} P^n(x, \{\Delta\})\tilde{\pi}(dx) + \breve{\pi}(\{\Delta\}) \int_{S-A_0} P^n(x, \{\Delta\})\tilde{\pi}(dx) \\
&= \breve{\pi}(\{\Delta\}) \int_S P^n(x, \{\Delta\})\tilde{\pi}(dx) = \breve{\pi}(\{\Delta\})\tilde{\pi}(\{\Delta\}),
\end{aligned}$$

where the first inequality follows by repeated application of (A.3) as $\breve{\pi}$ is assumed to be a subinvariant measure. The strict inequality follows from (A.4), and the facts that $\breve{\pi}(A_0) > \breve{\pi}(\{\Delta\})\tilde{\pi}(A_0)$ and $P^n(x, \{\Delta\}) > 0$ for all $x \in A_0$.

But $\tilde{\pi}(\{\Delta\}) = 1$ and so we get $\breve{\pi}(\{\Delta\}) > \breve{\pi}(\{\Delta\})$, which is a contradiction since $\breve{\pi}(\{\Delta\}) < \infty$. Thus $\breve{\pi}(A) = \breve{\pi}(\{\Delta\})\tilde{\pi}(A)$ for all $A \in \mathcal{S}$. □

Appendix C: Some technical results

Proposition 2. *Let $\{(X_n, Y_n)\}_{n\geq 0}$ be the Gibbs chain corresponding to the joint density $f(x, y)$ defined in (4.1) in Section 4. The Markov chain $\{(X_n, Y_n)\}_{n\geq 0}$ is regenerative and its subchains $\{X_n\}_{n\geq 0}$ and $\{Y_n\}_{n\geq 0}$ are both Harris recurrent Markov chains.*

Proof. Since $X_1|Y_0 \sim \mathrm{N}(Y_0, 1)$ and $Y_1|X_1 \sim \mathrm{N}(X_1, 1)$, we have $Y_1 = \eta_1 + \eta_2 + Y_0$, where (Y_0, η_1, η_2) are independent and η_1, η_2 are $\mathrm{N}(0, 1)$ random variables. By induction, $Y_n = Y_0 + \xi_1 + \cdots + \xi_n$ where $\{Y_0, \xi_1, \ldots, \xi_n\}$ are independent random variables and $\{\xi_i\}_{i\geq 1}$ are iid $\mathrm{N}(0, 2)$ random variables. Since $E|\xi_1|^2 < \infty$, $E(\xi_1) = 0$, and ξ_1 has an absolutely continuous distribution with mean 0, $\{Y_n\}_{n\geq 0}$ is Harris recurrent with Lebesgue measure as the reference measure [16, Proposition 9.4.5]. The Harris recurrence of $\{X_n\}_{n\geq 0}$ can be shown similarly. Let $B = (-1, 1)$. Since $\inf_{x\in B} P(Y_n \in B|X_n = x) = \inf_{x\in B}[\Phi(1+x)+\Phi(1-x)-1] > 0$ and Harris recurrence of $\{X_n\}_{n\geq 0}$ implies that $P(X_n \in B \text{ infinitely often}) = 1$ for any initial X_0, it follows that a recurrent atom can be constructed for $\{(X_n, Y_n)\}_{n\geq 0}$ and hence it is regenerative. □

Proposition 3. *The Gibbs sampler Markov chain $\{(X_n, Y_n)\}_{n\geq 0}$ corresponding to the joint density defined in (4.2) is regenerative.*

Proof. Since given (X_0, Y_0), the distribution of X_1 is exponential with mean $1/Y_0$, we can write $X_1 = (1/Y_0)\eta_1$, where η_1 is an exp(1) random variable and is independent of Y_0. Next, given X_1, the distribution of Y_1 is exponential with mean $1/X_1$ and hence we can write $Y_1 = \eta_2/X_1 = Y_0(\eta_2/\eta_1)$, where $\{Y_0, \eta_1, \eta_2\}$ are independent and η_1, η_2 are exp(1) random variables. By induction, $Y_n = Y_0\xi_1 \ldots \xi_n$ where $\{Y_0, \xi_1, \ldots, \xi_n\}$ are independent random variables and $\{\xi_i\}_{i\geq 1}$ are iid random variables such that ξ_1 has the same distribution as η_2/η_1. Since $E|\log \eta_1|^2 = \int_0^\infty |\log x|^2 \exp(-x)dx < \infty$, by Minkowski's inequality we have $E|\log \xi_1|^2 < \infty$. Also since $E|\log \eta_1| < \infty$, it follows that $E \log \xi_1 = E \log(\eta_2/\eta_1) = E \log(\eta_2) - E \log(\eta_1) = 0$. Now note that

$$\log Y_n = \log Y_0 + \sum_{i=1}^{n} \log \xi_i.$$

Then since $\log \xi_1$ has an absolutely continuous distribution with mean 0, as in Proposition 2, it implies that $\{\log Y_n\}_{n\geq 0}$ is Harris recurrent. This implies that $\{Y_n\}_{n\geq 0}$ is a Harris recurrent Markov chain with state space $(0, \infty)$. This in turn, as in the proof of Proposition 2, implies that $\{X_n, Y_n\}_{n\geq 0}$ is regenerative. □

Acknowledgments

The authors thank two reviewers and an associate editor for helpful comments and valuable suggestions which led to several improvements in the manuscript.

References

[1] Athreya, K. B. (1986). Darling and Kac revisited. *Sankhyā*, **48** 255–266. MR0905439

[2] Athreya, K. B. and Lahiri, S. N. (2006). *Measure Theory and Probability Theory*. Springer, New York. MR2247694

[3] Athreya, K. B. and Ney, P. (1978). A new approach to the limit theory of recurrent Markov chains. *Transactions of the American Mathematical Society*, **245** 493–501. MR0511425

[4] Breiman, L. (1992). *Probability*. SIAM, Philadelphia. MR1163370

[5] Casella, G. and George, E. (1992). Explaining the Gibbs sampler. *The American Statistician*, **46** 167–174. MR1183069

[6] Chung, K. L. (1968). *A Course in Probability Theory*. Harcourt, Brace & World, Inc., New York. MR0229268

[7] Durrett, R. (2010). *Probability: Theory and Examples*. 4th ed. Cambridge University Press, New York. MR2722836

[8] Feller, W. (1971). *An Introduction to Probability Theory and Its Applications*, vol. II. John Wiley & Sons, New York. MR0270403

[9] Gärtner, J. and Sun, R. (2009). A quenched limit theorem for the local time of random walks on $\mathbb{Z}^2$. *Stochastic Processes and Their Applications*, **119** 1198–1215. MR2508570

[10] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57** 97–109.

[11] Hobert, J. P. (2001). Stability relationships among the Gibbs sampler and its subchains. *Journal of Computational and Graphical Statistics*, **10** 185–205. MR1939696

[12] Hobert, J. P. and Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, **91** 1461–1473. MR1439086

[13] Hobert, J. P. and Casella, G. (1998). Functional compatibility, Markov chains, and Gibbs sampling with improper posteriors. *Journal of Computational and Graphical Statistics*, **7** 42–60. MR1628268

[14] Karlsen, H. A. and Tjøstheim, D. (2001). Nonparametric estimation in null recurrent time series. *The Annals of Statistics*, **29** 372–416. MR1863963

[15] Metropolis, N., Rosenbluth, A., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics*, **21** 1087–1092.

[16] Meyn, S. P. and Tweedie, R. L. (1993). *Markov Chains and Stochastic Stability*. Springer Verlag, London. MR1287609

[17] Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov chain samplers. *Journal of the American Statistical Association*, **90** 233–41. MR1325131

[18] Numelin, E. (1978). A splitting technique for Harris recurrent Markov chains. *Z. Wahrsch. Verw. Gebiete*, **43** 309–318. MR0501353

[19] Peköz, E. and Röllin, A. (2011). Exponential approximation for the nearly critical Galton-Watson process and occupation times of Markov chains. *Electronic Journal of Probability*, **16** 1381–1393. MR2827464

[20] R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

[21] Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. 2nd ed. Springer, New York. MR2080278

[22] Roberts, G. O., Sahu, S. K. and Wilks, W. R. (1995). Comment of "Bayesian computation and stochastic systems" by J. Besag, P. Green, D. Higdon and K. Mengersen. *Statistical Science*, **10** 49–51.

[23] Vineyard, G. H. (1963). The number of distinct sites visited in a random walk on a lattice. *Journal of Mathematical Physics*, **4** 1191–1193. MR0154664