Bernoulli **22**(3), 2016, 1808–1838
DOI: 10.3150/15-BEJ712

Markov Chain Monte Carlo confidence intervals

YVES F. ATCHADÉ

University of Michigan, 1085 South University, Ann Arbor, 48109, MI, United States.
E-mail: yvesa@umich.edu

For a reversible and ergodic Markov chain $\{X_n, n \geq 0\}$ with invariant distribution π, we show that a valid confidence interval for $\pi(h)$ can be constructed whenever the asymptotic variance $\sigma_P^2(h)$ is finite and positive. We do not impose any additional condition on the convergence rate of the Markov chain. The confidence interval is derived using the so-called fixed-b lag-window estimator of $\sigma_P^2(h)$. We also derive a result that suggests that the proposed confidence interval procedure converges faster than classical confidence interval procedures based on the Gaussian distribution and standard central limit theorems for Markov chains.

Keywords: Berry–Esseen bounds; confidence interval; lag-window estimators; martingale approximation; MCMC; reversible Markov chains

1. Introduction

Confidence intervals play an important role in Monte Carlo simulation (Robert and Casella [26], Asmussen and Glynn [1]). In Markov Chain Monte Carlo (MCMC), the existing literature requires the Markov chain to be geometrically ergodic for the validity of confidence interval procedures (Jones *et al.* [15], Flegal and Jones [8], Atchadé [3]). The main objective of this work is to simplify some of these assumptions. We show that for a reversible ergodic Markov chain, a valid confidence interval can be constructed whenever the asymptotic variance itself is finite. No additional convergence rate assumption on the Markov chain is required.

Let $\{X_n, n \geq 0\}$ be a reversible stationary Markov chain with invariant distribution π. For $h \in L^2(\pi)$, the asymptotic variance of h is denoted $\sigma_P^2(h)$ (see (2) below for the definition). A remarkable result by C. Kipnis and S. R. Varadhan (Kipnis and Varadhan [19]) says that if $0 < \sigma_P^2(h) < \infty$, then $\frac{1}{\sigma_P(h)\sqrt{n}} \sum_{i=1}^n (h(X_i) - \pi(h))$ converges weakly to $\mathbf{N}(0,1)$ where $\pi(h) \stackrel{\text{def}}{=} \int h(z)\pi(\mathrm{d}z)$. In order to turn this result into a confidence interval for $\pi(h)$, an estimator σ_n of $\sigma_P(h)$ is needed. A common practice consists in choosing σ_n as a consistent estimator of $\sigma_P(h)$. However, consistent estimation of $\sigma_P(h)$ typically requires further assumptions on the convergence rate of the Markov chain (typically geometric ergodicity), and on the function h. Instead of insisting on consistency, we consider the so-called fixed-b approach developed by Kiefer, Vogelsang and Bunzel [18], Kiefer and Vogelsang [17], where the proposed estimator σ_n is known to be inconsistent. Using this inconsistent estimator we show in Theorem 2.2 that a Studentized analog of the Kipnis–Varadhan's theorem holds: if $0 < \sigma_P^2(h) < \infty$, then $\mathbf{T}_n \stackrel{\text{def}}{=} \frac{1}{\sigma_n\sqrt{n}} \sum_{i=1}^n (h(X_i) - \pi(h))$ converges weakly to a (non-Gaussian) distribution. The theorem extends to nonstationary Markov chains that satisfy a very mild ergodicity assumption. To a certain extent, the result is a generalization of Atchadé and Cattaneo [4] which establishes the same limit theorem for geometrically ergodic (but not necessarily reversible) Markov chains. The result is particularly relevant for Markov chains with sub-geometric convergence rates. For such Markov chains, the author is not aware of any result that guarantees the asymptotic validity of confidence intervals. However, it is important to point out that the finiteness of $\sigma_P^2(h)$ carries some implications in terms of convergence rate of P, and is not always easy to check. But the main point of this work is that the finiteness of $\sigma_P^2(h)$ is all that is needed for consistent confidence interval.

As we shall see, Theorem 2.2 comes from the fact that there exists a pair of random variables (N, D), say, such that the joint process $(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(h(X_i) - \pi(h)), \sigma_n^2)$ converges weakly to $(\sigma_P(h)N, \sigma_P^2(h)D)$. As a result, $\sigma_P(h)$ cancels out in the limiting distribution of $\mathbf{T}_n$. This approach to confidence intervals is closely related to the standardized time series method of Schruben [29] (see also Glynn and Iglehart [9]), well known in operations research. Indeed in its simplest form, the standardized time series method is the analog of the fixed-b procedure using the batch-mean estimator with a fixed number of batches. Despite this close connection, this paper focuses only on the fixed-b confidence interval.

We also compare the fixed-b lag-window estimators with the more commonly used lag-window estimators. We limit this comparison to the case of geometrically ergodic Markov chains. We prove in Theorem 2.6 that the convergence rate of the fixed-b lag-window estimator is of order $\log(n)/\sqrt{n}$, better than the fastest rate achievable by the more commonly used lag-window estimator. Similar comparisons based on the convergence of $\mathbf{T}_n$ has been reported elsewhere in the literature. Jansson [13] studied stationary Gaussian moving average models and established that the rate of convergence of $\mathbf{T}_n$ is $n^{-1}\log(n)$. Sun, Phillips and Jin [30] obtained the rate n^{-1}, under the main assumption that the underlying process is Gaussian and stationary. It seems unlikely that the convergence rate n^{-1} will hold without the Gaussian assumption. However, it is unclear whether the convergence rate $\log(n)/\sqrt{n}$ obtained in Theorem 2.6 is tight.

We organize the paper as follows. Section 2 contains the main results, including the rate of convergence of the fixed-b lag-window estimator in Section 2.4. We present a simulation example to illustrate the finite sample properties of the confidence intervals in Section 2.5. All the main proofs are postponed to Section 3 and the Appendix.

1.1. Notation

Throughout the paper $(\mathsf{X}, \mathcal{B})$ denotes a measure space with a countably generated sigma-algebra $\mathcal{B}$ with a probability measure of interest π. We denote $L^2(\pi)$ the usual space of L^2-integrable functions with respect to π, with norm $\|\cdot\|$ and associated inner product $\langle\cdot\rangle$, and we denote $L_0^2(\pi)$ the subspace of $L^2(\pi)$ of functions orthogonal to the constants: $L_0^2(\pi) \stackrel{\text{def}}{=} \{f \in L^2(\pi)\colon \int f(x)\pi(\mathrm{d}x) = 0\}$.

For a measurable function $f:\mathsf{X}\to\mathbb{R}$, a probability measure ν on $(\mathsf{X}, \mathcal{B})$ and a Markov kernel Q on X, we use the notation: $\nu(f) \stackrel{\text{def}}{=} \int f(x)\nu(\mathrm{d}x)$, $\bar{f} \stackrel{\text{def}}{=} f - \pi(f)$, $Qf(x) \stackrel{\text{def}}{=} \int f(y)Q(x,\mathrm{d}y)$, and $Q^j f(x) \stackrel{\text{def}}{=} Q\{Q^{j-1}f\}(x)$, with $Q^0 f(x) = f(x)$. For $V:\mathsf{X}\to[0,\infty)$, we define $\mathcal{L}_V$ as the space of all measurable real-valued functions $f:\mathsf{X}\to\mathbb{R}$ s.t. $|f|_V \stackrel{\text{def}}{=} \sup_{x\in\mathsf{X}}|f(x)|/V(x) < \infty$.

For two probability measures ν_1, ν_2, we denotes $\|\nu_1 - \nu_2\|_{\text{tv}} \stackrel{\text{def}}{=} \sup_{|f|\le 1} |\nu_1(f) - \nu_2(f)|$, the total variation distance between ν_1 and ν_2, and $\|\nu_1 - \nu_2\|_V \stackrel{\text{def}}{=} \sup_{\{f, |f|_V \le 1\}} |\nu_1(f) - \nu_2(f)|$, its V-norm generalization.

For sequences $\{a_n, b_n\}$ of real nonnegative numbers, the notation $a_n \lesssim b_n$ means that $a_n \le c b_n$ for all n, and for some constant c that does not depend on n. For a random sequence $\{X_n\}$, we write $X_n = \mathrm{O}_p(a_n)$ if the sequence $|X_n|/a_n$ is bounded in probability. We say that $X_n = \mathrm{o}_p(a_n)$ if X_n/a_n converges in probability to zero as $n \to \infty$.

2. Monte Carlo confidence intervals for reversible Markov chains

Throughout the paper, P denotes a Markov kernel on $(\mathsf{X}, \mathcal{B})$ that is reversible with respect to π. This means that for any pair $f, g \in L^2(\pi)$, $\langle f, Pg\rangle = \langle g, Pf\rangle$. We assume that P satisfies the following.

A1 For π-almost all $x \in \mathsf{X}$,

$$\lim_{n\to\infty} \left\| P^n(x, \cdot) - \pi \right\|_{\text{tv}} = 0. \tag{1}$$

Remark 1. Assumption A1 is very basic. For instance, if P is ϕ-irreducible, and aperiodic (in addition to being reversible with respect to π), then A1 holds. If in addition P is Harris recurrent, then (1) holds *for all* $x \in \mathsf{X}$. If P is a Metropolis–Hastings kernel, Harris recurrence typically follows from π-irreducibility. All these statements can be found, for instance, in Tierney [31].

Throughout the section, unless stated otherwise, $\{X_n, n \ge 0\}$ is a (nonstationary) Markov chain on $(\mathsf{X}, \mathcal{B})$ with transition kernel P and started at some arbitrary (but fixed) point $x \in \mathsf{X}$ for which (1) holds. The Markov kernel P induces in the usual way a self-adjoint operator (also denoted P) on the Hilbert space $L^2_0(\pi)$ that maps $h \mapsto Ph$. This operator P admits a spectral measure $\mathcal{E}$ on $[-1, 1]$, and for $h \in L^2_0(\pi)$ we will write $\mu_h(\cdot) \stackrel{\text{def}}{=} \langle h, \mathcal{E}(\cdot)h\rangle$ for the associated nonnegative Borel measure on $[-1, 1]$. Assumption A1 implies that μ_h does not charge 1 or -1, that is $\mu_h(\{-1, 1\}) = 0$. This is Lemma 5 of Tierney [5].

2.1. Confidence interval for $\pi(h)$

Let $h \in L^2_0(\pi)$. We define

$$\sigma_P^2(h) \stackrel{\text{def}}{=} \int_{-1}^{1} \frac{1+\lambda}{1-\lambda} \mu_h(\mathrm{d}\lambda), \tag{2}$$

that we call the asymptotic variance of h. The terminology comes from the fact that if the Markov chain is assumed stationary, a calculation (see, e.g., Häggström and Rosenthal [11], Theorem 4) using the properties of the spectral measure μ_h gives

$$\lim_{n\to\infty} n\mathbb{E}\left[\left(n^{-1}\sum_{k=1}^{n} h(X_k)\right)^2\right] = \sigma_P^2(h). \tag{3}$$

For nonstationary Markov chains, such as the one considered in this paper, it is unclear whether (3) continues to hold in complete generality. The estimation of $\sigma_P^2(h)$ is often of interest because when (3) holds, $\sigma_P^2(h)/n$ approximates the mean squared error of the Monte Carlo estimate $n^{-1}\sum_{k=1}^{n} h(X_k)$. An estimate of $\sigma_P^2(h)$ is often also sought in order to exploit the Kipnis–Varadhan theorem for confidence interval purposes. It is known (Häggström and Rosenthal [11], Theorem 4) that $\sigma_P^2(h)$ can also be written as

$$\sigma_P^2(h) = \sum_{\ell=-\infty}^{+\infty} \gamma_{|\ell|}(h), \tag{4}$$

where for $\ell \ge 0$, $\gamma_\ell(h) \stackrel{\text{def}}{=} \langle h, P^\ell h\rangle$. This suggests the so-called lag-window estimator of $\sigma_P^2(h)$

$$\begin{aligned} \sigma_{b_n}^2 &\stackrel{\text{def}}{=} \sum_{\ell=-n+1}^{n-1} w\left(\frac{\ell}{b_n}\right)\gamma_{n,|\ell|}, \\ \text{where } \gamma_{n,\ell} &\stackrel{\text{def}}{=} n^{-1}\sum_{j=1}^{n-\ell}\big(h(X_j)-\hat{\pi}_n(h)\big)\big(h(X_{j+\ell})-\hat{\pi}_n(h)\big). \end{aligned} \tag{5}$$

In the above display, $\hat{\pi}_n(h) = n^{-1}\sum_{k=1}^{n} h(X_k)$, $1 \le b_n \le n$ is an integer such that $b_n \to \infty$, as $n \to \infty$, and $w:\mathbb{R}\to\mathbb{R}$ is an even function ($w(-x) = w(x)$) with support $[-1, 1]$, that is, $w(x) \neq 0$ on $(-1, 1)$ and $w(x) = 0$ for $|x| \ge 1$. Since w has support $[-1, 1]$, the actual range for ℓ in the summation defining $\sigma_{b_n}^2$ is $-b_n + 1 \le \ell \le b_n - 1$.

The lag-window estimator $\sigma_{b_n}^2$ can be applied more broadly in time series and the method has a long history. Some of the earlier work go back to the 1950s (Grenander and Rosenblatt [10], Parzen [24]). Convergence results specific to nonstationary Markov chains have been established recently (see, e.g., Damerdji [6], Flegal and Jones [8], Atchadé [3] and the references therein); however, under assumptions that are much stronger than A1. It remains an open problem whether $\sigma_{b_n}^2$ can be shown to converge to $\sigma_P^2(h)$ assuming only A1. In particular, the author is not aware of any result that establishes the consistency of $\sigma_{b_n}^2$ without assuming that P is geometrically ergodic.

However, if the goal is to construct a confidence interval for $\pi(h)$, we will now see that it is enough to assume A1 and $\sigma_P^2(h) < \infty$. Consider the lag-window estimator obtained by setting $b_n = n$. This writes

$$\sigma_n^2 \stackrel{\text{def}}{=} \sum_{\ell=-n+1}^{n-1} w\left(\frac{\ell}{n}\right)\gamma_{n,|\ell|}. \tag{6}$$

This estimator is well known to be inconsistent for estimating $\sigma_P^2(h)$, but has recently attracted a lot of interest in the Econometrics literature under the name of fixed-b asymptotics (Kiefer, Vogelsang and Bunzel [18], Kiefer and Vogelsang [17], Sun, Phillips and Jin [30], see also Neave [22] for some pioneer work). This paper takes inspiration from this literature. However, unlike these works, we exploit the Markov structure and we do not impose any stationary assumption. We introduce the function $v(t) \stackrel{\text{def}}{=} \int_0^1 w(t-u)\,\mathrm{d}u$, $t \in [0,1]$, and the kernel $\phi : [0,1] \times [0,1] \to \mathbb{R}$, where

$$\phi(s,t) = w(s-t) - v(s) - v(t) + \int_0^1 v(t)\,\mathrm{d}t, \qquad s,t \in [0,1]. \tag{7}$$

We say that a kernel $k : [0,1] \times [0,1] \to \mathbb{R}$ is positive definite if for all $n \geq 1$, all $a_1, \ldots, a_n \in \mathbb{R}$, and $t_1, \ldots, t_n \in [0,1]$, $\sum_{i=1}^n \sum_{j=1}^n a_i a_j k(t_i, t_j) \geq 0$. We will assume that the weight function w in (6) is such that the following holds.

A2 The function $w : \mathbb{R} \to \mathbb{R}$ is an even function, with support $[-1,1]$, and of class $\mathcal{C}^2$ on $(-1,1)$. Furthermore, the kernel ϕ defined in (7) is positive definite, and not identically zero.

Example 1. Assumption A2 holds for the function w given by $w(u) = (1-u^2)\mathbf{1}_{(-1,1)}(u)$. Indeed in this case, a simple calculation gives that $\phi(s,t) = 2(s-0.5)(t-0.5)$, which (by its multiplicative form) is clearly positive definite. In this particular case, solving $\int_0^1 \phi(s,t)u(t)\,\mathrm{d}t = \alpha u(s)$ yields the unique eigenvalue $\alpha = 2\int_0^1 (t-0.5)^2\,\mathrm{d}t = 1/6$.

A general approach to guarantee that ϕ as in (7) is positive definite is to start with a positive definite function w, as the next lemma shows.

Lemma 2.1. *Suppose that the kernel $[0,1] \times [0,1] \to \mathbb{R}$ defined by $(s,t) \mapsto w(s-t)$ is continuous and positive definite. Then ϕ as in* (7) *is also positive definite.*

Proof. By Mercer's theorem (see Theorem A.1), there exist nonnegative numbers $\{\lambda_j, j \geq 0\}$, orthonormal functions $\xi_j : [0,1] \to \mathbb{R}$ such that $\int_0^1 w(t-s)\xi_j(s)\,\mathrm{d}s = \lambda_j \xi_j(t)$, and

$$w(t-s) = \sum_{j\geq 0} \lambda_j \xi_j(t)\xi_j(s),$$

and the series converges uniformly and absolutely. It is easy to show that one can interchange integral and sum and write $v(t) = \int_0^1 w(t-s)\,\mathrm{d}s = \sum_{j\geq 0} \lambda_j \xi_j(t) \int_0^1 \xi_j(s)\,\mathrm{d}s$, $\int_0^1 v(t)\,\mathrm{d}t = \int_0^1 \int_0^1 w(t-s)\,\mathrm{d}s\,\mathrm{d}t = \sum_{j\geq 0} \lambda_j (\int_0^1 \xi_j(t)\,\mathrm{d}t)^2$, and then we get

$$\phi(s,t) = \sum_{j\geq 0} \lambda_j \left(\xi_j(t) - \int_0^1 \xi_j(t)\,\mathrm{d}t\right)\left(\xi_j(s) - \int_0^1 \xi_j(s)\,\mathrm{d}s\right).$$

This expression of ϕ easily shows that it is positive definite. □

The usual approach for showing that the kernel $(s,t) \mapsto w(s-t)$ is positive definite is by showing that the weight function $t \mapsto w(t)$ is a characteristic function (or more generally the Fourier transform of a positive measure) and applying Bochner's theorem. This approach shows that A2 holds for the Bartlett function $w(x) = (1-|x|)\mathbf{1}_{(-1,1)}(x)$, the Parzen function

$$w(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & \text{if } |x| \le \frac{1}{2}, \\ 2(1-|x|)^3, & \text{if } \frac{1}{2} \le |x| \le 1, \\ 0, & \text{if } |x| > 1, \end{cases}$$

and for a number of others weight functions (see, e.g., Hannan [12], pages 278–279 for details). In the case of the Bartlett function, the kernel ϕ is given by

$$\phi(s,t) = \tfrac{2}{3} - s(1-s) - t(1-t) - |s-t|.$$

For the Parzen function, we have

$$v(s) = \frac{3}{8} + s \wedge (1-s) - 2\bigl(s \wedge (1-s)\bigr)^3 + \bigl(s \wedge (1-s)\bigr)^4 \quad \text{and} \quad \int_0^1 v(t)\,\mathrm{d}t = \frac{23}{40},$$

where $a \wedge b \stackrel{\text{def}}{=} \min(a,b)$.

Assumption A2 implies that ϕ, considered as a linear operator on $L^2[0,1]$ ($\phi f(s) = \int_0^1 \phi(s,t) f(t)\,\mathrm{d}t$) is self-adjoint, compact and positive. Therefore, it has only nonnegative eigenvalues, and a countable number of positive eigenvalues. We denote $\{\alpha_j, j \in \mathsf{I}\}$ the set of positive eigenvalues of ϕ (each repeated according to its multiplicity). The index set $\mathsf{I} \subseteq \{1,2,\ldots\}$ is either finite or $\mathsf{I} = \{1,2,\ldots\}$. We introduce the random variable $\mathbf{T}_w$ defined as

$$\mathbf{T}_w \stackrel{\text{def}}{=} \frac{Z_0}{\sqrt{\sum_{i\in\mathsf{I}} \alpha_i Z_i^2}} \qquad \text{where } \{Z_0, Z_i, i \in \mathsf{I}\} \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0,1).$$

Here is the main result.

Theorem 2.2. *Assume* A1–A2, *and* $h \in L^2(\pi)$. *If* $0 < \sigma_P^2(h) < \infty$, *then as* $n \to \infty$,

$$\sigma_n^2 \stackrel{\mathsf{w}}{\to} \sigma_P^2(h) \sum_{i\in\mathsf{I}} \alpha_i Z_i^2 \quad \textit{and} \quad \mathbf{T}_n \stackrel{\text{def}}{=} \frac{1}{\sigma_n\sqrt{n}} \sum_{k=1}^n \bigl(h(X_k) - \pi(h)\bigr) \stackrel{\mathsf{w}}{\to} \mathbf{T}_w,$$

where $\{Z_i, i \in \mathsf{I}\} \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0,1)$.

Proof. See Section 3.1. □

The theorem implies that the confidence interval

$$\hat{\pi}_n(h) \pm t_{1-\alpha/2}\sqrt{\frac{\sigma_n^2}{n}}, \tag{8}$$

is an asymptotically valid Monte Carlo confidence interval for $\pi(h)$, where $t_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the distribution of $\mathbf{T}_w$. These quantiles are intractable in general but can be easily approximated by Monte Carlo simulation (see Section 2.3).

The assumption that $\sigma_P^2(h)$ is finite can be difficult to check. When P is known to satisfy a drift condition, one can find whole class of functions for which the asymptotic variance is finite, as the following proposition shows. The proposition uses Markov chain concepts that have not been defined above, and we refer the reader to Meyn and Tweedie [21] for details.

Proposition 2.3. *Suppose that P is ϕ-irreducible and aperiodic, with invariant distribution π. Suppose also that there exist measurable functions $V, f : \mathsf{X} \to [1, \infty)$, constant $b < \infty$, and some petite set $C \in \mathcal{B}$ such that*

$$PV(x) \le V(x) - f(x) + b\mathbf{1}_C(x), \qquad x \in \mathsf{X}. \tag{9}$$

If $\pi(fV) < \infty$, then for all $h \in \mathcal{L}_f$, $\sigma_P^2(h) < \infty$.

Proof. This is a well-known result. We give the proof only for completeness. Without any loss of generality, suppose that $\pi(h) = 0$. We recall that $\sigma_P^2(h) = \pi(h^2) + 2\sum_{j\ge1}\langle h, P^j h\rangle$. Since $|\langle h, P^j h\rangle| \le \int |h(x)||P^j h(x)|\pi(\mathrm{d}x)$, we obtain

$$\sum_{j\ge0}\bigl|\langle h, P^j h\rangle\bigr| \le |h|_f \int \bigl|h(x)\bigr| \Biggl\{\sum_{j\ge0}\bigl\|P^j(x,\cdot) - \pi(\cdot)\bigr\|_f\Biggr\}\pi(\mathrm{d}x).$$

Since P is ϕ-irreducible and aperiodic, and under the drift condition (9), Meyn and Tweedie [21], Theorem 14.0.1 implies that there exists a finite constant B such that $\sum_{j\ge0}\|P^j(x,\cdot) - \pi(\cdot)\|_f \le BV(x)$, $x \in \mathsf{X}$. We conclude that

$$\sigma_P^2(h) \le 2B|h|_f \int \bigl|h(x)\bigr|V(x)\pi(\mathrm{d}x) \le 2B|h|_f^2 \int f(x)V(x)\pi(\mathrm{d}x) < \infty. \qquad \square$$

Remark 2. Proposition 2.3 has a number of well-known special cases. The most common case is when $f = \lambda V$ for some $\lambda \in (0, 1)$, in which case P is geometrically ergodic and $\sigma_P^2(h) < \infty$ for all $h \in \mathcal{L}_{V^{1/2}}$. Another important special case is $f = V^\alpha$, for some $\alpha \in [0, 1)$. Such drift condition implies that the Markov chain converges at a polynomial rate. If $\alpha \ge 0.5$, then Proposition 2.3 implies that $\sigma_P^2(h) < \infty$ for all $h \in \mathcal{L}_{V^{\alpha-0.5}}$. To see this, notice that (9) with $f = V^\alpha$, and Jarner and Roberts [14], Lemma 3.5 imply that $PV^{1/2} \le V^{1/2} - cV^{\alpha-1/2} + b_1\mathbf{1}_C$. Since $\pi(V^\alpha) < \infty$, the claim follows from Proposition 2.3.

2.2. Example: Metropolis Adjusted Langevin Algorithm for smooth densities

We give another example where it is possible to check that $\sigma_P^2(h) < \infty$ without geometric ergodicity. Take $\mathsf{X} = \mathbb{R}^d$ equipped with the usual Euclidean inner product $\langle\cdot,\cdot\rangle_2$, norm $|\cdot|$, and the Lebesgue measure denoted $\mathrm{d}x$. We consider a probability measure π that has a density with respect to the Lebesgue measure, and in a slight abuse of notation we use the same symbol to represent π and its density: $\pi(x) = \mathrm{e}^{-u(x)}/Z$, for some function $u : \mathsf{X} \to \mathbb{R}$ that we assume is differentiable, with gradient ∇u.

Let $q_\sigma(x, \cdot)$ denotes the density of the Gaussian distribution $\mathbf{N}(x - \frac{\sigma^2}{2}\rho(x)\nabla u(x), \sigma^2 I_d)$, where the term $\rho(x) \geq 0$ is used to modulate the drift $-\frac{\sigma^2}{2}\nabla u(x)$, and $\sigma > 0$ is a scaling constant. We consider the Metropolis–Hastings algorithm that generates a Markov chain $\{X_n, n \geq 0\}$ with invariant distribution π as follows. Given $X_n = x$, we propose $Y \sim q_\sigma(x, \cdot)$. We either "accept" Y and set $X_{n+1} = Y$ with probability $\alpha(x, Y)$, or we "reject" Y and set $X_{n+1} = x$, where

$$\alpha(x, y) \stackrel{\text{def}}{=} \min\left(1, \frac{\pi(y)}{\pi(x)} \frac{q_\sigma(y, x)}{q_\sigma(x, y)}\right).$$

When $\rho(x) = 0$, we get the Random Walk Metropolis (RWM), and when $\rho(x) = 1$, we get the Metropolis Adjusted Langevin Algorithm (MaLa). However, we are mainly interested in the case where

$$\rho(x) \stackrel{\text{def}}{=} \frac{\tau}{\max(\tau, |\nabla u(x)|)}, \qquad x \in \mathsf{X} \tag{10}$$

for some given constant $\tau > 0$, which corresponds to the truncated MaLa proposed by Roberts and Tweedie [28]. The truncated MaLa combines the stability of the RWM and the mixing of the MaLa. It is known to be geometrically ergodic whenever RWM is geometrically ergodic (Atchadé [2]). However, checking in practice that the truncated MaLa is geometrically ergodic can be difficult, as this involves checking conditions on the curvature of the log-density. We show in the next result that if the gradient of the log-density u is Lipschitz and unbounded then P satisfies a drift condition of the type (9), and $\sigma_P^2(h)$ is guaranteed to be finite for certain functions.

B1 Suppose that u is bounded from below, continuously differentiable, and ∇u is Lipschitz, and

$$\limsup_{|x| \to \infty} \left|\nabla u(x)\right| = +\infty.$$

Theorem 2.4. *Assume* B1 *and* (10). *Set* $V(x) \stackrel{\text{def}}{=} a + u(x)$, *where* $a \in \mathbb{R}$ *is chosen such that* $V \geq 1$. *Then there exist* $b, r \in (0, \infty)$ *such that*

$$PV(x) \leq V(x) - \frac{\sigma^2}{4}\rho(x)\left|\nabla u(x)\right|^2 + b\mathbf{1}_{\{|x| \leq r\}}(x), \qquad x \in \mathsf{X}. \tag{11}$$

In particular, if $\int u(x)|\nabla u(x)|\mathrm{e}^{-u(x)}\,\mathrm{d}x < \infty$, *then* $\sigma_P^2(h) < \infty$ *for all* $h \in \mathcal{L}_f$, *where* $f(x) = \rho(x)|\nabla u(x)|^2$.

Proof. See Section 3.2. □

Remark 3. This result can be useful in contexts where the log-density u is known to have a Lipschitz gradient, but is too complicated to allow an easy verification of the geometric ergodicity conditions.

2.3. On the distribution of the random variable $\mathbf{T}_w$

It is clear that the limiting distribution $\mathbf{T}_w$ used for constructing the confidence interval (8) depends on the choice of w. More research is needed to explain how to best choose w in this regard. But from the limited simulations done in this paper, we found that weight functions w with large characteristic exponents lead to heavy-tailed limiting distributions $\mathbf{T}_w$, and wider confidence intervals. The characteristic exponent of a weight function w is the largest number $r > 0$ such that $\lim_{u\to 0} |u|^{-r}(1 - w(u)) \in (0, \infty)$. Overall, we recommend the use of the Bartlett weight function $w(u) = (1 - |u|)\mathbf{1}_{(-1,1)}(u)$, which has characteristic exponent 1, and has behaved very well in the simulations conducted.

Another issue is how to compute the quantiles of $\mathbf{T}_w$. As defined, the distribution of $\mathbf{T}_w$ is intractable in general, as it requires knowing the eigenvalues of ϕ. But the next result gives a straightforward method for approximate simulation from $\mathbf{T}_w$.

Proposition 2.5. *Let $\{Z_j, 1 \le j \le N\}$ be i.i.d. standard normal random variables. Then*

$$\mathbf{T}_w^{(N)} \stackrel{\text{def}}{=} \frac{\sum_{j=1}^N Z_j}{\sqrt{\sum_{i=1}^N \sum_{j=1}^N \phi(\frac{i-1}{N}, \frac{j-1}{N}) Z_i Z_j}} \stackrel{\text{w}}{\to} \mathbf{T}_w \qquad \text{as } N \to \infty.$$

Remark 4. As pointed out by a referee, one can also approximately sample from $\mathbf{T}_w$ by generating $X_{1:N} \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0, 1)$, and compute T_N, with $h(x) = x$. The approach in Proposition 2.5 is similar, but replaces σ_N^2 by $\check{\sigma}_N^2$ as defined in (17). By Lemma 3.4, the two approaches are essentially equivalent.

Proof of Proposition 2.5. Let $\{0, \alpha_j, j \in \mathsf{I}\}$ be the eigenvalues of ϕ, with associated eigenfunctions $\{\Psi_0, \Psi_j, j \in \mathsf{I}\}$ ($\Psi_0 \equiv 1$). By Mercer's theorem (see Theorem 14 in the Appendix),

$$\sum_{i=1}^N \sum_{k=1}^N \phi\left(\frac{i-1}{N}, \frac{k-1}{N}\right) Z_i Z_k = N \sum_{j\in\mathsf{I}} \alpha_j \left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \Psi_j\left(\frac{i-1}{N}\right) Z_i\right)^2.$$

Hence,

$$\mathbf{T}_w^{(N)} = \frac{1/\sqrt{N} \sum_{i=1}^N \Psi_0((i-1)/N) Z_i}{\sqrt{\sum_{j\in\mathsf{I}} \alpha_j (1/\sqrt{N} \sum_{i=1}^N \Psi_j((i-1)/N) Z_i)^2}}.$$

It is an application of Lemma 3.3 that as $N \to \infty$, $\{\frac{1}{\sqrt{N}} \sum_{i=1}^N \Psi_0(\frac{i-1}{N}) Z_i, \frac{1}{\sqrt{N}} \sum_{i=1}^N \Psi_j(\frac{i-1}{N}) Z_i, j \in \mathsf{I}\}$ converges weakly to $\{Z_0, Z_j, j \in \mathsf{I}\}$. The result then follows from the continuous mapping theorem. □

We use Proposition 2.5 to approximately simulate $\mathbf{T}_w$ for the function $w(u) = (1 - u^2)\mathbf{1}_{(-1,1)}(u)$, and for the Bartlett and Parzen functions. Table 1 reports the 95% and 97.5%

Table 1. Approximations of t such that $\mathbb{P}(\mathbf{T}_w > t) = \alpha/2$

	$\alpha = 10\%$	$\alpha = 5\%$
$w(u) = (1-u^2)_+$	15.49 (0.06)	31.21 (0.19)
Parzen	4.11 (0.01)	5.64
Bartlett	3.77 (0.005)	4.78 (0.01)

quantiles, computed based on 10 000 independent samples of $\mathbf{T}_w^{(N)}$, with $N = 3000$. We replicate these estimates 50 times to evaluate the Monte Carlo errors reported in parenthesis.

As explained in Example 1, in the case $w(u) = (1-u^2)\mathbf{1}_{(-1,1)}(u)$, $\mathbf{T}_w = \sqrt{6}T_1$, where T_ν denotes the student's distribution with ν degree of freedom; thus, is this case we can compute accurately the quantiles. In particular, the 95% and 97.5% quantiles are 15.465 and 31.123, respectively.

2.4. Rate of convergence of σ_n^2

An interesting question is understanding how the lag-window estimators σ_n^2 and $\sigma_{b_n}^2$ compare. On one hand, the asymptotic behavior of $\sigma_{b_n}^2$ is better understood. In the stationary case, the best rate of convergence of $\sigma_{b_n}^2$ towards $\sigma_P^2(h)$ is $n^{-q/(1+2q)}$ (see, e.g., Parzen [24], Theorem 5A–B), where q is the largest number $q \in (0, r]$ such that $\sum_{j\geq 1} j^q \gamma_j(h) < \infty$, where $\gamma_j(h) = \langle h, P^j h\rangle$, and r is the characteristic exponent of w. This optimal rate is achieved by choosing $b_n \propto n^{1/(1+2q)}$. Hence, the optimal rate in the case of a geometrically ergodic Markov chain is $n^{-r/(1+2r)}$. However, it is well documented (see, e.g., Newey and West [23]) that the finite sample properties of $\sigma_{b_n}^2$ are very sensitive to the actual constant in $b_n \propto n^{1/(1+2q)}$, and some tuning is often required in practice. On the other hand, the fixed-b framework has the advantage that it requires no tuning, since $b_n = n$. Furthermore, we establish in this section that σ_n^2 has a better convergence rate. Reversibility plays no role in this discussion. We further simplify the analysis by assuming that P satisfies a geometric ergodicity assumption:

(G) There exists a measurable function $V : \mathsf{X} \to [1, \infty)$ such that $\pi(V) < \infty$, and for all $\beta \in (0, 1]$,

$$\left\| P^n(x, \cdot) - \pi(\cdot) \right\|_{V^\beta} \leq C\rho^n \, V^\beta(x), \qquad n \geq 0, x \in \mathsf{X}. \tag{12}$$

Denote $\mathsf{Lip}_1(\mathbb{R})$ the set of all bounded Lipschitz functions $f : \mathbb{R} \to \mathbb{R}$ such that

$$|f|_{\mathrm{Lip}} \stackrel{\text{def}}{=} \sup_{x \neq y} \frac{|f(x) - f(y)|}{|x-y|} \leq 1.$$

For P, Q two probability measures on $\mathbb{R}$, we define

$$\mathsf{d}_1(P, Q) \stackrel{\text{def}}{=} \sup_{f \in \mathsf{Lip}_1(\mathbb{R})} \left| \int f \, \mathrm{d}P - \int f \, \mathrm{d}Q \right|.$$

$\mathsf{d}_1(P, Q)$ is the Wasserstein metric between P, Q. An upper bound on $\mathsf{d}_1(P_n, P)$ gives a Berry–Esseen-type bound on the rate of weak convergence of P_n to P. In a slight abuse of notation, if X, Y are random variables, and $X \sim P$ and $Y \sim Q$, we shall also write $\mathsf{d}_1(X, Y)$ to mean $\mathsf{d}_1(P, Q)$.

Theorem 2.6. *Suppose that* A2 *and* (G) *hold. Suppose also that* I *is finite. For* $\delta \in [0, 1/4)$, *let* $h \in \mathcal{L}_{V^\delta}$ *be such that* $\pi(h) = 0$, *and* $\sigma_P^2(h) = 1$. *Then*

$$\mathsf{d}_1\left(\sigma_n^2, \chi^2\right) \lesssim \frac{\log(n)}{\sqrt{n}} \qquad \text{as } n \to \infty, \tag{13}$$

where $\chi^2 = \sum_{i\in\mathsf{I}} \alpha_i Z_i^2$, $\{Z_i, i \in \mathsf{I}\}$ *are i.i.d.* $\mathbf{N}(0, 1)$, *and* $\{\alpha_i, i \in \mathsf{I}\}$ *is the set of positive eigenvalues of* ϕ.

Proof. See Section 3.3. □

Remark 5. The assumption that I is finite is mostly technical and it seems plausible that this result continues to hold without that assumption. For example, I is finite for the kernel $w(u) = (1 - u^2)\mathbf{1}_{(-1,1)}(u)$.

2.5. A simulation example

This section illustrates the finite sample behavior of the fixed-b confidence interval procedure. We will compare the fixed-b procedure and the standard confidence interval procedure based on $\sigma_{b_n}^2$ (using a Gaussian limit). As example, we consider the posterior distribution of a logistic regression model, and use the Random Walk Metropolis algorithm (Robert and Casella [26]).

Let $\mathsf{X} = \Theta = \mathbb{R}^d$ equipped with its Borel sigma-algebra, and π be absolutely continuous w.r.t. the Lebesgue measure $\mathrm{d}\theta$ with density still denoted by π. We write $|\theta|$ for the Euclidean norm of θ. Let q_Σ denotes the density of the normal distribution $\mathbf{N}(0, \Sigma)$ on Θ with covariance matrix Σ. The Random Walk Metropolis algorithm (RWMA) is a popular MCMC algorithm that generates a Markov chain with invariant distribution π and transition kernel given by

$$P_\Sigma(\theta, A) = \mathbf{1}_A(\theta) + \int_{\mathsf{X}} \alpha(\theta, \theta + z)\big(\mathbf{1}_A(\theta + z) - \mathbf{1}_A(\theta)\big) q_\Sigma(z)\,\mathrm{d}z, \qquad \theta \in \Theta, A \in \mathcal{B}(\Theta),$$

where $\mathbf{1}_A$ denotes the indicator function, and $\alpha(\theta, \vartheta) \stackrel{\text{def}}{=} \min(1, \frac{\pi(\vartheta)}{\pi(\theta)})$ is the acceptance probability.

We assume that π is the posterior distribution from a logistic regression model. More precisely, we assume that we have binary responses $y_i \in \{0, 1\}$, where

$$y_i \sim \mathcal{B}\big(p\big(x_i'\theta\big)\big), \qquad i = 1, \dots, n,$$

and $x_i \in \mathbb{R}^d$ is a vector of covariate, and $\theta \in \mathbb{R}^d$ is the vector of parameter. $\mathcal{B}(p)$ denotes the Bernoulli distribution with parameter $p \in (0, 1)$, and $p(x) = \frac{\mathrm{e}^x}{1+\mathrm{e}^x}$ is the cdf of the logistic distribution. Let $X \in \mathbb{R}^{n\times d}$ denote the matrix with ith row x_i'. Let $\ell(\theta|X)$ denotes the log-likelihood

Table 2. Coverage probability and half-length for fixed-b confidence intervals

	Coverage	Half-length
$w(u) = (1 - u^2)_+$	0.945 ± 0.03	0.10 ± 0.01
Parzen	0.94 ± 0.03	0.03 ± 0.002
Bartlett	0.955 ± 0.03	0.02 ± 0.001

function of the model. We assume a Gaussian prior $\mathbf{N}(0, s^2 I_d)$ for θ, with $s = 20$. The posterior distribution of θ then becomes

$$\pi(\theta|X) \propto \mathrm{e}^{\ell(\theta|X)} \mathrm{e}^{-1/(2s^2)|\theta|^2}.$$

It is known that for this target distribution the RWM is geometrically ergodic (see, e.g., Atchadé [3], Section 5.2). Therefore, for all polynomial functions Theorem 2.2 holds. It is also known that with an appropriate choice of b_n, $\sigma^2_{b_n}$ converges in probability to $\sigma^2_P(h)$ (see, e.g., Atchadé [3], Theorem 4.1, and Corollary 4.1). So we will compare the fixed-b confidence intervals and the classical confidence intervals based on $\sigma^2_{b_n}$.

We simulate a Gaussian dataset with $n = 250$, $d = 15$, and simulate the components of the true value of β from a $\mathsf{U}(-10, 10)$. We first run the adaptive chain for 10^6 iterations and take the sample posterior mean of β as the "true" posterior mean. We focus on the coefficient β_1. Each sampler is run for 30 000 iterations, with no burn-in period. For the RMW, we use a covariance matrix $\Sigma = c\mathrm{I}_{15}$, where c is chosen such that the acceptance probability in stationarity is about 30%, obtained from a preliminary run.

From each sampler, we compute the fixed-b 95% confidence interval, and a classical 95% confidence interval. To explore the range of behavior of the classical procedure, we use $b_n = n^\delta$ for different values of $\delta \in (0, 1)$. To estimate coverage probability and half-length of these confidence intervals, $K = 200$ replications are performed. The result is summarized in Table 2 for the fixed-b procedure, and in Figure 1 for the classical procedure.

We see from the results that using $b_n = n$ gives very good coverage, except for the choice $w(u) = (1 - u^2)_+$, which generates significantly wider intervals. This is somewhat expected given the very heavy tail of the limiting distribution. The result also shows that the confidence interval procedure based on $\sigma^2_{b_n}$ works equally well when b_n is carefully chosen, but can perform poorly otherwise.

We also test the conclusion of Theorem 2.6 by comparing the finite sample convergence rate of the two confidence interval procedures. Here, we use only the Bartlett function. For the standard procedure, we use the best choice of δ ($\delta \approx 0.66$), as given by the previous simulation. We compute the confidence intervals after MCMC runs of length n, where $n \in \{100, \dots, 10^4\}$. Each run is repeated 30 times to approximate the coverage probabilities and interval lengths. The result is plotted on Figure 2, and is consistent with Theorem 2.6 that the fixed-b procedure has faster convergence. The price to pay is a (slightly) wider interval length as seen on Figure 2.

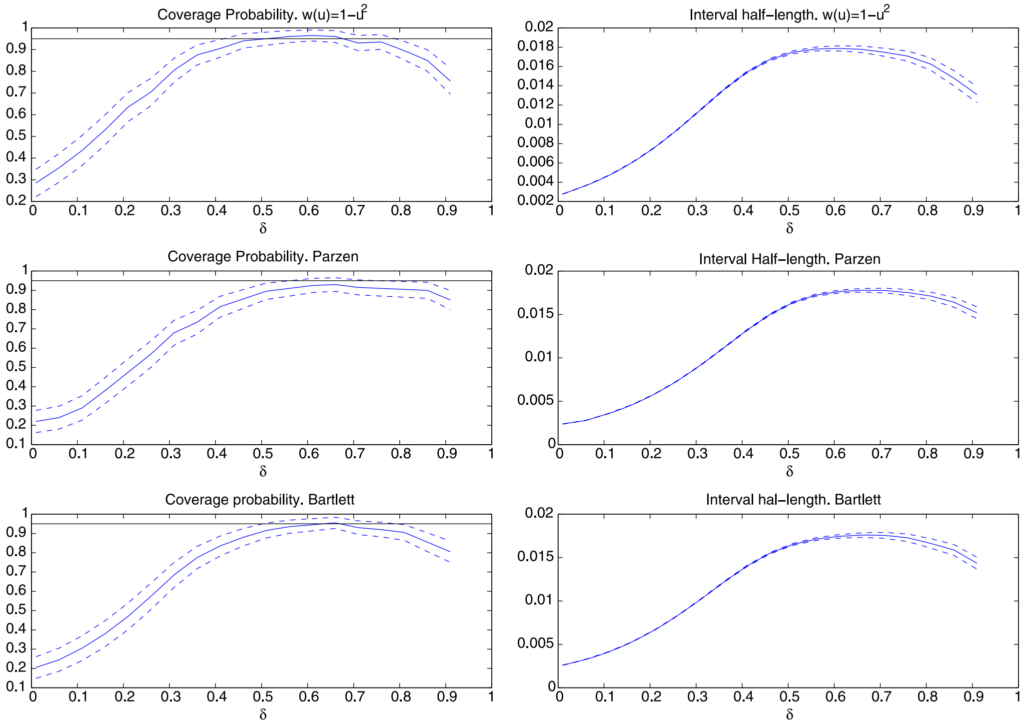

Figure 1. Coverage probability and confidence interval half-length for parameter β_1 for different values of δ using $\sigma^2_{b_n}$, and $b_n = n^{\delta}$. The dashed line is the 95% confidence band estimated from 200 replications.

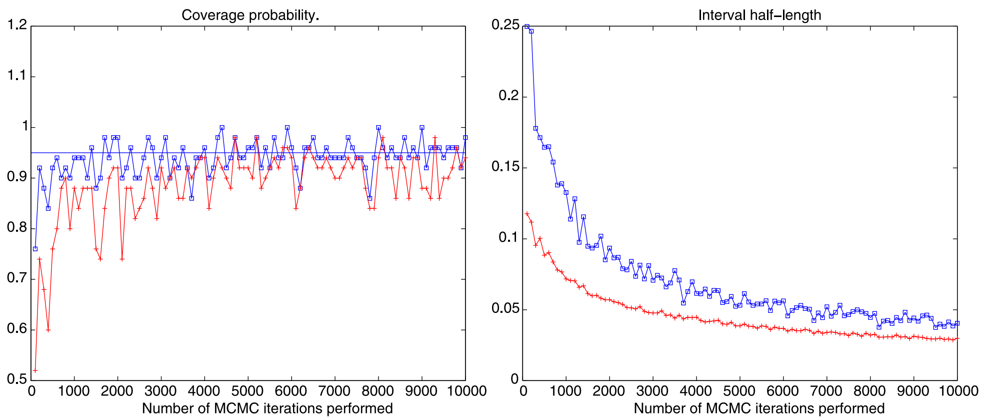

Figure 2. Coverage probability and confidence interval half-length for parameter β_1 as function of number of MCMC iterations. The square-line corresponds to using σ^2_n.

3. Proofs

3.1. Proof of Theorem 2.2

Let ϕ as in (7). Assumption A2 and Mercer's theorem implies that the kernel ϕ has a countable number of positive eigenvalues $\{\alpha_i, i \in \mathsf{I}\}$ with associated eigenfunctions $\{\Psi_j, j \in \mathsf{I}\}$ such that

$$\phi(s,t) = \sum_{j \in \mathsf{I}} \alpha_j \Psi_j(s)\Psi_j(t), \qquad (s,t) \in [0,1] \times [0,1], \tag{14}$$

where the convergence of the series is uniform on $[0,1] \times [0,1]$. Since $\int_0^1 \phi(s,t)\,\mathrm{d}t = 0$, 0 is also an eigenvalue of ϕ with eigenfunction $\Psi_0(x) \equiv 1$. Hence, we define $\bar{\mathsf{I}} = \{0\} \cup \mathsf{I}$, $\alpha = \{\alpha_j, j \in \bar{\mathsf{I}}\}$, with $\alpha_0 = 0$, and $\ell^2(\alpha)$ the associated Hilbert space of real numbers sequences $\{x_j, j \in \bar{\mathsf{I}}\}$ such that $\sum_j x_j^2 < \infty$, equipped with the norm $\|x\|_\alpha = \sqrt{\sum_j \alpha_j x_j^2}$ and the inner product $\langle x, y\rangle_\alpha \stackrel{\text{def}}{=} \sum_j \alpha_j x_j y_j$. We will need the differentiability of the eigenfunction Ψ_j. This is given by Kadota's theorem (Kadota [16]). Under the assumption that w is continuously twice differentiable, the eigenfunctions Ψ_j, $j \in \mathsf{I}$ are continuously differentiable (with derivative Ψ') and

$$\frac{\partial^2}{\partial s \partial t}\phi(s,t) = \sum_{j \in \mathsf{I}} \alpha_j \Psi_j'(s)\Psi_j'(t), \qquad (s,t) \in [0,1] \times [0,1], \tag{15}$$

where again the convergence of the series is uniform on $[0,1] \times [0,1]$. The expansions (14) and (15) easily imply that

$$\sum_{j \in \mathsf{I}} \alpha_j < \infty, \qquad \sup_{t \in [0,1]} \sum_{j \in \mathsf{I}} \alpha_j \left|\Psi_j(t)\right|^2 < \infty \quad \text{and} \\ \sup_{t \in [0,1]} \sum_{j \in \mathsf{I}} \alpha_j \left|\Psi_j'(t)\right|^2 < \infty. \tag{16}$$

It is easy to check that σ_n^2 can also be written as

$$\begin{aligned} \sigma_n^2 &= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n w\left(\frac{i-j}{n}\right)\bigl(\bar{h}(X_i) - \pi_n(\bar{h})\bigr)\bigl(\bar{h}(X_j) - \pi_n(\bar{h})\bigr) \\ &= \frac{1}{n}\sum_{i=1}^n \sum_{j=1}^n \left\{ w\left(\frac{i-j}{n}\right) - v_{n,i} - v_{n,j} + u_n \right\}\bar{h}(X_i)\bar{h}(X_j), \end{aligned}$$

where $v_{n,i} = n^{-1}\sum_{\ell=1}^n w(\frac{i-\ell}{n})$, and $u_n = n^{-2}\sum_{i=1}^n \sum_{=1}^n w(\frac{i-j}{n})$. Notice that $v_{n,i}$ is a Riemann sum approximation of $v(i/n)$, where $v(t) \stackrel{\text{def}}{=} \int_0^1 w(t-u)\,\mathrm{d}u$, and u_n approximates $\int_0^1\int_0^1 w(t -$

$u)\,\mathrm{d}u\,\mathrm{d}t = \int_0^1 v(t)\,\mathrm{d}t$. In view of this, we introduce

$$\begin{aligned}\check{\sigma}_n^2 &\stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\left\{w\left(\frac{i-j}{n}\right) - v\left(\frac{i-1}{n}\right) - v\left(\frac{j-1}{n}\right) + \int_0^1 v(t)\,\mathrm{d}t\right\}\bar{h}(X_i)\bar{h}(X_j) \\ &= \frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n \phi\left(\frac{i-1}{n}, \frac{j-1}{n}\right)\bar{h}(X_i)\bar{h}(X_j) = \sum_{\ell\in\mathsf{I}}\alpha_\ell\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\bar{h}(X_i)\right)^2.\end{aligned} \tag{17}$$

The last equality uses the Mercer's expansion for ϕ as given in (14). This implies that

$$\begin{aligned}\mathbf{T}_n &= \frac{\sum_{i=1}^n \bar{h}(X_i)}{\sigma_n\sqrt{n}} \\ &= \frac{1/(\sigma_P(h)\sqrt{n})\sum_{i=1}^n \Psi_0((i-1)/n)\bar{h}(X_i)}{\sqrt{\sum_{\ell\in\mathsf{I}}\alpha_\ell(1/(\sigma_P(h)\sqrt{n})\sum_{i=1}^n \Psi_\ell((i-1)/n)\bar{h}(X_i))^2 + (\sigma_n^2 - \check{\sigma}_n^2)/\sigma_P^2(h)}}.\end{aligned}$$

Hence, the proof of the theorem boils down to the limiting behavior of the $\ell^2(\alpha)$-valued process

$$\left\{\frac{1}{\sigma_P(h)\sqrt{n}}\sum_{i=1}^n \Psi_j\left(\frac{i-1}{n}\right)\bar{h}(X_i),\ j\in\bar{\mathsf{I}}\right\},$$

and the remainder $(\sigma_n^2 - \check{\sigma}_n^2)$. In Lemma 3.5, we show that $\{\frac{1}{\sigma_P(h)\sqrt{n}}\sum_{i=1}^n \Psi_\ell(\frac{i-1}{n})\bar{h}(X_i), \ell\in\bar{\mathsf{I}}\}$ converges weakly to $\{Z_\ell, \ell\in\bar{\mathsf{I}}\}$, and that $\sigma_n^2 - \check{\sigma}_n^2$ converges in probability to zero. This is done first in the stationary case in Lemmas 3.3–3.4, and in the nonstationary case in Lemma 3.5. Hence, the theorem follows by applying Slutszy's theorem and the continuous mapping theorem. Everything rely on a refinement of the martingale approximation of Kipnis and Varadhan [19] that we establish first in Lemma 3.2.

3.1.1. *Martingale approximation for Markov chains*

Throughout this section, unless stated otherwise, $\{X_n, n\geq 0\}$ denotes a stationary reversible Markov chain with invariant distribution π and transition kernel P, and we fix $h\in L_0^2(\pi)$. We denote $\mathcal{F}_n \stackrel{\text{def}}{=} \sigma(X_0,\ldots,X_n)$. We introduce the probability measure $\bar{\pi}(\mathrm{d}x,\mathrm{d}y) = \pi(\mathrm{d}x)P(x,\mathrm{d}y)$ on $\mathsf{X}\times\mathsf{X}$, and we denote $L^2(\bar{\pi})$ the associated L^2-space with norm $|||f|||^2 \stackrel{\text{def}}{=} \iint |f(x,y)|^2 \times \pi(\mathrm{d}x)P(x,\mathrm{d}y)$. For $\varepsilon > 0$, define

$$U_\varepsilon(x) \stackrel{\text{def}}{=} \sum_{j\geq 0}\frac{1}{(1+\varepsilon)^{j+1}}P^j h(x), \qquad G_\varepsilon(x,y) \stackrel{\text{def}}{=} U_\varepsilon(y) - PU_\varepsilon(x).$$

Since P is a contraction of $L_0^2(\pi)$, it is clear that $U_\varepsilon\in L^2(\pi)$, and $G_\varepsilon\in L^2(\bar{\pi})$. Furthermore, for all $\varepsilon > 0$,

$$\|U_\varepsilon\| \leq \varepsilon^{-1}\|h\| \quad \text{and} \quad |||G_\varepsilon||| \leq 2\|U_\varepsilon\|. \tag{18}$$

When $\sigma_P^2(h) < \infty$ a stronger conclusion is possible, and this is the key observation made by Kipnis and Varadhan [19], Theorem 1.3. We summarize their result as follows.

Lemma 3.1 (Kipnis and Varadhan [19]). *Suppose that* $h \in L_0^2(\pi)$, *and* $\sigma_P^2(h) < \infty$. *Then for any sequence* $\{\varepsilon_n, n \ge 0\}$ *of positive numbers such that* $\lim_n \varepsilon_n = 0$,

$$\lim_{n\to\infty} \sqrt{\varepsilon_n} \|U_{\varepsilon_n}\| = 0.$$

Furthermore, there exists $G \in L^2(\bar{\pi})$, *with* $\int P(x, \mathrm{d}z)G(x, z) = 0$ (π-*a.e.*) *such that* $\sigma_P^2(h) = |||G|||^2$, *and* $\lim_n |||G_{\varepsilon_n} - G||| = 0$.

For $n \ge 1$, define the process

$$B_n(t) = \frac{1}{\sigma_P(h)\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} G(X_i, X_{i-1}), \qquad 0 \le t \le 1,$$

and let $\{B(t), 0 \le t \le 1\}$ denotes the standard Brownian motion. It is an easy consequence of Lemma 3.1 that $\{G(X_i, X_{i-1}), 1 \le i \le n\}$ is a stationary martingale difference sequence with finite variance. Therefore, by the weak invariance principle for stationary martingales, $B_n \stackrel{\mathrm{w}}{\to} B$ in $D[0, 1]$ equipped with the Skorohod metric. In Corollary 1.5, [19], it is shown that the Markov chain $\{X_n, n \ge 0\}$ inherits this weak invariance principle. For the purpose of this paper, we need some refinements of this result. Let $\{a_{n,k}, 0 \le k \le n\}$ be a sequence of real numbers. Set $|a_n|_\infty \stackrel{\mathrm{def}}{=} \sup_{0\le k\le n} |a_{n,k}|$, and $|a_n|_{\mathrm{tv}} \stackrel{\mathrm{def}}{=} \sum_{k=1}^n |a_{n,k} - a_{n,k-1}|$.

Lemma 3.2. *Let* $h \in L_0^2(\pi)$ *be such that* $\sigma_P^2(h) < \infty$.

(1) *If* $|a_n|_\infty + |a_n|_{\mathrm{tv}}$ *is bounded in* n, *then*

$$\sum_{i=1}^n a_{n,i-1} h(X_i) = \sum_{i=1}^n a_{n,i-1} G(X_i, X_{i-1}) + R_n, \tag{19}$$

where $n^{-1}\mathbb{E}(|R_n|^2) \to 0$ *as* $n \to \infty$.

(2) *If* $f : [0, 1] \to \mathbb{R}$ *is a continuously differentiable function, then* $\frac{1}{\sigma_P(h)\sqrt{n}} \sum_{i=1}^n f(\frac{i-1}{n}) h(X_i)$ *converges weakly to* $\int_0^1 f(t)\, \mathrm{d}B(t)$, *as* $n \to \infty$.

Proof. Set $S_n \stackrel{\mathrm{def}}{=} \sum_{i=1}^n a_{n,i-1} h(X_i)$. The function U_ε satisfies $(1 + \varepsilon)U_\varepsilon(x) - PU_\varepsilon(x) = h(x)$, π-a.e. $x \in \mathsf{X}$. This is used to write

$$\begin{aligned} a_{n,k-1} h(X_k) &= a_{n,k-1}\big(\varepsilon U_\varepsilon(X_k) + U_\varepsilon(X_k) - PU_\varepsilon(X_k)\big) \\ &= a_{n,k-1}\varepsilon U_\varepsilon(X_k) + a_{n,k-1}\big(U_\varepsilon(X_k) - PU_\varepsilon(X_{k-1})\big) \\ &\quad + \big(a_{n,k-1} PU_\varepsilon(X_{k-1}) - a_{n,k} PU_\varepsilon(X_k)\big) + (a_{n,k} - a_{n,k-1}) PU_\varepsilon(X_k). \end{aligned}$$

It follows that

$$S_n = \varepsilon \sum_{k=1}^{n} a_{n,k-1} U_\varepsilon(X_k) + \sum_{k=1}^{n} a_{n,k-1} G(X_k, X_{k-1}) + \sum_{k=1}^{n} a_{n,k-1}\bigl(G_\varepsilon(X_k, X_{k-1}) - G(X_k, X_{k-1})\bigr) + \bigl(a_{n,0} PU_\varepsilon(X_0) - a_{n,n} PU_\varepsilon(X_n)\bigr) + \sum_{k=1}^{n} (a_{n,k} - a_{n,k-1}) PU_\varepsilon(X_k),$$

which is valid for any $\varepsilon > 0$. In particular with $\varepsilon = \varepsilon_n = 1/n$, we have

$$S_n = \sum_{k=1}^{n} a_{n,k-1} G(X_k, X_{k-1}) + \sum_{k=1}^{n} a_{n,k-1}\bigl(G_{\varepsilon_n}(X_k, X_{k-1}) - G(X_k, X_{k-1})\bigr) + R_n^{(1)} + R_n^{(2)} + R_n^{(3)},$$

where

$$R_n^{(1)} \stackrel{\text{def}}{=} \varepsilon_n \sum_{k=1}^{n} a_{n,k-1} U_{\varepsilon_n}(X_k), \qquad R_n^{(2)} \stackrel{\text{def}}{=} \bigl(a_{n,0} PU_{\varepsilon_n}(X_0) - a_{n,n} PU_{\varepsilon_n}(X_n)\bigr) \quad \text{and}$$

$$R_n^{(3)} \stackrel{\text{def}}{=} \sum_{k=1}^{n} (a_{n,k} - a_{n,k-1}) PU_{\varepsilon_n}(X_k).$$

By stationarity and the martingale property,

$$\frac{1}{n}\mathbb{E}\left[\left(\sum_{k=1}^{n} a_{n,k-1}\bigl(G_\varepsilon(X_k, X_{k-1}) - G(X_k, X_{k-1})\bigr)\right)^2\right] = |\!|\!| G_{\varepsilon_n} - G |\!|\!|^2 \frac{1}{n}\sum_{k=1}^{n} a_{n,k-1}^2 \to 0,$$

using Lemma 3.1, and the assumption on a_n. The other remainders are also easily dealt with.

$$\frac{1}{\sqrt{n}}\mathbb{E}^{1/2}\bigl(\bigl|R_n^{(3)}\bigr|^2\bigr) \le \frac{1}{\sqrt{n}}\sum_{k=1}^{n} |a_{n,k} - a_{n,k-1}| \mathbb{E}^{1/2}\bigl(\bigl|PU_{\varepsilon_n}(X_k)\bigr|^2\bigr) = \sqrt{\varepsilon_n}\|U_{\varepsilon_n}\| |a_n|_{\mathrm{tv}} \to 0,$$

using Lemma 3.1 and the assumption on a_n. Similarly,

$$\frac{1}{\sqrt{n}}\mathbb{E}^{1/2}\bigl(\bigl|R_n^{(2)}\bigr|^2\bigr) \le 2|a_n|_\infty \sqrt{\varepsilon_n}\|U_{\varepsilon_n}\| \to 0 \quad \text{and}$$

$$\frac{1}{\sqrt{n}}\mathbb{E}^{1/2}\bigl(\bigl|R_n^{(1)}\bigr|^2\bigr) \le \sqrt{\varepsilon_n}\|U_{\varepsilon_n}\| \frac{1}{n}\sum_{k=1}^{n} |a_{n,k-1}| \to 0.$$

This proves part (1) of the lemma. For part (2), we use part (1) with $a_{n,i} = f(i/n)$ to conclude that

$$\begin{aligned}\frac{1}{\sigma_P(h)\sqrt{n}}&\sum_{i=1}^{n} f\left(\frac{i-1}{n}\right)h(X_i)\\&=\frac{1}{\sigma_P(h)\sqrt{n}}\sum_{i=1}^{n} f\left(\frac{i-1}{n}\right)G(X_i, X_{i-1}) + \mathrm{o}_p(1)\\&=\int_0^1 f(t)\,\mathrm{d}B_n(t) + \mathrm{o}_p(1),\end{aligned}$$

where $A_n = \mathrm{o}_p(1)$ means that A_n converges in probability to zero as $n \to \infty$. To conclude the proof, it suffices to show that $\int_0^1 f(t)\,\mathrm{d}B_n(t)$ converges weakly to $\int_0^1 f(t)\,\mathrm{d}B(t)$. This follows from the weak convergence continuous mapping theorem by noticing that B has continuous sample path (almost surely), and the map $D[0,1] \to \mathbb{R}$, $x \mapsto \int_0^1 f(t)\,\mathrm{d}x(t)$ is continuous at all points $x_0 \in C[0,1]$, where the integral $\int_0^1 f(t)\,\mathrm{d}x(t)$ is understood as a Riemann–Stietjes integral. To see the continuity, take $\{x_n\}$ a sequence of elements in $D[0,1]$ that converges to $x_0 \in C[0,1]$ in the Skorohod metric. Since $x_0 \in C[0,1]$, the sequence $\{x_n\}$ converges to x_0 in $C[0,1]$ as well. By integration by part, $\int_0^1 f(t)\,\mathrm{d}x_n(t) = f(1)x_n(1) - f(0)x_n(0) - \int_0^1 x_n(t)f'(t)\,\mathrm{d}t$, and

$$\left|\int_0^1 f(t)\,\mathrm{d}x_n(t) - \int_0^1 f(t)\,\mathrm{d}x_0(t)\right| \le |x_n - x_0|_\infty\left(2|f|_\infty + \int_0^1 \left|f'(t)\right|\mathrm{d}t\right) \to 0,$$

as $n \to \infty$. □

Lemma 3.3. *Let $h \in L_0^2(\pi)$ be such that $\sigma_P^2(h) < \infty$. Define*

$$Z^{(n)} \stackrel{\text{def}}{=} \left\{\frac{1}{\sigma_P(h)\sqrt{n}}\sum_{i=1}^{n}\Psi_j\left(\frac{i-1}{n}\right)\bar{h}(X_i),\ j \in \bar{\mathrm{I}}\right\} \quad \text{and} \quad Z \stackrel{\text{def}}{=} \left\{\int_0^1 \Psi_j(t)\,\mathrm{d}B(t),\ j \in \bar{\mathrm{I}}\right\}.$$

Then as $n \to \infty$, $Z^{(n)}$ converges weakly to Z in $\ell^2(\alpha)$.

Proof. We need to show that for all $u \in \ell^2(\alpha)$, $\langle Z^{(n)}, u\rangle_\alpha \stackrel{\mathrm{w}}{\to} \langle Z, u\rangle_\alpha$, and that $\{Z^{(n)}\}$ is tight.

For $u \in \ell^2(\alpha)$, $\langle Z^{(n)}, u\rangle_\alpha = \frac{1}{\sigma_P(h)\sqrt{n}}\sum_{i=1}^n f_u(\frac{i-1}{n})\bar{h}(X_i)$, where $f_u(t) = \sum_j \alpha_j u_j \Psi_j(t)$. From basic results in calculus, it follows from Kadota's theorem that f_u is continuously differentiable on $[0,1]$. Hence, by Lemma 3.2, part (2), $\langle Z^{(n)}, u\rangle_\alpha \stackrel{\mathrm{w}}{\to} \int_0^1 f_u(t)\,\mathrm{d}B(t) = \langle u, Z\rangle_\alpha$. To show that $\{Z^{(n)}\}$ is tight, it suffices to show that

$$\lim_{N\to\infty}\sup_{n\ge 1}\mathbb{E}\left(\sum_{j=N}^{\infty}\left\langle Z^{(n)}, e_j\right\rangle_\alpha^2\right) = 0. \tag{20}$$

We have

$$\mathbb{E}\left(\langle Z^{(n)}, e_j\rangle_\alpha^2\right) = \frac{\alpha_j}{\sigma_P^2(h)n}\sum_{i=1}^n\sum_{k=1}^n \Psi_j\left(\frac{i-1}{n}\right)\Psi_j\left(\frac{k-1}{n}\right)\pi\left(hP^{|i-k|}h\right)$$
$$= \frac{\alpha_j}{\sigma_P^2(h)n}\int_{-1}^1\sum_{i=1}^n\sum_{k=1}^n \Psi_j\left(\frac{i-1}{n}\right)\Psi_j\left(\frac{k-1}{n}\right)\lambda^{|i-k|}\mu_h(\mathrm{d}\lambda).$$

By Fubini's theorem, for $N \geq 1$,

$$\mathbb{E}\left(\sum_{j=N}^\infty \langle Z^{(n)}, e_j\rangle_\alpha^2\right) = \frac{1}{\sigma_P^2(h)n}\int_{-1}^1\sum_{i=1}^n\sum_{k=1}^n\sum_{j=N}^\infty \alpha_j\Psi_j\left(\frac{i-1}{n}\right)\Psi_j\left(\frac{k-1}{n}\right)\lambda^{|i-k|}\mu_h(\mathrm{d}\lambda).$$

Let $\varepsilon > 0$. By uniform convergence of the series $\sum_j \alpha_j\Psi_j(s)\Psi_j(t)$, we can find N_0 such that for any $N \geq N_0$ and for all $s, t \in [0,1]$, $|\sum_{\ell\geq N}\alpha_\ell\Psi_\ell(t)\Psi_\ell(s)| \leq \varepsilon$. So that for all $n \geq 1$,

$$\mathbb{E}\left(\sum_{\ell=N}^\infty \langle Z^{(n)}, e_\ell\rangle_\alpha^2\right) \leq \frac{\varepsilon}{\sigma_P^2(h)n}\int_{-1}^1\sum_{i=1}^n\sum_{j=1}^n \lambda^{|i-j|}\mu_h(\mathrm{d}\lambda) \leq \frac{\varepsilon}{\sigma_P^2(h)}\int_{-1}^1 \frac{1+\lambda}{1-\lambda}\mu_h(\mathrm{d}\lambda) = \varepsilon,$$

since $\varepsilon > 0$ is arbitrary, this proves (20). □

Lemma 3.4. *Let $h \in L_0^2(\pi)$ be such that $\sigma_P^2(h) < \infty$. Then as $n \to \infty$, $\mathbb{E}(|\sigma_n^2 - \check{\sigma}_n^2|) = \mathrm{O}(1/n)$. Hence $\sigma_n^2 - \check{\sigma}_n^2$ converges in probability to 0, as $n \to \infty$.*

Proof. Comparing the expression of σ_n^2 and $\check{\sigma}_n^2$, we see that

$$\sigma_n^2 - \check{\sigma}_n^2 = \left(u_n - \int_0^1 v(t)\,\mathrm{d}t\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{h}(X_i)\right)^2 - 2\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{h}(X_i)\right)\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(v_{n,i} - v\left(\frac{i-1}{n}\right)\right)\bar{h}(X_i)\right). \tag{21}$$

Since the sequence $\mathbb{E}[(\frac{1}{\sqrt{n}}\sum_{i=1}^n \bar{h}(X_i))^2]$ converges to the finite limit $\sigma^2(h)$ by assumption, it is bounded, and there exists a finite constant c_1 such that

$$\mathbb{E}\left(|\sigma_n^2 - \check{\sigma}_n^2|\right) \leq c_1^2\left|u_n - \int_0^1 v(t)\,\mathrm{d}t\right| + \frac{2c_1}{n}\mathbb{E}^{1/2}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n n\left(v_{n,i} - v\left(\frac{i-1}{n}\right)\right)h(X_i)\right)^2\right].$$

Set $a_{n,0} = 0$, $a_{n,i} \stackrel{\text{def}}{=} n(v_{n,i} - v(\frac{i-1}{n}))$. We recall that $v_{n,i} = n^{-1}\sum_{\ell=1}^{n} w(\frac{i-\ell}{n})$, and $v(t) = \int_0^1 w(t-u)\,\mathrm{d}u$, and write

$$\begin{aligned} a_{n,i} &= n\sum_{\ell=1}^{n}\int_{(\ell-1)/n}^{\ell/n}\left[w\left(\frac{i-1}{n}-\frac{\ell-1}{n}\right)-w\left(\frac{i-1}{n}-u\right)\right]\mathrm{d}u \\ &= n\sum_{\ell=1}^{n}\int_{(\ell-1)/n}^{\ell/n}\left(\frac{\ell-1}{n}-u\right)\int_0^1 w'\left(\frac{i-1}{n}-\frac{\ell-1}{n}-t\left(u-\frac{\ell-1}{n}\right)\right)\mathrm{d}t\,\mathrm{d}u. \end{aligned}$$

Using this expression, it is easy to show that $|a_n|_\infty \le |w'|_\infty/2$. And since w is of class C^2, a mean-value theorem on w' using the above expression shows that $|a_n|_{\mathrm{tv}} = |a_{n,1}| + \sum_{i=2}^{n}|a_{n,i} - a_{n,i-1}| \le (|w'|_\infty + |w''|_\infty)/2$. We are then in position to apply Lemma 3.2(1) to obtain

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}a_{n,i}\bar{h}(X_i)\right)^2\right] = \mathrm{O}(1).$$

By similar arguments as above, and since $u_n = n^{-2}\sum_{i=1}^{n}\sum_{=1}^{n} w(\frac{i-j}{n})$ is a Riemann sum approximation of $\int_0^1 v(t)\,\mathrm{d}t$, we obtain that $|u_n - \int_0^1 v(t)\,\mathrm{d}t| = \mathrm{O}(\frac{1}{n})$. In conclusion,

$$\mathbb{E}\left(\left|\sigma_n^2 - \check{\sigma}_n^2\right|\right) = \mathrm{O}\left(\frac{1}{n}\right). \tag{22}$$

□

Lemma 3.5. *Assume* A1. *Suppose that the Markov chain* $\{X_n, n \ge 0\}$ *starts at* $X_0 = x$ *for* $x \in \mathsf{X}$ *such that* (1) *holds. Let* $h \in L_0^2(\pi)$ *be such that* $\sigma_P^2(h) < \infty$. *Then as* $n \to \infty$, $\sigma_n^2 - \check{\sigma}_n^2$ *converges in probability to zero, and* $Z^{(n)} \xrightarrow{\mathsf{w}} Z$ *in* $\ell^2(\alpha)$.

Proof. Ergodicity is equivalent to the existence of a successful coupling of the Markov chain and its stationary copy. More precisely, we can construct a process $\{(X_n, \tilde{X}_n), n \ge 0\}$ such that $\{X_n, n \ge 0\}$ is a Markov chain with initial distribution δ_x and transition kernel P, $\{\tilde{X}_n, n \ge 0\}$ is a Markov chain with initial distribution π and transition kernel P, and there exists a finite (coupling) time τ such that $X_n = \tilde{X}_n$ for all $n \ge \tau$. For a proof of this result, see for instance Lindvall [20], Theorem 14.10; see also Roberts and Rosenthal [27], Proposition 28. We use a wide "tilde" to denote quantities computed from the stationary chain $\{\tilde{X}_n, n \ge 0\}$.

Since $X_n = \tilde{X}_n$ for all $n \ge \tau$, and in view of the expression of $\sigma_n^2 - \check{\sigma}_n^2$ given in (21), it is straightforward to check that $\sigma_n^2 - \check{\sigma}_n^2 - (\widetilde{\sigma_n^2 - \check{\sigma}_n^2})$ converges to zero in probability. The convergence of $\|Z^{(n)} - \widetilde{Z^{(n)}}\|_\alpha$ is handled similarly.

$$\begin{aligned} &\left\|Z^{(n)} - \widetilde{Z^{(n)}}\right\|_\alpha^2 \\ &\quad = \sum_{\ell\in\mathsf{I}}\alpha_\ell\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\Psi_\ell\left(\frac{k}{n}\right)\left(h(X_k) - h(\tilde{X}_k)\right)\right)^2 \end{aligned}$$

$$= \sum_{\ell \in \mathsf{I}} \alpha_\ell \left(\frac{1}{\sqrt{n}} \sum_{k=1}^{\tau-1} \Psi_\ell\left(\frac{k}{n}\right)\bigl(h(X_k) - h(\tilde{X}_k)\bigr)\right)^2$$

$$\le \frac{\tau}{n}\left(\sup_{t\in[0,1]} \sum_{\ell \in \mathsf{I}} \alpha_\ell \bigl|\Psi_\ell(t)\bigr|^2\right)\left(\sum_{k=1}^{\tau}\bigl(h(X_k) - h(\tilde{X}_k)\bigr)^2\right),$$

which converges almost surely to zero, given (16), and since τ is finite almost surely. □

3.2. Proof of Theorem 2.4

Since u is bounded from below, we can choose $a = 1 - \inf_{x\in\mathsf{X}} u(x)$ such that $V(x) \stackrel{\text{def}}{=} a + u(x) \ge 1$. Let $q_\sigma(x, y)$ be the density of the proposal $\mathbf{N}(x - \frac{\sigma^2}{2}\rho(x)\nabla u(x), \sigma^2 I_d)$, and define $\mathsf{R}(x) \stackrel{\text{def}}{=} \{y \in \mathbb{R}^p\colon \alpha(x, y) < 1\}$. We have

$$\begin{aligned} PV(x) - V(x) &= \int \alpha(x, y)\bigl(V(y) - V(x)\bigr)q_\sigma(x, y)\,\mathrm{d}y \\ &= \int_{\mathsf{R}(x)} \bigl[\alpha(x, y) - 1\bigr]\bigl(V(y) - V(x)\bigr)q_\sigma(x, y)\,\mathrm{d}y \\ &\quad + \int \bigl(V(y) - V(x)\bigr)q_\sigma(x, y)\,\mathrm{d}y. \end{aligned} \tag{23}$$

Since ∇u is Lipschitz, with Lipschitz constant L, say, we have by Taylor expansion

$$V(y) - V(x) \le \bigl\langle \nabla u(x), y - x\bigr\rangle_2 + \frac{L}{2}|y - x|^2.$$

Integrating both sides, and using the fact that $\rho(x)|\nabla u(x)| \le \tau$, we get

$$\begin{aligned} \int \bigl(V(y) - V(x)\bigr)q_\sigma(x, y)\,\mathrm{d}y &\le -\frac{\sigma^2}{2}\rho(x)\bigl|\nabla u(x)\bigr|^2 + \frac{L}{2}\left(\frac{\sigma^4}{4}\rho(x)^2\bigl|\nabla u(x)\bigr|^2 + \mathrm{d}\sigma^2\right) \\ &\le -\frac{\sigma^2}{2}\rho(x)\bigl|\nabla u(x)\bigr|^2 + \frac{L}{2}\left(\frac{\tau^2\sigma^4}{4} + \mathrm{d}\sigma^2\right). \end{aligned} \tag{24}$$

We also have

$$\begin{aligned} &\frac{\pi(y)}{\pi(x)} \frac{q_\sigma(y, x)}{q_\sigma(x, y)} \\ &\quad = \exp\left(V(x) - V(y) - \frac{1}{2\sigma^2}\left|x - y + \frac{\sigma^2}{2}\rho(y)\nabla u(y)\right|^2 + \frac{1}{2\sigma^2}\left|y - x - \frac{\sigma^2}{2}\rho(x)\nabla u(x)\right|^2\right). \end{aligned}$$

If $y \in \mathsf{R}(x)$, we necessarily have $\frac{\pi(y)}{\pi(x)}\frac{q_\sigma(y,x)}{q_\sigma(x,y)} < 1$, which translates to

$$V(y) - V(x) > -\frac{1}{2\sigma^2}\left|x - y + \frac{\sigma^2}{2}\rho(y)\nabla u(y)\right|^2 + \frac{1}{2\sigma^2}\left|y - x - \frac{\sigma^2}{2}\rho(x)\nabla u(x)\right|^2.$$

Hence, if $y \in \mathsf{R}(x)$,

$$\begin{aligned}
&\big[\alpha(x,y) - 1\big]\big(V(y) - V(x)\big) \\
&\quad\le \big[\alpha(x,y) - 1\big]\left(-\frac{1}{2\sigma^2}\left|x - y + \frac{\sigma^2}{2}\rho(y)\nabla u(y)\right|^2 + \frac{1}{2\sigma^2}\left|y - x - \frac{\sigma^2}{2}\rho(x)\nabla u(x)\right|^2\right) \\
&\quad= \big[1 - \alpha(x,y)\big]\frac{\sigma^2}{8}\Big(\rho^2(y)\big|\nabla u(y)\big|^2 - \rho^2(x)\big|\nabla u(x)\big|^2 \\
&\qquad\qquad - \frac{2}{\sigma^2}\big\langle y - x, \rho(x)\nabla u(x) + \rho(y)\nabla u(y)\big\rangle\Big) \\
&\quad\le \frac{\sigma^2}{8}\left(\tau^2 + \frac{4\tau}{\sigma^2}|y - x|\right).
\end{aligned}$$

Hence,

$$\int_{\mathsf{R}(x)} \big[\alpha(x,y) - 1\big]\big(V(y) - V(x)\big)q_\sigma(x,y)\,\mathrm{d}y \le \frac{\sigma^2\tau^2}{8} + \frac{\tau}{2}\sqrt{\mathrm{d}\sigma^2 + \frac{\sigma^2\tau^2}{2}}. \tag{25}$$

We combine (23)–(25) to conclude that

$$PV(x) - V(x) \le -\frac{\sigma^2}{2}\rho(x)\big|\nabla u(x)\big|^2 + K,$$

where $K = \frac{L}{2}(\frac{\tau^2\sigma^4}{4} + \mathrm{d}\sigma^2) + \frac{\sigma^2\tau^2}{8} + \frac{\tau}{2}\sqrt{\mathrm{d}\sigma^2 + \frac{\sigma^2\tau^2}{2}}$. Since $f(x) \stackrel{\text{def}}{=} \frac{\sigma^2}{2}\rho(x)|\nabla u(x)|^2$ is continuous and $f(x) \to \infty$, as $\|x\| \to \infty$ by assumption, the results follow readily.

3.3. Proof Theorem 2.6

We follow Dedecker and Rio [7], Theorem 2.1. With the geometric ergodicity assumption, the martingale approximation to $\sum_{i=1}^n h(X_i)$ can be constructed more explicitly than in Lemmas 3.1 and 3.2. Define

$$g(x) = \sum_{j\ge 0} P^j\bar{h}(x), \qquad x \in \mathsf{X}.$$

By the geometric ergodicity assumption, g is well-defined and belongs to $\mathcal{L}_{V^\delta}$. Then we define $D_0 = 0$, and $D_k \stackrel{\text{def}}{=} g(X_k) - Pg(X_{k-1})$, $k \ge 1$. It is easy to see that $\{D_k, k \ge 0\}$ is a martingale-difference sequence with respect to the natural filtration of $\{X_n, n \ge 0\}$. Using this martingale, we define

$$\bar{\sigma}_n^2 \stackrel{\text{def}}{=} \sum_{\ell\in\mathsf{I}} \alpha_\ell \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right) D_i \right)^2,$$

and we recall that $\check{\sigma}_n^2 \stackrel{\text{def}}{=} \sum_{\ell\in\mathsf{I}} \alpha_\ell (\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell(\frac{i-1}{n}) h(X_i))^2$. Hence,

$$\sigma_n^2 = \sum_{\ell\in\mathsf{I}} \alpha_\ell \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right) D_i \right)^2 + \left(\sigma_n^2 - \check{\sigma}_n^2\right) + \left(\check{\sigma}_n^2 - \bar{\sigma}_n^2\right).$$

Although the martingales are constructed differently, the argument in Lemma 3.4 carries through and shows that $\mathbb{E}(|\sigma_n^2 - \check{\sigma}_n^2|) = \mathrm{O}(1/n)$. The proof is similar to the proof of Lemma 3.4 and is omitted. Also $\mathbb{E}(|\check{\sigma}_n^2 - \bar{\sigma}_n^2|) = \mathrm{O}(1/\sqrt{n})$. To see this, use the Cauchy–Schwarz inequalities for sequences in $\ell^2(\alpha)$ and for random variables to write

$$\begin{aligned}
&\mathbb{E}\left(\left|\check{\sigma}_n^2 - \bar{\sigma}_n^2\right|\right) \\
&\quad= \mathbb{E}\left[\left| \sum_{\ell\in\mathsf{I}} \alpha_\ell \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\left(h(X_i) - D_i\right)\right) \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\left(h(X_i) + D_i\right)\right) \right|\right] \\
&\quad\le \mathbb{E}\left[\left\{ \sum_{\ell\in\mathsf{I}} \alpha_\ell \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\left(h(X_i) - D_i\right)\right)^2 \right\}^{1/2} \right. \\
&\qquad\quad \left. \times \left\{ \sum_{\ell\in\mathsf{I}} \alpha_\ell \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\left(h(X_i) + D_i\right)\right)^2 \right\}^{1/2} \right] \\
&\quad\le \left\{ \sum_{\ell\in\mathsf{I}} \alpha_\ell \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\left(h(X_i) - D_i\right)\right)^2\right] \right\}^{1/2} \\
&\qquad \times \left\{ \sum_{\ell\in\mathsf{I}} \alpha_\ell \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\left(h(X_i) + D_i\right)\right)^2\right] \right\}^{1/2}.
\end{aligned}$$

By the martingale approximation, we have

$$\begin{aligned}
\sum_{i=1}^n \Psi_\ell\left(\frac{i-1}{n}\right)\left(h(X_i) - D_i\right) &= \Psi_\ell(0) Pg(X_0) - \Psi\left(\frac{n-1}{n}\right) Pg(X_n) \\
&\quad + \sum_{i=2}^n \left(\Psi_\ell\left(\frac{i-1}{n}\right) - \Psi_\ell\left(\frac{i-2}{n}\right)\right) Pg(X_{i-1}).
\end{aligned}$$

The details of these calculations can be found for instance in [4], Proposition A1. It is then easy to show that

$$\sum_{\ell\in\mathsf{I}}\alpha_\ell\mathbb{E}\left[\left(\sum_{i=1}^n\Psi_\ell\left(\frac{i-1}{n}\right)\big(h(X_i)-D_i\big)\right)^2\right] \\ \leq\left(6\sup_{0\leq t\leq 1}\sum_{\ell\in\mathsf{I}}\alpha_\ell\left|\Psi_\ell(t)\right|^2+3\sup_{0\leq t\leq 1}\sum_{\ell\in\mathsf{I}}\alpha_\ell\left|\Psi_\ell'(t)\right|^2\right)|h|_{V^\delta}^2.$$

For the second term, notice that

$$\sum_{i=1}^n\Psi_\ell\left(\frac{i-1}{n}\right)\big(h(X_i)+D_i\big)=2\sum_{i=1}^n\Psi_\ell\left(\frac{i-1}{n}\right)D_i+\sum_{i=1}^n\Psi_\ell\left(\frac{i-1}{n}\right)\big(h(X_i)-D_i\big).$$

Hence, with similar calculations, we obtain

$$\sum_{\ell\in\mathsf{I}}\alpha_\ell\mathbb{E}\left[\left(\sum_{i=1}^n\Psi_\ell\left(\frac{i-1}{n}\right)\big(h(X_i)+D_i\big)\right)^2\right] \\ \leq 2|h|_{V^\delta}^2 n\sup_{0\leq t\leq 1}\sum_{\ell\in\mathsf{I}}\alpha_\ell\left|\Psi_\ell(t)\right|^2 \\ +6\left(2\sup_{0\leq t\leq 1}\sum_{\ell\in\mathsf{I}}\alpha_\ell\left|\Psi_\ell(t)\right|^2+\sup_{0\leq t\leq 1}\sum_{\ell\in\mathsf{I}}\alpha_\ell\left|\Psi_\ell'(t)\right|^2\right)|h|_{V^\delta}^2.$$

Given (16), these calculations show that $\mathbb{E}(|\check{\sigma}_n^2-\bar{\sigma}_n^2|)=\mathrm{O}(1/\sqrt{n})$. We conclude that

$$\sigma_n^2=\sum_{\ell\in\mathsf{I}}\alpha_j\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\Psi_\ell\left(\frac{i-1}{n}\right)D_i\right)^2+\mathrm{O}_p\left(\frac{1}{\sqrt{n}}\right),$$

which implies that

$$\mathsf{d}_1\big(\sigma_n^2,\chi^2\big)\lesssim\mathsf{d}_1\big(\bar{\sigma}_n^2,\chi^2\big)+\frac{1}{\sqrt{n}}. \tag{26}$$

Therefore, we only need to focus on the term $\mathsf{d}_1(\bar{\sigma}_n^2,\chi^2)$.

On the Euclidean space $\mathbb{R}^{\mathsf{I}}$, we define the norms $\|x\|_\alpha^2=\sum_{i\in\mathsf{I}}\alpha_i x_i^2$, $\|x\|^2=\sum_{i\in\mathsf{I}}x_i^2$ and the inner-products $\langle x,y\rangle_\alpha=\sum_{i\in\mathsf{I}}\alpha_i x_i y_i$, and $\langle x,y\rangle=\sum_{i\in\mathsf{I}}x_i y_i$. For a sequence $(a_1,a_2,\ldots)$, we use the notation $a_{i:k}=(a_i,\ldots,a_k)$ (and $a_{i:k}$ is the empty set if $i>k$). We introduce new random variables $\{Z_{i,j},i\in\mathsf{I},1\leq j\leq n\}$ which are i.i.d. $\mathbf{N}(0,1)$, and set $S_{\ell:k}\overset{\text{def}}{=}(\sum_{j=\ell}^k Z_{1j},\ldots,\sum_{j=\ell}^k Z_{\mathsf{I}j})^\mathsf{T}\in\mathbb{R}^{\mathsf{I}}$, so that

$$\chi^2\overset{\text{dist.}}{=}\sum_{i\in\mathsf{I}}\alpha_i\left(\frac{1}{\sqrt{n}}\sum_{j=1}^n Z_{i,j}\right)^2=\left\|\frac{1}{\sqrt{n}}S_{1:n}\right\|_\alpha^2.$$

For $1 \le \ell \le k \le n$, and omitting the dependence on n, we set $\mathbf{B}_{\ell:k}$ as the $\mathbb{R}^{\mathsf{I}\times(k-\ell+1)}$ matrix

$$\mathbf{B}_{\ell:k}(i,j) = \Psi_i\left(\frac{j}{n}\right), \qquad i \in \mathsf{I}, \ell \le j \le k.$$

By the Mercer's expansion for ϕ, we have

$$\bar{\sigma}_n^2 = \sum_{i\in\mathsf{I}} \alpha_i \left(\frac{1}{\sqrt{n}}\sum_{k=1}^n \Psi_i\left(\frac{k}{n}\right) D_k\right)^2 = \left\| \frac{1}{\sqrt{n}} \mathbf{B}_{1:n} D_{1:n}\right\|_\alpha^2.$$

For $f \in \mathsf{Lip}_1(\mathbb{R})$, we introduce the function $f_\alpha : \mathbb{R}^{|\mathsf{I}|} \to \mathbb{R}$, defined as $f_\alpha(x) = f(\|x\|_\alpha^2)$. As a matter of telescoping the sums, we have

$$\begin{aligned}
&\mathbb{E}\left[f\left(\bar{\sigma}_n^2\right) - f\left(\chi^2\right)\right] \\
&\quad= \mathbb{E}\left[f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:n}D_{1:n}\right) - f_\alpha\left(\frac{1}{\sqrt{n}}S_{1:n}\right)\right] \\
&\quad= \sum_{\ell=1}^n \mathbb{E}\left[f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell}D_{1:\ell} + \frac{1}{\sqrt{n}}S_{\ell+1:n}\right) - f_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{1}{\sqrt{n}}S_{\ell:n}\right)\right] \\
&\quad= \sum_{\ell=1}^n \mathbb{E}\left[f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{1}{\sqrt{n}}\mathbf{B}_{\ell:\ell}D_\ell\right)\right. \\
&\qquad\qquad\left. - f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{1}{\sqrt{n}}S_{\ell:\ell}\right)\right],
\end{aligned}$$

where we define

$$f_{\alpha,n,\ell}(x) \stackrel{\text{def}}{=} \mathbb{E}\left[f_\alpha\left(x + \frac{1}{\sqrt{n}}S_{\ell:n}\right)\right] \quad \text{and set } f_{\alpha,n,n+1}(x) = f_\alpha(x).$$

First, we claim that $f_{\alpha,n,\ell}$ is differentiable everywhere on $\mathbb{R}^{\mathsf{I}}$. To prove this, it suffices to obtain the almost everywhere differentiability of $z \in \mathbb{R}^{\mathsf{I}} \mapsto f_\alpha(x+z)$ for any $x \in \mathbb{R}^{\mathsf{I}}$. By Rademacher's theorem, f as a Lipschitz function is differentiable almost everywhere on $\mathbb{R}$. If E is the set of points where f is not differentiable, we conclude that f_α is differentiable at all points $z \notin \{z \in \mathbb{R}^{\mathsf{I}}: \|x+z\|_\alpha^2 \in E\}$. Now by Ponomarëv [25], Theorem 2, the Lebesgue measure of the set $\{z \in \mathbb{R}^{\mathsf{I}}: \|x+z\|_\alpha^2 \in E\}$ is zero, which proves the claim.

As a result, the function $x \mapsto f_{\alpha,n,\ell}(x)$ is differentiable with derivative

$$\nabla f_{\alpha,n,\ell}(x)\cdot h = 2\mathbb{E}\left[f_\alpha'\left(x + \frac{1}{\sqrt{n}}S_{\ell:n}\right)\left\langle x + \frac{1}{\sqrt{n}}S_{\ell:n}, h\right\rangle_\alpha\right].$$

By writing this expectation wrt the distribution of $x + \frac{1}{\sqrt{n}}S_{\ell:n}$, we get

$$\nabla f_{\alpha,n,\ell}(x)\cdot h = 2\int f_\alpha'(z)\langle z,h\rangle_\alpha \exp\left(-\frac{n}{2(n-\ell+1)}\left(\|x\|^2 - 2\langle x,z\rangle\right)\right)\mu_{n,\ell}(\mathrm{d}z),$$

where $\mu_{n,\ell}$ is the distribution of $\frac{1}{\sqrt{n}}S_{\ell:n}$. This implies that $f_{\alpha,n,\ell}$ is infinitely differentiable with second derivatives given by

$$\begin{aligned}
&\nabla^{(2)} f_{\alpha,n,\ell}(x)\cdot(h_1,h_2)\\
&\quad= -2\left(\frac{n}{n-\ell+1}\right)\\
&\qquad\times\int f'_\alpha(z)\langle z,h_1\rangle_\alpha\langle x-z,h_2\rangle\exp\left(-\frac{n}{2(n-\ell+1)}\left(\|x\|^2-2\langle x,z\rangle\right)\right)\mu_{n,\ell}(\mathrm{d}z)\\
&\quad= 2\mathbb{E}\left[f'_\alpha\left(x+\frac{1}{\sqrt{n}}S_{\ell:n}\right)\left\langle x\sqrt{\frac{n}{n-\ell+1}}+\frac{S_{\ell:n}}{\sqrt{n-\ell+1}},h_1\right\rangle_\alpha\left\langle\frac{S_{\ell:n}}{\sqrt{n-\ell+1}},h_2\right\rangle\right],
\end{aligned}$$

which implies after some easy calculations that for $h\in\mathbb{R}^{\mathsf{l}}$,

$$\left|\nabla^{(2)} f_{\alpha,n,\ell}(x)\cdot(h,h)\right|\lesssim\|h\|^2\left(1+\sqrt{\frac{n}{n-\ell+1}}\|x\|_\alpha\right). \tag{27}$$

Similarly for $h\in\mathbb{R}^{\mathsf{l}}$,

$$\left|\nabla^{(3)} f_{\alpha,n,\ell}(x)\cdot(h,h,h)\right|\lesssim\sqrt{\frac{n}{n-\ell+1}}\|h\|^3\left(1+\sqrt{\frac{n}{n-\ell+1}}\|x\|_\alpha\right). \tag{28}$$

Now, by Taylor expansion we have

$$\begin{aligned}
&f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{1}{\sqrt{n}}\mathbf{B}_{\ell:\ell}D_\ell\right)-f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}+\frac{1}{\sqrt{n}}S_{\ell:\ell}\right)\\
&\quad=\frac{1}{\sqrt{n}}\nabla f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}\right)\cdot(\mathbf{B}_{\ell:\ell}D_\ell-S_{\ell:\ell})\\
&\qquad+\frac{1}{2n}\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}\right)\cdot\left[(\mathbf{B}_{\ell:\ell}D_\ell,\mathbf{B}_{\ell:\ell}D_\ell)-(S_{\ell:\ell},S_{\ell:\ell})\right]+\varrho^{(3)}_{n,\ell},
\end{aligned}$$

where, using (28),

$$\left|\varrho^{(3)}_{n,\ell}\right|\lesssim\sqrt{\frac{n}{n-\ell+1}}n^{-3/2}\left(1+\sqrt{\frac{\ell-1}{n-\ell+1}}\left\|\frac{\mathbf{B}_{1:\ell-1}D_{1:\ell-1}}{\sqrt{\ell-1}}\right\|_\alpha\right)\left(\|\mathbf{B}_{\ell:\ell}D_\ell\|^3_\alpha+\|S_{\ell:\ell}\|^3_\alpha\right).$$

It follows that

$$\sum_{\ell=1}^{n-1}\mathbb{E}\left(\left|\varrho^{(3)}_{n,\ell}\right|\right)\lesssim n^{-1}\sum_{\ell=1}^{n}\frac{1}{\sqrt{\ell}}+n^{-1/2}\sum_{\ell=1}^{n}\frac{1}{\ell}\lesssim n^{-1/2}\log(n). \tag{29}$$

By first conditioning on $\mathcal{F}_{\ell-1}$, we have

$$\mathbb{E}\left[\nabla f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}\right)\cdot(\mathbf{B}_{\ell:\ell}D_\ell-S_{\ell:\ell})\right]=0.$$

Writing $K_{n,\ell} \stackrel{\text{def}}{=} \frac{1}{2}\nabla^{(2)} f_{\alpha,n,\ell}(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1})$, we have

$$\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}\right)\cdot\left[(\mathbf{B}_{\ell:\ell}D_\ell,\mathbf{B}_{\ell:\ell}D_\ell)-(S_{\ell:\ell},S_{\ell:\ell})\right]$$
$$= D_\ell^2\sum_{i,j}\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)K_{n,\ell}(i,j)-\sum_{i,j}\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)K_{n,\ell}(i,j)Z_{i,\ell}Z_{j\ell}.$$

Therefore,

$$\mathbb{E}\left(\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}\right)\cdot\left[(\mathbf{B}_{\ell:\ell}D_\ell,\mathbf{B}_{\ell:\ell}D_\ell)-(S_{\ell:\ell},S_{\ell:\ell})\right]|\mathcal{F}_{\ell-1}\right)$$
$$=\sum_{i,j}\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)K_{n,\ell+1}(i,j)\left[\mathbb{E}\left(D_\ell^2|\mathcal{F}_{\ell-1}\right)-\delta_{ij}\right],$$

where $\delta_{ij}=1$ if $i=j$ and zero otherwise. We claim that the proof will be finished if we show that for all $i,j\in\mathsf{I}$, and $1\le\ell\le n$,

$$\mathbb{E}^{1/2}\left[\left(K_{n,\ell}(i,j)-K_{n,\ell+1}(i,j)\right)^2\right]\lesssim\frac{\sqrt{n}}{n-\ell+1}. \tag{30}$$

To prove this claim, it suffice to use (30) to show that $|n^{-1}\sum_{\ell=1}^n\Psi_i(\frac{\ell}{n})\Psi_j(\frac{\ell}{n})\mathbb{E}(K_{n,\ell+1}(i,j))|\lesssim n^{-1/2}\log(n)$ for $i\neq j$, and $|n^{-1}\sum_{\ell=1}^n\Psi_i(\frac{\ell}{n})\Psi_j(\frac{\ell}{n})\mathbb{E}(K_{n,\ell+1}(i,j)[\mathbb{E}(D_\ell^2|\mathcal{F}_{\ell-1})-1])|\lesssim n^{-1/2}\log(n)$ for all $i,j\in\mathsf{I}$. To show this, write

$$\frac{1}{n}\sum_{\ell=1}^{n-1}\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)\mathbb{E}\left(K_{n,\ell+1}(i,j)\right)$$
$$=\left\{\frac{1}{n}\sum_{\ell=1}^{n-1}\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)\right\}\mathbb{E}\left(K_{n,n}(i,j)\right)$$
$$+\frac{1}{n}\sum_{\ell=1}^{n-1}\left[\frac{1}{n}\sum_{k=1}^{\ell-1}\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)\right]\left[\mathbb{E}\left(K_{n,\ell}(i,j)-K_{n,\ell+1}(i,j)\right)\right].$$

By the convergence of Riemann sums, $|\frac{1}{n}\sum_{\ell=1}^{n-1}\Psi_i(\frac{\ell}{n})\Psi_j(\frac{\ell}{n})|\lesssim n^{-1}$. Combined with (27) and (30), this implies that

$$\left|\frac{1}{n}\sum_{\ell=1}^{n}\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)\mathbb{E}\left(K_{n,\ell+1}(i,j)\right)\right|\le\frac{1}{n}\left(\sqrt{n}+\sqrt{n}\sum_{k=1}^{n}\frac{1}{k}\right)\lesssim\frac{\log(n)}{\sqrt{n}}.$$

For the second term, notice from the definition of D_ℓ at the beginning of the proof that $\mathbb{E}(D_\ell^2|\mathcal{F}_{\ell-1})-1=G(X_{\ell-1})-\pi(G)$, where $G(x)=Pg^2(x)-(Pg(x))^2$. Since $h\in\mathcal{L}_{V^\delta}$ for

$\delta < 1/4$, $G \in \mathcal{L}_{V^{2\delta}}$, and $2\delta < 1/2$. Therefore, by geometric ergodicity, the solution of the Poisson equation for G defined as $U(x) = \sum_{j\geq 0} P^j(G(x) - \pi(G))$ is well-defined, $U \in \mathcal{L}_{V^{2\delta}}$, and we have almost surely

$$U(X_{\ell-1}) - PU(X_{\ell-1}) = \mathbb{E}\big(D_\ell^2|\mathcal{F}_{\ell-1}\big) - 1.$$

Notice that, since $2\delta < 1/2$, for any $p \geq 2$ such that $2p\delta \leq 1$, the geometric ergodicity assumption (G) implies that $\sup_{k\geq 1} \mathbb{E}(|U(X_k)|^p) < \infty$. Now we use the usual martingale approximation trick (see, e.g., Atchadé and Cattaneo [4], Proposition A1) to write

$$\begin{aligned}
&\frac{1}{n}\sum_{\ell=1}^{n-1} \Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)\mathbb{E}\big(K_{n,\ell+1}(i,j)\big[\mathbb{E}\big(D_\ell^2|\mathcal{F}_{\ell-1}\big) - 1\big]\big) \\
&\quad = \frac{1}{n}\Psi_i\left(\frac{1}{n}\right)\Psi_j\left(\frac{1}{n}\right)\mathbb{E}\big(K_{n,2}(i,j)U(X_0)\big) \\
&\qquad - \frac{1}{n}\Psi_i\left(1-\frac{1}{n}\right)\Psi_j\left(1-\frac{1}{n}\right)\mathbb{E}\big(K_{n,n}(i,j)U(X_{n-1})\big) \\
&\qquad + \frac{1}{n}\sum_{\ell=1}^{n-1}\mathbb{E}\bigg[\bigg\{\Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)K_{n,\ell+1}(i,j) \\
&\qquad\qquad - \Psi_i\left(\frac{\ell-1}{n}\right)\Psi_j\left(\frac{\ell-1}{n}\right)K_{n,\ell}(i,j)\bigg\}U(X_{\ell-1})\bigg].
\end{aligned}$$

We now use the fact that $\Psi_i\Psi_j$ is of class C^1 (see Theorem A.1(ii)), (27), and (30) to conclude that

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{\ell=1}^{n-1} \Psi_i\left(\frac{\ell}{n}\right)\Psi_j\left(\frac{\ell}{n}\right)\mathbb{E}\big(K_{n,\ell+1}(i,j)\big[\mathbb{E}\big(D_\ell^2|\mathcal{F}_{\ell-1}\big) - 1\big]\big)\right| \\
&\quad \lesssim \frac{1}{\sqrt{n}} + \frac{1}{n}\sum_{\ell=1}^{n-1}\mathbb{E}^{1/2}\big(\big|K_{n,\ell+1}(i,j) - K_{n,\ell+2}(i,j)\big|^2\big) \lesssim \frac{\log(n)}{\sqrt{n}}.
\end{aligned}$$

This proves the claim. It remains to establish (30). Write $\mathbb{E}_\ell$ to denote the expectation operator wrt $n^{-1/2}\mathbf{S}_{\ell:n}$. We then have for any $h_1, h_2 \in \mathbb{R}^{\mathsf{l}}$,

$$\begin{aligned}
&2K_{n,\ell}\cdot(h_1,h_2) \\
&\quad = \nabla^{(2)} f_{\alpha,n,\ell}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1}\right)\cdot(h_1,h_2) \\
&\quad = 2\left(\frac{n}{n-\ell+1}\right) \\
&\quad\quad \times \mathbb{E}_\ell\left[f'_\alpha\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{S_{\ell:n}}{\sqrt{n}}\right)\left\langle\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{S_{\ell:n}}{\sqrt{n}}, h_1\right\rangle_\alpha\left\langle\frac{S_{\ell:n}}{\sqrt{n}}, h_2\right\rangle\right]
\end{aligned}$$

$$= \left(\frac{n-\ell}{n-\ell+1}\right)\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{S_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2)$$

$$+ \left(\frac{n}{n-\ell+1}\right)\mathrm{O}\left(\frac{1}{\sqrt{n}}\right).$$

Therefore,

$$\begin{aligned}
&2(K_{n,\ell} - K_{n,\ell+1})\cdot(h_1,h_2)\\
&\quad = \nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{S_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2)\\
&\qquad - \nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{\mathbf{B}_\ell D_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2)\\
&\qquad - \frac{1}{n-\ell+1}\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{S_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2) + \left(\frac{n}{n-\ell+1}\right)\mathrm{O}\left(\frac{1}{\sqrt{n}}\right)\\
&\quad = \nabla^{(3)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + t\frac{S_\ell}{\sqrt{n}} + (1-t)\frac{\mathbf{B}_\ell D_\ell}{\sqrt{n}}\right)\cdot\left(h_1,h_2,\frac{S_\ell}{\sqrt{n}} - \frac{\mathbf{B}_\ell D_\ell}{\sqrt{n}}\right)\\
&\qquad - \frac{1}{n-\ell+1}\nabla^{(2)} f_{\alpha,n,\ell+1}\left(\frac{1}{\sqrt{n}}\mathbf{B}_{1:\ell-1}D_{1:\ell-1} + \frac{S_\ell}{\sqrt{n}}\right)\cdot(h_1,h_2) + \left(\frac{n}{n-\ell+1}\right)\mathrm{O}\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}$$

for some $t \in (0,1)$. Using (27) and (28), (30) follows from the above.

Appendix: Mercer's theorem

We recall Mercer's theorem below. Part (i) is the standard Mercer's theorem, and part (ii) is a special case of a result due to T. Kadota (Kadota [16]).

Theorem A.1 (Mercer's theorem). (i) *Let* $k:[0,1]\times[0,1]\to\mathbb{R}$ *be a continuous positive semidefinite kernel. Then there exist nonnegative numbers* $\{\lambda_j, j\ge 0\}$, *and orthonormal functions* $\{\phi_j, j\ge 0\}$, $\phi_j \in L^2([0,1])$, *such that* $\int_0^1 k(x,y)\phi_j(y)\,\mathrm{d}y = \lambda_j\phi_j(x)$ *for all* $x\in[0,1]$, $j\ge 0$, *and*

$$\lim_{n\to\infty}\sup_{x,y\in[0,1]}\left|k(x,y) - \sum_{j=0}^n \lambda_j\phi_j(x)\phi_j(y)\right| = 0.$$

Furthermore, if $\lambda_j \neq 0$, ϕ_j *is continuous.*

(ii) *Let* k *as above. If in addition* k *is of class* C^2 *on* $[0,1]\times[0,1]$, *then for* $\lambda_j\neq 0$, ϕ_j *is of class* C^1 *on* $[0,1]$ *and*

$$\lim_{n\to\infty}\sup_{x,y\in[0,1]}\left|\frac{\partial^2}{\partial x\partial y}k(x,y) - \sum_{j=0}^n \lambda_j\phi_j'(x)\phi_j'(y)\right| = 0.$$

By setting $x = y$, in both expansions, it follows that

$$\sup_{0\le x\le 1}\sum_{j\ge 0}\lambda_j\left|\phi_j(x)\right|^2 \le \sup_{0\le x\le 1} k(x,x) < \infty \tag{A.1}$$

and

$$\sup_{0\le x\le 1}\sum_{j\ge 0}\lambda_j\left|\phi_j'(x)\right|^2 \le \sup_{0\le x\le 1}\left|\frac{\partial^2}{\partial u\partial v}k(u,v)|_{u=x,v=x}\right| < \infty. \tag{A.2}$$

Acknowledgments

I am grateful to Matias Cattaneo and Shane Henderson for very helpful discussions. This work is partly supported by NSF Grants DMS-09-06631, NSF-SES 1229261 and DMS-12-28164.

References

[1] Asmussen, S. and Glynn, P.W. (2007). *Stochastic Simulation: Algorithms and Analysis. Stochastic Modelling and Applied Probability* **57**. New York: Springer. MR2331321

[2] Atchadé, Y.F. (2006). An adaptive version for the Metropolis adjusted Langevin algorithm with a truncated drift. *Methodol. Comput. Appl. Probab.* **8** 235–254. MR2324873

[3] Atchadé, Y.F. (2011). Kernel estimators of asymptotic variance for adaptive Markov chain Monte Carlo. *Ann. Statist.* **39** 990–1011. MR2816345

[4] Atchadé, Y.F. and Cattaneo, M.D. (2014). A martingale decomposition for quadratic forms of Markov chains (with applications). *Stochastic Process. Appl.* **124** 646–677. MR3131309

[5] Chan, K.S. and Geyer, C. (1994). Comment on "Markov chains for exploring posterior distributions." *Ann. Statist.* **22** 1747–1758.

[6] Damerdji, H. (1995). Mean-square consistency of the variance estimator in steady-state simulation output analysis. *Oper. Res.* **43** 282–291. MR1327416

[7] Dedecker, J. and Rio, E. (2008). On mean central limit theorems for stationary sequences. *Ann. Inst. Henri Poincaré Probab. Stat.* **44** 693–726. MR2446294

[8] Flegal, J.M. and Jones, G.L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *Ann. Statist.* **38** 1034–1070. MR2604704

[9] Glynn, P.W. and Iglehart, D.L. (1990). Simulation output analysis using standardized time series. *Math. Oper. Res.* **15** 1–16. MR1038232

[10] Grenander, U. and Rosenblatt, M. (1953). Statistical spectral analysis of time series arising from stationary stochastic processes. *Ann. Math. Statist.* **24** 537–558. MR0058901

[11] Häggström, O. and Rosenthal, J.S. (2007). On variance conditions for Markov chain CLTs. *Electron. Commun. Probab.* **12** 454–464 (electronic). MR2365647

[12] Hannan, E.J. (1970). *Multiple Time Series*. New York: Wiley. MR0279952

[13] Jansson, M. (2004). The error in rejection probability of simple autocorrelation robust tests. *Econometrica* **72** 937–946. MR2051441

[14] Jarner, S.F. and Roberts, G.O. (2002). Polynomial convergence rates of Markov chains. *Ann. Appl. Probab.* **12** 224–247. MR1890063

[15] Jones, G.L., Haran, M., Caffo, B.S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *J. Amer. Statist. Assoc.* **101** 1537–1547. MR2279478
[16] Kadota, T.T. (1967). Term-by-term differentiability of Mercer's expansion. *Proc. Amer. Math. Soc.* **18** 69–72. MR0203397
[17] Kiefer, N.M. and Vogelsang, T.J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory* **21** 1130–1164. MR2200988
[18] Kiefer, N.M., Vogelsang, T.J. and Bunzel, H. (2000). Simple robust testing of regression hypotheses. *Econometrica* **68** 695–714. MR1769382
[19] Kipnis, C. and Varadhan, S.R.S. (1986). Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions. *Comm. Math. Phys.* **104** 1–19. MR0834478
[20] Lindvall, T. (1992). *Lectures on the Coupling Method. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*. New York: Wiley. MR1180522
[21] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability. Communications and Control Engineering Series*. London: Springer. MR1287609
[22] Neave, H.R. (1970). An improved formula for the asymptotic variance of spectrum estimates. *Ann. Math. Statist.* **41** 70–77. MR0251876
[23] Newey, W.K. and West, K.D. (1994). Automatic lag selection in covariance matrix estimation. *Rev. Econ. Stud.* **61** 631–653. MR1299308
[24] Parzen, E. (1957). On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28** 329–348. MR0088833
[25] Ponomarëv, S.P. (1987). Submersions and pre-images of sets of measure zero. *Sibirsk. Mat. Zh.* **28** 199–210. MR0886871
[26] Robert, C.P. and Casella, G. (2004). *Monte Carlo Statistical Methods*, 2nd ed. *Springer Texts in Statistics*. New York: Springer. MR2080278
[27] Roberts, G.O. and Rosenthal, J.S. (2004). General state space Markov chains and MCMC algorithms. *Probab. Surv.* **1** 20–71. MR2095565
[28] Roberts, G.O. and Tweedie, R.L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273
[29] Schruben, L. (1983). Confidence interval estimation using standardized time series. *Oper. Res.* **31** 1090–1108.
[30] Sun, Y., Phillips, P.C.B. and Jin, S. (2008). Optimal bandwidth selection in heteroskedasticity-autocorrelation robust testing. *Econometrica* **76** 175–194. MR2374985
[31] Tierney, L. (1994). Markov chains for exploring posterior distributions. *Ann. Statist.* **22** 1701–1762. MR1329166

Received February 2014 and revised October 2014