Rigorous confidence bounds for MCMC under a geometric drift condition*

Krzysztof Łatuszyński and Wojciech Niemiro

K. Łatuszyński
Department of Statistics
University of Warwick
CV4 7AL, Coventry, UK
e-mail: latuch@gmail.com
url: http://www2.warwick.ac.uk/fac/sci/statistics/staff/research/latuszynski/

W. Niemiro
Faculty of Mathematics and Computer Science
Nicolaus Copernicus University
Chopina 12/18, 87-100 Toruń, Poland
e-mail: wniem@mat.uni.torun.pl

Abstract: We assume a drift condition towards a small set and bound the mean square error of estimators obtained by taking averages along a single trajectory of a Markov chain Monte Carlo algorithm. We use these bounds to construct fixed-width nonasymptotic confidence intervals. For a possibly unbounded function $f : \mathcal{X} \to R$, let $I = \int_{\mathcal{X}} f(x)\pi(x)dx$ be the value of interest and $\hat{I}_{t,n} = (1/n)\sum_{i=t}^{t+n-1} f(X_i)$ its MCMC estimate. Precisely, we derive lower bounds for the length of the trajectory n and burn-in time t which ensure that

$$P(|\hat{I}_{t,n} - I| \leq \varepsilon) \geq 1 - \alpha.$$

The bounds depend only and explicitly on drift parameters, on the $V-$norm of f, where V is the drift function and on precision and confidence parameters ε, α. Next we analyse an MCMC estimator based on the median of multiple shorter runs that allows for sharper bounds for the required total simulation cost. In particular the methodology can be applied for computing Bayesian estimators in practically relevant models. We illustrate our bounds numerically in a simple example.

AMS 2000 subject classifications: Primary 60J10, 65C05; secondary 62F15.
Keywords and phrases: MCMC estimation, confidence intervals, mean square error, Markov chain, convergence rate, V-uniform ergodicity, drift condition, simulation cost.

1. Introduction

An essential part of many problems in Bayesian inference is the computation of analytically intractable integral

$$I = \int_{\mathcal{X}} f(x)\pi(x)dx,$$

*Work partially supported by Polish Ministry of Science and Higher Education Grant No. N N201387234.

where $f(x)$ is the target function of interest, $\mathcal{X}$ is often a region in high-dimensional space and the probability distribution π over $\mathcal{X}$ is usually known up to a normalizing constant and direct simulation from π is not feasible (see e.g. [8], [27]). A common approach to this problem is to simulate an ergodic Markov chain $(X_n)_{n\geq 0}$, using a transition kernel P with stationary distribution π, which ensures that $X_n \to \pi$ in distribution. Thus, for a "large enough" n_0, X_n for $n \geq n_0$ is approximately distributed as π. Since a simple and powerful algorithm has been introduced in 1953 by Metropolis et al. in a very seminal paper [29], various sampling schemes and approximation strategies have been developed and analyzed ([27], [8]) and the method is referred to as Markov chain Monte Carlo (MCMC).

The standard method is to use average along a single trajectory of the underlying Markov chain and discard the initial part to reduce bias. In this case the estimate is of the form

$$\hat{I}_{t,n} = \frac{1}{n} \sum_{i=t}^{t+n-1} f(X_i) \tag{1}$$

and t is called the burn-in time. Asymptotic validity of (1) is ensured by a law of large numbers that holds in this setting under very mild assumptions [32]. Various results justify the choice of (1). In particular, for reversible chains, Geyer in [14] shows that subsampling is ineffective (in terms of asymptotic variance) and Chan and Yue in [9] consider asymptotic efficiency of (1) in a class of linear estimators (in terms of mean square error). Asymptotic behaviour of $\hat{I}_{t,n}$ is usually examined via a Central Limit Theorem (CLT) for Markov chains c.f. [14, 21, 32]. One constructs asymptotic confidence intervals, based on the CLT and consistent estimators of the asymptotic variance, as described in [14, 22, 18, 6]. Asymptotic behaviour of the mean square error of $\hat{I}_{0,n}$ in the $V-$uniformly ergodic setting has been also studied by Mathé in [28] using arguments from interpolation theory.

The goal of this paper is to derive explicit lower bounds for n and t in (1) that ensure the following condition:

$$P(|\hat{I}_{t,n} - I| \leq \varepsilon) \geq 1 - \alpha, \tag{2}$$

where ε is the precision of estimation and $1-\alpha$, the confidence level. We insist on obtaining bounds which depend only on ε, α and computable characteristics of the transition kernel P and function f. To decrease the total simulation cost, apart from $\hat{I}_{t,n}$, we also consider a nonlinear estimator based on the median of multiple shorter runs.

Results of this or related type have been obtained for finite or compact state space $\mathcal{X}$ and bounded target function f in [2, 16, 37]. Niemiro and Pokarowski in [31] give results for relative precision estimation. For uniformly ergodic chains and bounded function f, Hoeffding type inequalities are available in [17, 25, 26] and can easily lead to (2).

Tail inequalities for bounded functionals of Markov chains that are not uniformly ergodic were considered in [10], [1] and [11] using regeneration techniques.

Computing explicit bounds from these results may be possible with additional work, but we do not pursue it here.

If the target function f is not bounded and the Markov chain is not uniformly ergodic, rigorous nonasymptotic results about finite sample behaviour of $\hat{I}_{t,n}$ are scarce. Tail inequalities valid in this setup have been established by Bertail and Clémençon in [7] by regenerative approach and using truncation arguments. However, they involve non-explicit constants and can not be directly applied to derive lower bounds on t and n. In [24] a result analogous to (2) is established for a sequential-regenerative estimator (instead of $\hat{I}_{t,n}$). The approach of [24] requires identification of regeneration times. In many problems of practical interest, especially in high dimension, regeneration schemes are difficult to implement [15, 38].

Our approach is to assume a version of the well known drift condition towards a small set (Assumption 2.1 in Section 2), which is the typical setting when dealing with integrals of unbounded functions on noncompact sets. Under this assumption in Section 3 we bound the mean square error of $\hat{I}_{t,n}$. Our main Theorem 3.1 exploits the result of Baxendale [3]. In Section 4 we study confidence estimation (2) and obtain explicit lower bounds on n and t in terms of drift parameters defined in Assumption 2.1, the $V-$norm of f, where V is the drift function (for definitions see Section 1.1) and estimation parameters ε, α. Our bounds are designed to minimise the total simulation cost $t+n$. The estimation scheme is then refined via an elementary exponential inequality for a nonlinear estimator, a median of multiple shorter runs. In Section 5 we give an illustrative toy example.

The emphasis in our paper is on unbounded f, noncompact $\mathcal{X}$ and nonuniformly ergodic Markov chains, because this is a setting which usually arises when computing Bayesian estimators in many practically relevant models. We note that drift conditions required to apply our approach have been established in particular for the important hierarchical random effects models in [23] and for a more general family of linear models in [20].

1.1. Notation and Basic Definitions

Throughout this paper, π represents the probability measure of interest, defined on some measurable state space $(\mathcal{X},\mathcal{F})$ and $f:\mathcal{X}\to R$, the target function. Let $(X_n)_{n\geq 0}$ be a time homogeneous Markov chain on $(\mathcal{X},\mathcal{F})$ with transition kernel P. By π_0 denote its initial distribution and by π_t its distribution at time t. Let $I=\int_{\mathcal{X}} f(x)\pi(dx)$ be the value of interest and $\hat{I}_{t,n}=\frac{1}{n}\sum_{i=t}^{t+n-1} f(X_i)$ its MCMC estimate along one walk.

For a probability measure μ and a transition kernel Q, by μQ we denote a probability measure defined by $\mu Q(\cdot):=\int_{\mathcal{X}} Q(x,\cdot)\mu(dx)$. In this convention $\pi_t=\pi_0 P^t$. Furthermore if g is a real-valued function on $\mathcal{X}$, let $Qg(x):=\int_{\mathcal{X}} g(y)Q(x,dy)$ and $\mu g:=\int_{\mathcal{X}} g(x)\mu(dx)$. We will also use $E_\mu g$ for μg. If $\mu=\delta_x$ we will write E_x instead of E_μ. For transition kernels Q_1 and Q_2, Q_1Q_2 is also a transition kernel defined by $Q_1Q_2(x,\cdot):=\int_{\mathcal{X}} Q_2(y,\cdot)Q_1(x,dy)$.

Let $V : \mathcal{X} \to [1, \infty)$ be a measurable function. For a measurable function $g : \mathcal{X} \to R$ define its *V-norm* as

$$|g|_V := \sup_{x \in \mathcal{X}} \frac{|g(x)|}{V(x)}.$$

To evaluate the distance between two probability measures μ_1 and μ_2 we use the *V-norm distance*, defined as

$$\|\mu_1 - \mu_2\|_V := \sup_{|g| \leq V} |\mu_1 g - \mu_2 g| .$$

Note that for $V \equiv 1$ the $V-$norm distance $||\cdot||_V$ amounts to the well known total variation distance, precisely $\|\mu_1 - \mu_2\|_V = 2||\mu_1 - \mu_2||_{\text{tv}} := 2 \sup_{A \in \mathcal{F}} |\mu_1(A) - \mu_2(A)|$.

Finally for two transition kernels Q_1 and Q_2 the *V-norm distance* between Q_1 and Q_2 is defined by

$$|||Q_1 - Q_2|||_V := \big|\|Q_1(x, \cdot) - Q_2(x, \cdot)\|_V\big|_V = \sup_{x \in \mathcal{X}} \frac{\|Q_1(x, \cdot) - Q_2(x, \cdot)\|_V}{V(x)}.$$

For a probability distribution μ, define a transition kernel $\mu(x, \cdot) := \mu(\cdot)$, to allow for writing $|||Q - \mu|||_V$ and $|||\mu_1 - \mu_2|||_V$. Define also

$$B_V := \{f : f : \mathcal{X} \to R, |f|_V < \infty\}.$$

Now if $|||Q_1 - Q_2|||_V < \infty$, then $Q_1 - Q_2$ is a bounded operator from B_V to itself, and $|||Q_1 - Q_2|||_V$ is its operator norm. See [30] for details.

In the sequel we will work with geometrically ergodic Markov chains. A Markov chain is said to be *geometrically ergodic* if

$$\|\delta_x P^n - \pi\|_{\text{tv}} \leq M(x)\tilde{\gamma}^n, \quad \text{for} \quad \pi - \text{a.e. } x, \quad \text{and for some} \quad \tilde{\gamma} < 1.$$

In particular, if $M(x) \leq M$ then the chain is said to be uniformly ergodic. Geometric ergodicity is equivalent to existence of a drift function V towards a small set (see [32] and c.f. Assumption 2.1) and consequently also to $V-$*uniform ergodicity* which is defined by the following condition.

$$\|\delta_x P^n - \pi\|_V \leq MV(x)\gamma^n \quad \text{or equivalently} \quad |||P^n - \pi|||_V \leq M\gamma^n,$$

for some $M < \infty$ and some $\gamma < 1$.

2. A Drift Condition and Preliminary Lemmas

We analyze the MCMC estimation under the following assumption of a drift condition towards a small set, c.f. [3].

Assumption 2.1.

(A.1) Small set. There exist $C \subseteq \mathcal{X}$, $\tilde{\beta} > 0$ *and a probability measure* ν *on* $\mathcal{X}$, *such that for all* $x \in C$ *and* $A \subseteq \mathcal{X}$

$$P(x, A) \geq \tilde{\beta}\nu(A).$$

(A.2) Drift. There exist a function $V : \mathcal{X} \to [1, \infty)$ *and constants* $\lambda < 1$ *and* $K < \infty$ *satisfying*

$$PV(x) \leq \begin{cases} \lambda V(x), & \text{if} \quad x \notin C, \\ K, & \text{if} \quad x \in C. \end{cases}$$

(A.3) Strong Aperiodicity. There exists $\beta > 0$ *such that* $\tilde{\beta}\nu(C) \geq \beta$.

In the sequel we refer to $\tilde{\beta}, V, \lambda, K, \beta$ as drift parameters.

This type of drift condition is often assumed and widely discussed in Markov chains literature since it implies geometric ergodicity and a CLT for a convenient class of target functions, see [30] for details and definitions. Computable bounds for geometric ergodicity parameters under drift conditions allow to control the burn-in time t and the bias of MCMC estimators in practically relevant models. Substantial effort has been devoted to establishing such bounds, c.f. the survey paper by Roberts and Rosenthal [32] and references therein. Particular references include e.g. Rosenthal [35] or Roberts and Tweedie [34] for bounds on the total variation distance. Since we deal with unbounded functions, in the sequel we make use of the $V-$uniform ergodicity convergence bounds obtained by Baxendale in [3] (c.f. Douc *at al.* [12] and Fort [13]). In the drift condition setting and using explicit convergence bounds, our goal is to control not only the burn-in time t, but also the length of simulation n.

Theorem 2.2 ([30],[3])**.** *Under Assumption 2.1* $(X)_{n\geq 0}$ *has a unique stationary distribution* π *and* $\pi V < \infty$ *([30]). Moreover (Theorem 1.1 of [3]), there exists* $\rho < 1$ *depending only and explicitly on* $\tilde{\beta}, \beta, \lambda$ *and* K *such that whenever* $\rho < \gamma < 1$ *there exists* $M < \infty$ *depending only and explicitly on* $\gamma, \tilde{\beta}, \beta, \lambda$ *and* K *such that for all* $n \geq 0$

$$|||P^n - \pi|||_V \leq M\gamma^n. \tag{3}$$

Formulas for $\rho = \rho(\tilde{\beta}, \lambda, K, \beta)$ and $M = M(\gamma, \tilde{\beta}, \lambda, K, \beta)$ established in [3] are given in Appendix A and are used in Section 5. To our knowledge the abovementioned theorem gives the best available explicit constants. However this is a topic of ongoing research (c.f. [4]). We note that improving ergodicity constants in Theorem 2.2 will automatically result in tightening bounds established in our paper.

Corollary 2.3. *Under Assumption 2.1*

$$\|\pi_0 P^n - \pi\|_V \leq \min\{\pi_0 V, \|\pi_0 - \pi\|_V\} M\gamma^n,$$

where M *and* γ *are such as in Theorem 2.2.*

Proof. From Theorem 2.2 we have $\|P^n(x,\cdot)-\pi(\cdot)\|_V \leq M\gamma^n V(x)$, which yields

$$\begin{aligned}\pi_0 V M\gamma^n &\geq \int_{\mathcal{X}} \|P^n(x,\cdot)-\pi(\cdot)\|_V \pi_0(dx) \geq \sup_{|g|\leq V} \int_{\mathcal{X}} |P^n(x,\cdot)g-\pi g|\pi_0(dx)\\ &\geq \sup_{|g|\leq V} |\pi_0 P^n g - \pi g| = \|\pi_0 P^n - \pi\|_V.\end{aligned}$$

Now let $b_V = \inf_{x\in\mathcal{X}} V(x)$ and let μ_1, μ_2 be measures. Clearly $\|\mu_1(x,\cdot) - \mu_2(x,\cdot)\|_V$ is constant in x and therefore

$$|||\mu_1-\mu_2|||_V = \sup_x \frac{\|\mu_1(x,\cdot)-\mu_2(x,\cdot)\|_V}{V(x)} = \frac{\|\mu_1-\mu_2\|_V}{b_V}.$$

Since $|||\cdot|||_V$ is an operator norm and π is invariant for P, we have

$$\begin{aligned}\|\pi_0 P^n - \pi\|_V &= b_V|||\pi_0 P^n - \pi|||_V = b_V|||(\pi_0-\pi)(P^n-\pi)|||_V\\ &\leq b_V|||\pi_0-\pi|||_V|||P^n-\pi|||_V = \|\pi_0-\pi\|_V|||P^n-\pi|||_V.\\ &\leq \|\pi_0-\pi\|_V M\gamma^n.\end{aligned}$$

□

Next we focus on the following simple but useful observation.

Lemma 2.4. *If for a Markov chain $(X_n)_{n\geq 0}$ on $\mathcal{X}$ with transition kernel P Assumption 2.1 holds with parameters $\tilde{\beta}, V, \lambda, K, \beta$, it holds also with $\tilde{\beta}_r := \tilde{\beta}$, $V_r := V^{1/r}$, $\lambda_r := \lambda^{1/r}$, $K_r := K^{1/r}$, $\beta_r := \beta$ for every $r > 1$.*

Proof. It is enough to check (A.2). For $x \notin C$ by Jensen inequality we have

$$\lambda V(x) \geq \int_{\mathcal{X}} V(y)P(x,dy) \geq \left(\int_{\mathcal{X}} V(y)^{1/r} P(x,dy)\right)^r$$

and hence $PV_r(x) \leq \lambda^{1/r} V_r(x)$, as claimed. Similarly for $x \in C$ we obtain $PV_r(x) \leq K^{1/r}$. □

Lemma 2.4 together with Theorem 2.2 yield the following corollary.

Corollary 2.5. *Under Assumption 2.1 we have*

$$|||P^n-\pi|||_{V^{1/r}} \leq M_r\gamma_r^n,$$

where M_r and γ_r are constants defined as in Theorem 2.2 resulting from drift parameters defined in Lemma 2.4.

Integrating the drift condition with respect to π yields the following bound on πV.

Lemma 2.6. *Under Assumption 2.1*

$$\pi V \leq \pi(C)\frac{K-\lambda}{1-\lambda} \leq \frac{K-\lambda}{1-\lambda}.$$

Let $f_c = f - \pi f$. The next lemma provides a bound on $||f_c|^p|_V$ in terms of $||f|^p|_V$ without additional effort.

Lemma 2.7. *Under Assumption 2.1*

$$||f_c|^p|_V^{2/p} \leq \left(||f|^p|_V^{1/p} + \frac{\pi(C)}{b_V^{1/p}} K_{p,\lambda}\right)^2 \leq \left(||f|^p|_V^{1/p} + K_{p,\lambda}\right)^2,$$

where $b_V = \inf_{x\in\mathcal{X}} V(x)$ *and* $K_{p,\lambda} = \frac{K^{1/p}-\lambda^{1/p}}{1-\lambda^{1/p}}$.

Proof. Note that $\pi V^{1/p} \leq \pi(C)K_{p,\lambda} \leq K_{p,\lambda}$ by Lemma 2.6 and proceed:

$$\begin{aligned} ||f_c|^p|_V &= \sup_{x\in\mathcal{X}} \frac{|f(x)-\pi f|^p}{V(x)} \leq \sup_{x\in\mathcal{X}} \frac{\left(||f|^p|_V^{1/p} V^{1/p}(x) + \pi|f|\right)^p}{V(x)} \\ &\leq \sup_{x\in\mathcal{X}} \frac{\left(||f|^p|_V^{1/p} V^{1/p}(x) + \pi(C)K_{p,\lambda}\right)^p}{V(x)} \leq ||f|^p|_V \left(1 + \frac{\pi(C)K_{p,\lambda}}{b_V^{1/p}||f|^p|_V^{1/p}}\right)^p. \end{aligned}$$

□

3. MSE Bounds

By $MSE(\hat{I}_{t,n})$ we denote the mean square error of $\hat{I}_{t,n}$, i.e.

$$MSE(\hat{I}_{t,n}) = E_{\pi_0}[\hat{I}_{t,n} - I]^2.$$

Nonasymptotic bounds on $MSE(\hat{I}_{t,n})$ are essential to establish confidence estimation (2) and are also of independent interest. The main result of this section is the following

Theorem 3.1 (MSE Bounds)**.** *Assume the Drift Condition 2.1 holds and* $X_0 \sim \pi_0$. *Then for every measurable function* $f : \mathcal{X} \to R$, *every* $p \geq 2$ *and every* $r \in [\frac{p}{p-1}, p]$

$$MSE(\hat{I}_{0,n}) \leq \frac{||f_c|^p|_V^{2/p}}{n} \left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right) \left(\pi V + \frac{M \min\{\pi_0 V, \|\pi_0 - \pi\|_V\}}{n(1-\gamma)}\right), \tag{4}$$

where $f_c = f - \pi f$ *and constants* M, γ, M_r, γ_r *depend only and explicitly on* $\tilde{\beta}, \beta, \lambda$ *and* K *from Assumption 2.1 as in Theorem 2.2 and Corollary 2.4.*

We emphasise the most important special case for $p = r = 2$ as a corollary.

Corollary 3.2. *In the setting of Theorem 3.1, we have in particular*

$$MSE(\hat{I}_{0,n}) \leq \frac{|f_c^2|_V}{n} \left(1 + \frac{2M_2\gamma_2}{1-\gamma_2}\right) \left(\pi V + \frac{M \min\{\pi_0 V, \|\pi_0 - \pi\|_V\}}{n(1-\gamma)}\right). \tag{5}$$

Remark 3.3. The formulation of the foregoing Theorem 3.1 is motivated by a trade-off between small V and small λ in Assumption 2.1. It should be intuitively clear that establishing the drift condition for a quickly increasing V should result in smaller λ at the cost of bigger πV. So it may be reasonable to look for a valid drift condition with $V \geq C||f_c|^p|$ for some $p > 2$ instead of the natural choice of $p = 2$. Lemma 2.4 should strengthen this intuition.

Remark 3.4. For evaluating $\min\{\pi_0 V, \|\pi_0 - \pi\|_V\}$ one will often use the obvious bound $\min\{\pi_0 V, \|\pi_0 - \pi\|_V\} \leq \pi_0 V$, because $\pi_0 V$ depends on π_0 which is users choice, e.g. a deterministic point. Also, in some cases a fairly small bound for πV should be possible to obtain by direct calculations, e.g. if π is exponentially concentrated and V is a polynomial of degree 2. However, in absence of a better bound for πV, Lemma 2.6 is at hand. Similarly Lemma 2.7 bounds the unknown value $||f_c|^p|_V^{2/p}$ in terms of $||f|^p|_V$. Note that in applications both f and V have explicit formulas known to the user and $||f|^p|_V$ can be evaluated directly or easily bounded.

Remark 3.5. Let $\sigma^2_{\text{as}}(f)$ denote the asymptotic variance from the CLT for Markov chains (see e.g. [32, 5]). Since in the drift condition setting

$$\frac{nMSE(\hat{I}_{0,n})}{\sigma^2_{\text{as}}(f)} \to 1 \quad \text{as} \quad n \to \infty,$$

we see that the bounds in Theorem 3.1 and Corollary 3.2 have the correct asymptotic dependence on n and are easy to interpret. In particular $\pi V|f_c^2|_V$ in Corollary 3.2 should be close to $Var_\pi f$ for an appropriate choice of V, the term $2M_2\gamma_2/(1-\gamma_2)$ corresponds to the autocorrelation of the chain and $M\min\{\pi_0 V, \|\pi_0 - \pi\|_V\}/n(1-\gamma)$ is the price for nonstationarity of the initial distribution. In fact Theorem 3.1 with $\pi_0 = \pi$ yields the following bound on the asymptotic variance

$$\sigma^2_{\text{as}}(f) = \lim_{n\to\infty} nE_\pi[\hat{I}_{0,n} - I]^2 \leq \pi V||f_c|^p|_V^{2/p}\left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right).$$

Proof of Theorem 3.1. Note that $|f|^r_{V^{1/r}} = ||f|^r|_V$. Without loss of generality consider f_c instead of f and assume $||f_c|^p|_V = 1$. In this setting $|f_c^2|_V \leq 1$, $Var_\pi f_c = \pi f_c^2 \leq \pi V$, $MSE(\hat{I}_{0,n}) = E_{\pi_0}(\hat{I}_{0,n})^2$, and also for every $r \in [\frac{p}{p-1}, p]$,

$$|f_c|_{V^{1/r}} \leq ||f_c|^{p/r}|_{V^{1/r}} = 1 \quad \text{and} \quad |f_c|_{V^{1-1/r}} \leq ||f_c|^{p-p/r}|_{V^{1-1/r}} = 1.$$

Obviously

$$nMSE(\hat{I}_{0,n}) \quad = \quad \frac{1}{n}\sum_{i=0}^{n-1} E_{\pi_0} f_c(X_i)^2 + \frac{2}{n}\sum_{i=0}^{n-2}\sum_{j=i+1}^{n-1} E_{\pi_0} f_c(X_i) f_c(X_j). \quad (6)$$

We start with a bound for the first term of the right hand side of (6). Since $f_c^2(x) \leq V(x)$, we use Corollary 2.3 for f_c^2. Let $C = \min\{\pi_0 V, \|\pi_0 - \pi\|_V\}$ and

proceed

$$\frac{1}{n}\sum_{i=0}^{n-1} E_{\pi_0} f_c(X_i)^2 = \frac{1}{n}\sum_{i=0}^{n-1} \pi_0 P^i f_c^2 \leq \pi f_c^2 + \frac{1}{n}\sum_{i=0}^{n-1} CM\gamma^i \leq \pi V + \frac{CM}{n(1-\gamma)}. \quad (7)$$

To bound the second term of the right hand side of (6) note that $|f_c| \leq V^{1/r}$ and use Corollary 2.5.

$$\begin{aligned}
\frac{2}{n}\sum_{i=0}^{n-2}\sum_{j=i+1}^{n-1} E_{\pi_0} f_c(X_i) f_c(X_j) &= \frac{2}{n}\sum_{i=0}^{n-2}\sum_{j=i+1}^{n-1} \pi_0\left(P^i\left(f_c P^{j-i} f_c\right)\right) \\
&\leq \frac{2}{n}\sum_{i=0}^{n-2}\sum_{j=i+1}^{n-1} \pi_0\left(P^i\left(|f_c||P^{j-i} f_c|\right)\right) \\
&\leq \frac{2M_r}{n}\sum_{i=0}^{n-2}\sum_{j=i+1}^{\infty} \gamma_r^{j-i}\pi_0\left(P^i\left(|f_c|V^{1/r}\right)\right) \\
&\leq \frac{2M_r\gamma_r}{n(1-\gamma_r)}\sum_{i=0}^{n-2} \pi_0\left(P^i\left(|f_c|V^{1/r}\right)\right) = \spadesuit
\end{aligned}$$

Since $|f_c| \leq V^{1/r}$ and $|f_c| \leq V^{1-1/r}$, also $|f_c V^{1/r}| \leq V$ and we use Corollary 2.3 for $|f_c|V^{1/r}$.

$$\spadesuit \leq \frac{2M_r\gamma_r}{n(1-\gamma_r)}\sum_{i=0}^{n-2}\left(\pi\left(|f_c|V^{1/r}\right) + CM\gamma^i\right) \leq \frac{2M_r\gamma_r}{1-\gamma_r}\left(\pi V + \frac{CM}{n(1-\gamma)}\right). \quad (8)$$

Combine (7) and (8) to obtain

$$MSE(\hat{I}_{0,n}) \leq \frac{||f_c|^p|_V^{2/p}}{n}\left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right)\left(\pi V + \frac{CM}{n(1-\gamma)}\right).$$

□

Theorem 3.1 is explicitly stated for $\hat{I}_{0,n}$, but the structure of the bound is flexible enough to cover most typical settings as indicated below.

Corollary 3.6. *In the setting of Theorem 3.1,*

$$MSE(\hat{I}_{0,n}) \leq \frac{\pi V||f_c|^p|_V^{2/p}}{n}\left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right), \quad \text{if} \quad \pi_0 = \pi, \quad (9)$$

$$MSE(\hat{I}_{0,n}) \leq \frac{||f_c|^p|_V^{2/p}}{n}\left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right)\left(\pi V + \frac{MV(x)}{n(1-\gamma)}\right), \quad \text{if} \quad \pi_0 = \delta_x, \quad (10)$$

$$MSE(\hat{I}_{t,n}) \leq \frac{||f_c|^p|_V^{2/p}}{n}\left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right)\left(\pi V + \frac{M^2\gamma^t V(x)}{n(1-\gamma)}\right), \quad \text{if} \quad \pi_0 = \delta_x. \quad (11)$$

Proof. Only (11) needs a proof. Note that $X_t \sim \delta_x P^t$. Now use Theorem 2.2 to see that $\|\delta_x P^t - \pi\|_V \leq M\gamma^t V(x)$, and apply Theorem 3.1 with $\pi_0 = \delta_x P^t$. □

Bound (9) corresponds to the situation when a perfect sampler is available and used instead of burn-in. For deterministic start without burn-in and with burn-in, (10) and (11) should be applied respectively.

4. Confidence Estimation

Confidence estimation is an easy corollary of MSE bounds by the Chebyshev inequality.

Theorem 4.1 (Confidence Estimation)**.** *Under Assumption 2.1, let*

$$b = \frac{\pi V|f_c|^p|_V^{2/p}}{\varepsilon^2\alpha}\left(1+\frac{2M_r\gamma_r}{1-\gamma_r}\right), \tag{12}$$

$$c = \frac{M\min\{\pi_0 V, \|\pi_0-\pi\|_V\}||f_c|^p|_V^{2/p}}{\varepsilon^2\alpha(1-\gamma)}\left(1+\frac{2M_r\gamma_r}{1-\gamma_r}\right), \tag{13}$$

$$c(t) = \frac{M^2\gamma^t V(x)||f_c|^p|_V^{2/p}}{\varepsilon^2\alpha(1-\gamma)}\left(1+\frac{2M_r\gamma_r}{1-\gamma_r}\right), \tag{14}$$

$$n(t) = \frac{b+\sqrt{b^2+4c(t)}}{2}, \tag{15}$$

$$\tilde{c} = \frac{M^2 V(x)||f_c|^p|_V^{2/p}}{\varepsilon^2\alpha(1-\gamma)}\left(1+\frac{2M_r\gamma_r}{1-\gamma_r}\right). \tag{16}$$

Then

$$P(|\hat{I}_{0,n}-I|\leq\varepsilon)\geq 1-\alpha, \quad \textit{if} \quad X_0\sim\pi_0, \quad n\geq\frac{b+\sqrt{b^2+4c}}{2}. \tag{17}$$

$$P(|\hat{I}_{t,n}-I|\leq\varepsilon)\geq 1-\alpha, \quad \textit{if} \quad \begin{cases} X_0\sim\delta_x, \\ t\geq\max\left\{0,\log_\gamma\left(\frac{2+\sqrt{4+b^2\ln^2\gamma}}{\tilde{c}\ln^2\gamma}\right)\right\}, \\ n\geq n(t). \end{cases} \tag{18}$$

Remark 4.2 (Leading term). The above bounds in (18) give the minimal length of the trajectory $(t+n)$ resulting from (11). The leading term of the bound on n is

$$b = \frac{\pi V|f_c^2|_V}{\varepsilon^2\alpha}(1+\frac{2M_2\gamma_2}{1-\gamma_2})$$

(where we took $p=r=2$ for simplicity). Quantity $\pi V|f_c^2|_V$ should be of the same order as $Var_\pi f$, thus a term of this form is inevitable in any bound on n. Next, ε^{-2} which results from Chebyshev's inequality, is typical and inevitable, too. The factor α^{-1} will be reduced later in this section to $\log(\alpha^{-1})$ for small α by Lemma 4.4 and Algorithm 4.5. The term $1+\frac{2M_2\gamma_2}{1-\gamma_2}$ which roughly speaking bounds the autocorrelation of the chain, is the bottleneck of the approach. Here

good bounds on γ and the somewhat disregarded in literature $M = M(\gamma)$ are equally important. Improvements in Baxendale-type convergence bounds may lead to dramatic improvement of the bounds on the total simulation cost (e.g. by applying the preliminary results of [4]).

Remark 4.3. The formulation of Theorem 4.1 indicates how the issue of a sufficient burn-in should be understood. The common approach is to describe t as *time to stationarity* and to require that $t^* = t(x, \tilde{\varepsilon})$ should be such that $\rho(\pi, \delta_x P^{t^*}) \leq \tilde{\varepsilon}$ (where $\rho(\cdot,\cdot)$ is a distance function for probability measures, e.g. total variation distance, or $V-$norm distance). This approach seems not appropriate for such a natural goal as fixed precision of estimation at fixed confidence level. The optimal burn-in time can be much smaller then t^* and in particular cases it can be 0. Also we would like to emphasise that in the typical drift condition setting, i.e. if $\mathcal{X}$ is not compact and the target function f is not bounded, $||\pi_t - \pi||_{\text{tv}} \to 0$ does not even imply $\pi_t f \to \pi f$. Therefore a $V-$norm with $|f|_V < \infty$ should be used as a measure of convergence.

Proof of Theorem 4.1. From the Chebyshev's inequality we get

$$\begin{aligned} P(|\hat{I}_{t,n} - I| \leq \varepsilon) &= 1 - P(|\hat{I}_{t,n} - I| \geq \varepsilon) \\ &\geq 1 - \frac{MSE(\hat{I}_{t,n})}{\varepsilon^2} \geq 1 - \alpha \quad \text{if} \quad MSE(\hat{I}_{t,n}) \leq \varepsilon^2 \alpha. \end{aligned} \tag{19}$$

To prove (17) set $C = \min\{\pi_0 V, \|\pi_0 - \pi\|_V\}$, and combine (19) with (4) to get

$$n^2 - n\frac{\pi V||f_c|^p|_V^{2/p}}{\varepsilon^2\alpha}\left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right) - \frac{MC||f_c|^p|_V^{2/p}}{\varepsilon^2\alpha(1-\gamma)}\left(1 + \frac{2M_r\gamma_r}{1-\gamma_r}\right) \geq 0,$$

and hence $n \geq \frac{b+\sqrt{b^2+4c}}{2}$, where b and c are defined by (12) and (13) respectively. The only difference in (18) is that now we have $c(t)$ defined by (14) instead of c. It is easy to check that the best bound on t and n (i.e. which minimizes $t+n$) is such that

$$n \geq n(t) \qquad \text{and} \qquad t \geq \max\left\{0, \min\{t \in N : n'(t) \geq -1\}\right\},$$

where $n(t)$ is defined by (15) and $n'(t) = \frac{d}{dt}n(t)$. Standard calculations show that

$$\min\{t \in N : n'(t) \geq -1\} = \min\{t \in N : (\gamma^t)^2 \tilde{c}^2 \ln^2\gamma - \gamma^t 4\tilde{c} - b^2 \leq 0\},$$

where $\tilde{c}$ is defined by (16). Hence we obtain

$$t \geq \max\left\{0, (\ln\gamma)^{-1} \ln\left(\frac{2 + \sqrt{4 + b^2\ln^2\gamma}}{\tilde{c}\ln^2\gamma}\right)\right\} \qquad \text{and} \qquad n \geq n(t).$$

This completes the proof. □

Next we consider an alternative nonlinear estimation scheme, the so called "median trick" (introduced in [19] in the computational complexity context and further developed in [31]) that allows for sharper bounds for the total simulation cost needed to obtain confidence estimation for small α. The following simple lemma is taken from a more general setting of Section 2 in [31].

Lemma 4.4. *Let $m \in N$ be an odd number and let $\hat{I}_1, \ldots, \hat{I}_m$ be independent random variables, such that $P(|\hat{I}_k - I| \leq \varepsilon) \geq 1 - a > 1/2$, for $k = 1, \ldots, m$. Define $\hat{I} := \text{med}\{\hat{I}_1, \ldots, \hat{I}_m\}$. Then*

$$P(|\hat{I} - I| \leq \varepsilon) \geq 1 - \alpha, \quad \text{if} \quad m \geq \frac{2\ln(2\alpha)}{\ln[4a(1-a)]}. \tag{20}$$

Hence confidence estimation with parametesrs ε, α can be obtained by the following Algorithm 4.5.

Algorithm 4.5 (MA: median of averages)**.**

1. *Simulate m independent runs of length $t + n$ of the underlying Markov chain,*
$$X_0^{(k)}, \ldots, X_{t+n-1}^{(k)}, \quad k = 1, \ldots, m.$$
2. *Calculate m estimates of I, each based on a single run,*
$$\hat{I}_k = \hat{I}_{t,n}^{(k)} = \frac{1}{n} \sum_{i=t}^{t+n-1} f(X_i^{(k)}), \quad k = 1, \ldots, m.$$
3. *For the final estimate take*
$$\hat{I} = \text{med}\{\hat{I}_1, \ldots, \hat{I}_m\}.$$

Theorem 4.1 should be used to find t and n that guarantee confidence estimation with parameters ε, a and m results from Lemma 4.4. The total cost of Algorithm 4.5 amounts to $m(t + n)$ and depends on a (in addition to previous parameters). The optimal a can be found numerically, however $a = 0.11969$ is an acceptable arbitrary choice (cf. [31]).

5. A Toy Example - Contracting Normals

To illustrate the results of previous sections we analyze the *contracting normals* example studied by Baxendale in [3] (see also [34], [33] and [36]), where Markov chains with transition probabilities $P(x, \cdot) = N(\theta x, 1 - \theta^2)$ for some parameter $\theta \in (-1, 1)$ are considered.

Similarly as in [3] we take a drift function $V(x) = 1 + x^2$ and a small set $C = [-d, d]$ with $d > 1$, which allows for $\lambda = \theta^2 + \frac{2(1-\theta^2)}{1+d^2} < 1$ and $K = 2 + \theta^2(d^2 - 1)$. We also use the same minorization condition with ν concentrated on C, such that $\tilde{\beta}\nu(dy) = \min_{x \in C}(2\pi(1-\theta^2))^{-1/2} \exp(-\frac{(\theta x - y)^2}{2(1-\theta^2)})dy$. This yields

$\tilde{\beta} = 2[\Phi(\frac{(1+|\theta|)d}{\sqrt{1-\theta^2}}) - \Phi(\frac{|\theta|d}{\sqrt{1-\theta^2}})]$, where Φ denotes the standard normal cumulative distribution function.

Baxendale in [3] indicated that the chain is reversible with respect to its invariant distribution $\pi = N(0,1)$ for all $\theta \in (-1,1)$ and it is reversible and positive for $\theta > 0$. Moreover, in Lemma 5.1 we observe a relationship between marginal distributions of the chain with positive and negative values of θ. By $\mathcal{L}(X_n|X_0,\theta)$ denote the distribution of X_n given the starting point X_0 and the parameter value θ.

Lemma 5.1.

$$\mathcal{L}(X_n|X_0,\theta) = \mathcal{L}(X_n|(-1)^n X_0, -\theta). \tag{21}$$

Proof. Let $Z_1, Z_2, \ldots$ be an iid $N(0,1)$ sequence, then

$$\begin{aligned}
\mathcal{L}(X_n|X_0,\theta) &= \mathcal{L}\Big(\theta^n X_0 + \sum_{k=1}^{n} \theta^{n-k}\sqrt{1-\theta^2} Z_k\Big) \\
&= \mathcal{L}\Big((-\theta)^n(-1)^n X_0 + \sum_{k=1}^{n} (-\theta)^{n-k}\sqrt{1-\theta^2} Z_k\Big) \\
&= \mathcal{L}(X_n|(-1)^n X_0, -\theta),
\end{aligned}$$

and we used the fact that Z_k and $-Z_k$ have the same distribution. □

Therefore, if $\theta \geq 0$ then from Theorem 2.2 we have

$$||\mathcal{L}(X_n|X_0,\theta) - \pi||_V \leq M\gamma^n V(X_0) = M\gamma^n(1+X_0^2), \tag{22}$$

with M and γ computed for reversible and positive Markov chains (see Appendix A.3 for formulas). For $\theta < 0$ we get the same bound (22) with exactly the same M, γ by Lemma 5.1 and the fact that $V(x)$ is symmetric.

The choice of $V(x) = 1 + x^2$ allows for confidence estimation of $\int_{\mathcal{X}} f(x)\pi(dx)$ if $|f^2|_V < \infty$ for the possibly unbounded function f. In particular the MCMC works for all linear functions on $\mathcal{X}$. We take $f(x) = x$ where $|f^2|_V = 1$ as an example. We have to provide parameters and constants required for Theorem 4.1. In this case the optimal starting point is $X_0 = 0$ since it minimizes $V(x)$. Although in this example we can compute $\pi V = 2$ and $|f_c^2|_V = 1$, we also consider bounding πV and $|f_c^2|_V$ using Lemma 2.6 and Lemma 2.7 respectively.

		setting 1				setting 2				reality			
α	algorithm	m	t	n	total cost	m	t	n	total cost	m	t	n	total cost
.1	one walk	1	218	6.46e+09	6.46e+09	1	229	1.01e+08	1.01e+08	1	0	811	811
	MA	-	-	-	-	-	-	-	-	-	-	-	-
10^{-3}	one walk	1	218	6.46e+11	6.46e+11	1	229	1.01e+10	1.01e+10	1	0	3248	3248
	MA	15	218	5.40e+09	8.10e+10	15	229	8.45e+07	1.27e+09	7	0	726	5082
10^{-5}	one walk	1	218	6.46e+13	6.46e+13	1	229	1.01e+12	1.01e+12	1	0	5853	5853
	MA	27	218	5.40e+09	1.46e+11	27	229	8.45e+07	2.28e+09	13	0	726	9438

Table 1. Bounds for the *one walk algorithm* and the *median of averages Algorithm 4.5* (MA) for $\theta = .5$, precision parameter $\varepsilon = .1$ and different values of the confidence parameter α. Baxendale's $V-$uniform ergodicity parameters in this example are $\rho = .895$, $\rho_2 = .899$. Optimizing the total simulation cost results in $\gamma = .915$, $\gamma_2 = .971$, $M = 3.64e+04$, $M_2 = 748$. **Setting 1** uses Lemmas 2.6 and 2.7, whereas in **setting 2**, πV and $|f_c^2|_V$ are computed directly. The bounds are compared to **reality** obtained empirically in a simulation study.

Examples of bounds for t and n for the one walk estimator, or t, n and m for the median of averages (MA) estimator are given in Table 1. The bounds are computed for $C = [-d, d]$ with $d = 1.6226$ which minimizes ρ_2 (rather than ρ) for $\theta = 0.5$. Then a grid search is performed to find optimal values of γ and γ_2 that minimize the total simulation cost. Note that in Baxendale's result, the constant M depends on γ and goes relatively quickly to ∞ as $\gamma \to \rho$. This is the reason why optimal γ and γ_2 are far from ρ and ρ_2 and turns out to be the bottleneck of Baxendale's bounds in applications (c.f. Remark 4.2). Also for small $\alpha = 10^{-5}$, the $m = 27$ shorter runs have a significantly lower bound on the required total simulation effort then the single long run. MA is thus more mathematically tractable. However, in reality MA is about $\pi/2$ times less efficient then the one walk estimator - a phenomenon that can be inferred from the standard asymptotic theory.

R functions for computing this example and also the general bounds resulting from Theorem 4.1 are available at http://www2.warwick.ac.uk/fac/sci/statistics/staff/research/latuszynski/

6. Concluding Remarks

The main message of our paper is a very positive one: current theoretical knowledge of Markov chains reached the stage when for many MCMC algorithms of practical relevance applied to difficult problems, i.e. estimating expectations of unbounded functions, we are able to provide a rigorous, nonasymptotic, a priori analysis of the quality of estimation. This is much more then the often used in practice visual assessment of convergence by looking at a graph, more sophisticated a posteriori convergence diagnostics, bounding only burn in time or even using asymptotic confidence intervals, and should replace it, where possible.

The bounds derived in our paper are admittedly conservative, as observed in Section 5. We note that this is the case also for explicit bounds on convergence in total variation norm established under drift conditions. Nevertheless drift conditions remain the main and most universal tool in obtaining nonasymptotic results for general state space Markov chains.

For regenerative algorithms alternative bounds established in [24] are typically tighter then those resulting from our Section 4. However, the algorithms proposed there are more difficult to implement in practically relevant examples.

Appendix A: Formulas for ρ and M

For the convenience of the reader we repeat here the formulas from [3] that play a key role in our considerations.

In the sequel the term *atomic case* and *nonatomic case* refers to $\tilde{\beta}=1$ and $\tilde{\beta}<1$ respectively. If $\tilde{\beta}<1$, define

$$\alpha_1 = 1+\frac{\log\frac{K-\tilde{\beta}}{1-\beta}}{\log\lambda^{-1}}, \quad \alpha_2 = \begin{cases} 1, & \text{if } \nu(C)=1, \\ 1+\frac{\log\tilde{K}}{\log\lambda^{-1}}, & \text{if } \nu(C)+\int_{C^c}V d\nu \le \tilde{K}, \\ 1+\left(\log\frac{K}{\tilde{\beta}}\right)/(\log\lambda^{-1}), & \text{otherwise.} \end{cases}$$

Then let

$$R_0 = \min\{\lambda^{-1},(1-\tilde{\beta})^{-1/\alpha_1}\}, \qquad L(R) = \begin{cases} \frac{\tilde{\beta}R^{\alpha_2}}{1-(1-\tilde{\beta})R^{\alpha_1}}, & \text{if } \ 1<R<R_0, \\ \infty & \text{if } \ R=R_0. \end{cases}$$

A.1. Formulas for general operators

For $\beta>0$, $R>1$ and $L>1$, let $R_1=R_1(\beta,R,L)$ be the unique solution $r\in(1,R)$ of the equation

$$\frac{r-1}{r(\log(R/r))^2} = \frac{e^2\beta(R-1)}{8(L-1)}$$

and for $1<r<R_1$, define

$$K_1(r,\beta,R,L) = \frac{2\beta+2(\log N)(\log(R/r))^{-1}-8Ne^{-2}(r-1)r^{-1}(\log(R/r))^{-2}}{(r-1)[\beta-8Ne^{-2}(r-1)r^{-1}(\log(R/r))^{-2}]},$$

where $N=(L-1)/(R-1)$.

For the *atomic case* we have $\rho=1/R_1(\beta,\lambda^{-1},\lambda^{-1}K)$ and for $\rho<\gamma<1$,

$$\begin{aligned} M &= \frac{\max(\lambda,K-\lambda/\gamma)}{\gamma-\lambda}+\frac{K(K-\lambda/\gamma)}{\gamma(\gamma-\lambda)}K_1(\gamma^{-1},\beta,\lambda^{-1},\lambda^{-1}K) \\ &\quad +\frac{(K-\lambda/\gamma)\max(\lambda,K-\lambda)}{(\gamma-\lambda)(1-\lambda)}+\frac{\lambda(K-1)}{(\gamma-\lambda)(1-\lambda)}. \end{aligned} \tag{23}$$

For the *nonatomic case* let $\tilde{R}=\arg\max_{1<R<R_0}R_1(\beta,R,L(R))$. Then we have $\rho=1/R_1(\beta,\tilde{R},L(\tilde{R}))$ and for $\rho<\gamma<1$,

$$\begin{aligned} M &= \frac{\gamma^{-\alpha_2-1}(K\gamma-\lambda)}{(\gamma-\lambda)[1-(1-\tilde{\beta})\gamma^{-\alpha_1}]^2}\times\left(\frac{\tilde{\beta}\max(\lambda,K-\lambda)}{1-\lambda}+\frac{(1-\tilde{\beta})(\gamma^{-\alpha_1}-1)}{\gamma^{-1}-1}\right) \\ &\quad +\frac{\max(\lambda,K-\lambda/\gamma)}{\gamma-\lambda}+\frac{\tilde{\beta}\gamma^{-\alpha_2-2}K(K\gamma-\lambda)}{(\gamma-\lambda)[1-(1-\tilde{\beta})\gamma^{-\alpha_1}]^2}K_1(\gamma^{-1},\beta,\tilde{R},L(\tilde{R})) \\ &\quad +\frac{\gamma^{-\alpha_2}\lambda(K-1)}{(1-\lambda)(\gamma-\lambda)[1-(1-\tilde{\beta})\gamma^{-\alpha_1}]}+\frac{K[K\gamma-\lambda-\tilde{\beta}(\gamma-\lambda)]}{\gamma^2(\gamma-\lambda)[1-(1-\tilde{\beta})\gamma^{-\alpha_1}]} \\ &\quad +\frac{K-\lambda-\tilde{\beta}(1-\lambda)}{(1-\lambda)(1-\gamma)}\left((\gamma^{-\alpha_2}-1)+(1-\tilde{\beta})(\gamma^{-\alpha_1}-1)/\tilde{\beta}\right). \end{aligned} \tag{24}$$

A.2. Formulas for self-adjoint operators

A Markov chain is said to be reversible with respect to π if $\int_{\mathcal{X}} Pf(x)g(x)\pi(dx) = \int_{\mathcal{X}} f(x)Pg(x)\pi(dx)$ for all $f, g \in L^2(\pi)$. For reversible Markov chains the following tighter bounds are available.

For the *atomic case* define

$$R_2 = \begin{cases} \min\left\{\lambda^{-1}, r_s\right\}, & \text{if} \quad K > \lambda + 2\beta, \\ \lambda^{-1}, & \text{if} \quad K \leq \lambda + 2\beta, \end{cases}$$

where r_s is the unique solution of $1 + 2\beta r = r^{1+(\log K)(\log \lambda^{-1})}$. Then $\rho = R_2^{-1}$ and for $\rho < \gamma < 1$ take M as in (23) with $K_1(\gamma^{-1}, \beta, \lambda^{-1}, \lambda^{-1}K)$ replaced by $K_2 = 1 + 1/(\gamma - \rho)$.

For the *nonatomic case* let

$$R_2 = \begin{cases} r_s, & \text{if} \quad L(R_0) > 1 + 2\beta R_0, \\ R_0, & \text{if} \quad L(R_0) \leq 1 + 2\beta R_0, \end{cases}$$

where r_s is the unique solution of $1+2\beta r = L(r)$. Then $\rho = R_2^{-1}$ and for $\rho < \gamma < 1$ take M as in (24) with $K_1(\gamma^{-1}, \beta, \tilde{R}, L(\tilde{R}))$ replaced by $K_2 = 1 + \sqrt{\tilde{\beta}}/(\gamma - \rho)$.

A.3. Formulas for self-adjoint positive operators

A Markov chain is said to be positive if $\int_{\mathcal{X}} Pf(x)f(x)\pi(dx) \geq 0$ for every $f \in L^2(\pi)$. For reversible and positive markov chains take M's as in Section A.2 with $\rho = \lambda$ in the *atomic case* and $\rho = R_0^{-1}$ in the *nonatomic case*.

References

[1] R. Adamczak (2008): A tail inequality for suprema of unbounded empirical processes with applications to Markov chains. *Electronic Journal of Probability.* 34, 1000–1034.

[2] Aldous D., 1987, *On the Markov Chain Simulation Method for Uniform Combinatorial Distributions and Simulated Annealing.* Probability in the Engineering and Informational Sciences 1, 33-46.

[3] Baxendale P. H., 2005. *Renewal Theory and Computable Convergence Rates for Geometrically Ergodic Markov Chains.* Ann. Appl. Prob. 15, 700-738.

[4] Bednorz, W., (2009) *On the Kendall Theorem and its Applicatios to the Geometrical Ergodicity of Markov Chains.* Preprint.

[5] W. Bednorz, R. Latała and K. Łatuszyński (2008): A Regeneration Proof of the Central Limit Theorem for Uniformly Ergodic Markov Chains. *Elect. Comm. in Probab.* 13, 85–98.

[6] P. Bertail, S. Clémençon (2006): Regeneration-based statistics for Harris recurrent Markov chains, pages 1–54. Number 187 in *Lecture notes in Statistics.* Springer.

[7] P. Bertail, S. Clémençon (2009): Sharp bounds for the tail of functionals of Markov chains, to appear *Probability Theory and its applications.*

[8] Casella G., Robert C. P., 1999. *Monte Carlo Statistical Methods.* Springer-Verlag, New York.

[9] Chan K. S., Yue H., 1996, "Asymptotic Efficiency of the Sample Mean in Markov Chain Monte Carlo Schemes," *Journal of the Royal Statistical Society, Series B.* 58 (3), 525-539.

[10] S.J.M. Clémençon (2001): Moment and probability inequalities for sums of bounded functionals of regular Markov chains via the Nummelin splitting technique. *Statist. Probab. Lett.* 55, 227–238.

[11] R. Douc, A. Guillin and E. Moulines (2008): Bounds on regeneration times and limit theorems for subgeometric Markov chains, *Ann. Inst. H. Poincar Probab. Statist.* 44, 239–257.

[12] Douc R., Moulines E., Rosenthal J.S. (2003): Quantitative bounds on convergence of time-inhomogeneous Markov Chains. *Ann. Appl. Prob.* 14, 1643-1665.

[13] G. Fort (2002): Computable bounds for V-geometric ergodicity of Markov transition kernels. Preprint.

[14] Geyer C. J., 1992, *Practical Markov Chain Monte Carlo.* Stat. Sci. 7 (4), 473-511.

[15] Gilks W.R., Roberts G.O., Sahu S.K. (1998): Adaptive Markov chain Monte Carlo through regeneration. *J. Amer. Statist. Assoc.* 93(443), 1045–1054.

[16] Gillman D., 1998, *A Chernoff Bound for Random Walks on Expander Graphs.* SIAM J. Comput. 27 (4), 1203-1220.

[17] Glynn P. W., Ormoneit D. 2002 *Hoeffding's Inequality for Uniformly Ergodic Markov Chains.* Statist. and Probab. Lett. 56, 143-146.

[18] J.P. Hobert, G.L. Jones: Honest Exploration of Intractable Probability Distributions via Markov Chain Monte Carlo. *Statistical Science* 16(4), pp. 312–334, 2001.

[19] M.R. Jerrum, L.G. Valiant, V.V. Vizirani: Random generation of combinatorial structures fro, a uniform distribution. *Theoretical Computer Science* 43, 169–188, 1986.

[20] Johnson A. A., Jones G. L., (2007) *Gibbs Sampling for a Bayesian Hierarchical Version of the General Linear Mixed Model.* Preprint. arXiv:0712.3056v3

[21] Jones, G. L. (2005). On the Markov chain central limit theorem. *Probability Surveys* **1** 299–320.

[22] Jones, G. L., Haran, M., Caffo, B. S., Neath, R. (2006), "Fixed-Width Output Analysis for Markov Chain Monte Carlo," *Journal of the American Statatistical Association,* 101, 1537-1547.

[23] Jones G.L., Hobert J. P., 2004, *Sufficient Burn-in for Gibbs Samplers for a Hierarchical Random Effects Model.* The Annals of Statistics 32 (2), 784-817.

[24] Łatuszyński K., Miasojedow B., Niemiro W. (2009): Nonasymptotic bounds on the estimation error for regenerative MCMC algorithms. *Submitted*, arXiv:0907.4915v1

[25] Kontoyiannis I., Lastras-Montano L., Meyn S. P. 2005 *Relative Entropy and Exponential Deviation Bounds for General Markov Chains.* 2005 IEEE International Symposium on Information Theory.
[26] León C. A., Perron F., 2004. *Optimal Chernoff Bounds for Finite Reversible Markov Chains.* Ann. Appl. Prob. 14, 958-970.
[27] Liu J. S., 2001. *Monte Carlo Strategies in Scientific Computing.* Springer.
[28] Mathé P. (2004): Numerical integration using $V-$uniformly ergodic Markov chains. *J. Appl. Prob.* 41, 1104–1112.
[29] Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953. *Equations of state calculations by fast computing machines.* J. Chem. Phys. 21, 1087-1091.
[30] Meyn S. P., Tweedie R. L., 1993. *Markov Chains and Stochastic Stability.* Springer-Verlag.
[31] W. Niemiro, P. Pokarowski (2009): Fixed precision MCMC Estimation by Median of Products of Averages. *J. Appl. Probab.* 46 (2), 309–329.
[32] Roberts G. O., Rosenthal J. S., 2005. *General state space Markov chains and MCMC algorithms.* Probability Surveys 1:20-71.
[33] Roberts G. O., Rosenthal J. S., 1997. *Shift-coupling and convergence rates of ergodic averages.* Comm. in Stat. - Stoch. Models 13, 147-165.
[34] Roberts., G. O., Tweedie, R. L., 1999, *Bounds on Regeneration Times and Convergence Rates for Markov Chains.* Stochastic Process. Appl. 91, 337-338.
[35] Rosenthal, J. S., 1995, *Minorization Conditions and Convergence Rates for Markov Chain Monte Carlo.* Journal of the American Statistical Association, 90, 558-566.
[36] Rosenthal, J. S., 1995, *Rates of Convergence for Gibbs Sampling for Variance Component Models.* The Annals of Statistics, 23, 740-761.
[37] D. Rudolf (2008): Explicit error bounds for lazy reversible Markov chain Monte Carlo. *J. of Complexity.* 25, 11–24.
[38] Sahu S.K., Zhigljavsky A.A., (2003): Self-regenerative Markov Chain Monte Carlo with adaptation. *Bernoulli* 9, 395-422.