Electronic Journal of Statistics
Vol. 18 (2024) 2450–2476
ISSN: 1935-7524
https://doi.org/10.1214/24-EJS2257

Multivariate strong invariance principles in Markov chain Monte Carlo

Arka Banerjee[1] **and Dootika Vats**[2]

[1] *Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur, India 208016, e-mail:* arkabee20@iitk.ac.in

[2] *Department of Mathematics and Statistics, Indian Institute of Technology, Kanpur, India 208016, e-mail:* dootika@iitk.ac.in

Abstract: Strong invariance principles in Markov chain Monte Carlo are crucial to theoretically grounded output analysis. Using the wide-sense regenerative nature of ergodic Markov chains, we obtain explicit bounds on the almost sure convergence rates for partial sums of multivariate ergodic Markov chains. Further, we present results on the existence of strong invariance principles for both polynomially and geometrically ergodic Markov chains without requiring a 1-step minorization condition. Our tight and explicit rates have a direct impact on output analysis, as it allows the verification of important conditions in the strong consistency of variance estimators.

Keywords and phrases: Batch-means estimator, wide-sense regeneration.

Received March 2023.

Contents

1. Introduction

Markov chain Monte Carlo (MCMC) is the workhorse computational algorithm for Bayesian inference. In MCMC, given a target distribution – typically a multi-dimensional Bayesian posterior – an ergodic Markov chain is constructed such that its stationary distribution is the desired target distribution. Specifically, let π be a distribution with support $\mathcal{X}$, equipped with a countably generated σ-field, $\mathcal{B}(\mathcal{X})$. In a typical Bayesian problem, π denotes a posterior distribution and for a function $f : \mathcal{X} \to \mathbb{R}^d$, interest is in estimating $\mathbf{E}_\pi(f(X))$. Having constructed a π-stationary Markov chain, $\{X_t\}_{t\geq 1}$, $\mathbf{E}_\pi(f(X))$ may be estimated using the Monte Carlo average

$$\hat{f}_n := \frac{S_n}{n} := \frac{1}{n}\sum_{t=1}^{n} f(X_t)\,.$$

Naturally, the quality of estimation depends critically on the variability of $\hat{f}_n$ and its limiting distribution. Until recently, focus was on the simpler case of $d = 1$ [16, 24], but the increasing complexity of modern data and modeling strategies warrants the need for ensuring multivariate quality assessment of MCMC averages. [40, 43] highlighted that a multivariate central limit theorem holds for $\hat{f}_n$ if there exists a $d\times d$ positive-definite matrix Σ_f such that as $n \to \infty$

$$\sqrt{n}\left(\hat{f}_n - \mathbf{E}_\pi(f(X)\right) \xrightarrow{d} N(0, \Sigma_f)\,.$$

Estimators of Σ_f are then employed to assess the estimation quality in $\hat{f}_n$ and determine a sufficient Monte Carlo sample size; see [37, 41] for recent reviews. In [19], the authors highlight that strong consistency of variance estimators is necessary for valid sequential stopping rules in simulation. Strong consistency results of estimators of MCMC variances often require the assumption of a strong invariance principle (SIP) on the underlying Markov chain [6, 34, 43].

A popular estimator of Σ_f is the batch-means estimator of [7], which we now describe. Let the Monte Carlo sample size n be such that $n = a_n b_n$ where a_n denotes the number of batches and b_n is the batch-size. For $k = 1, \dots, a_n$, define the mean vector of the k^{th} batch as, $\bar{f}_k = b_n^{-1}\sum_{t=(k-1)b_n+1}^{kb_n} f(X_t)$. Then the batch-means estimator of Σ_f is

$$\hat{\Sigma}_{\text{BM}} := \frac{b_n}{a_n - 1}\sum_{k=1}^{a_n}\left(\bar{f}_k - \hat{f}_n\right)\left(\bar{f}_k - \hat{f}_n\right)^\top\,. \tag{1}$$

[11, 24] studied the strong consistency of the popular batch-means estimator when f is univariate. The path to establishing strong consistency typically assumes the existence of a SIP and thus many works have focused on establishing sufficient conditions under which a univariate SIP holds. Using established results on univariate SIPs for mixing processes, [24] arrived at verifiable conditions for strong consistency for univariate batch-means estimators. Consequently, sufficient conditions on a_n and b_n are available, providing guidelines to users guaranteeing almost sure convergence. On the other hand, verifiable

sufficient conditions are not known for multivariate inference. The best known multivariate SIP results for Markov chains employed in MCMC are due to [43], where the conditions are not verifiable due to unknown constants in the SIP rate of convergence. This makes it challenging for practitioners to employ theoretically grounded simulation techniques in their Bayesian inference models. Utilizing *wide-sense regenerative* properties of ergodic Markov chains, we present conditions for when a multivariate SIP holds under polynomial and geometric ergodicity. Our rates are explicit and the tightest known in this framework.

Let $\|\cdot\|$ denote the Euclidean norm. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the underlying probability space that is assumed to be suitably rich, implying that independent copies of random variables can exist. Let L be a $d \times d$ positive-definite matrix, and $\kappa : \mathbb{N} \to \mathbb{R}^+$ be an increasing function. A multivariate SIP holds if on a suitably rich probability space, one can construct $\{f(X_t)\}_{t\geq 1}$ along with a d-dimensional Wiener process $\{W(t) : t \geq 0\}$ so that as $n \to \infty$,

$$\|S_n - \mathbf{E}(S_n) - LW(n)\| = \mathcal{O}(\kappa(n)) \qquad \text{with probability } 1\,.$$

The rate $\kappa(n)$ often depends on the moments and the amount of correlation in the process. For independent and identically distributed (iid) univariate processes exhibiting a moment generating function, [25, 26] obtained the rate $\kappa(n) = \log(n)$; if $f(X_1)$ has r moments for $r > 3$, they obtain rate $\kappa(n) = n^{1/r}$; we will refer to this result as the KMT result after authors Komlós, Major, and Tusnády. These are the tightest rates possible. For correlated sequences, [35] collate the various known SIP rates for ϕ-mixing, α-mixing, non-stationary, and regenerative processes. The rates obtained are of the form $\kappa(n) = n^{1/2-\lambda}$ where λ is known.

The situation is quite different in the multivariate case. Although for iid vectors [15] extend the KMT result, the proof techniques used in the univariate case for stochastic processes do not, in general, yield explicit rates in the multidimensional one [32]. For general ϕ-mixing processes, [4, 14, 28] obtain a rate of $n^{1/2-\lambda}$ for some unknown $0 < \lambda < 1/2$. The best known result is due to [29] of order $\kappa(n) = n^{1/r}$ for some $2 < r < 4$, and we use their main results to facilitate our result for Markov chains.

For the univariate case, [24] obtain explicit rates for uniformly ergodic Markov chains. For polynomially ergodic univariate continuous-time Markov processes with r moments, [34] obtain rate $\max\{n^{1/4}\log(n), n^{1/r}\log^2(n)\}$. If the Markov chain exhibits a 1-step minorization, it is classically regenerative. Let p denote the moments of the regeneration time and let the partial sum of a regenerative tour exhibit $2+\delta$ finite moments; then [8] obtain rate $n^{\max\{1/(2+\delta),1/(2p),1/4\}}\log(n)$. This work forms the basis of obtaining explicit rates for Markov chains with [24] obtaining an expression for geometrically ergodic Markov chains and [30] obtaining the KMT result for geometrically ergodic bounded Markov chains. For univariate Markov chains exhibiting an l-step minorization, [39] obtain a rate of $n^{1/2-\lambda}$ for some unknown $0 < \lambda < 1/2$. The best known result under this setting is that of [13] and [45], who obtain rate $n^{\max\{1/(2+\delta),1/p,1/4\}}\log(n)$ for univariate and multivariate processes, respectively, under moment assumptions on the regeneration time.

We obtain multivariate SIP rates for MCMC that match the existing univariate rates, while also significantly weakening the conditions. Specifically, the univariate literature requires the Markov chain to exhibit a 1-step minorization, a condition that can be challenging to verify, even when it holds. We require an l-step minorization condition for any $l \geq 1$; this condition is trivially satisfied by all Harris ergodic Markov chains. Our reliance on an l-step minorization also allows us to reformulate the asymptotic covariance matrix, Σ_f, yielding a wide-sense regeneration based estimator.

The remainder of the paper is organized as follows. In Section 2 we define our notations, and present our main results. Our main SIP result requires finite moments on the regeneration time, and we show when these moments results exist for both polynomially and geometrically ergodic Markov chains. In Section 3, we leverage our results to arrive at batch-sizes choices that guarantee strong consistency of $\hat{\Sigma}_{\text{BM}}$. Our SIP result also allows the construction of a new regeneration based estimator of Σ_f. We explore the practicality of this new estimator in Section 4 via two examples, and also demonstrate the performance of the batch-means estimator under conditions where strong consistency is both known and unknown.

2. Definitions and main result

2.1. Markov chains and wide-sense regeneration

Consider a π-Harris ergodic [1] Markov chain with a one-step Markov transition kernel

$$P(x, A) := \Pr(X_{k+1} \in A \mid X_k = x) \quad \text{for } k \geq 1, x \in \mathcal{X}, \text{ and } A \in \mathcal{B}(\mathcal{X})\,.$$

Similarly, for $n \geq 1$, the n-step Markov transition kernel is

$$P^n(x, A) := \Pr(X_{k+n} \in A \mid X_k = x) \quad \text{for } k \geq 1, x \in \mathcal{X}, \text{ and } A \in \mathcal{B}(\mathcal{X})\,.$$

Our SIP results will apply to Markov chains exhibiting certain rates of convergence measured via the total variation distance:

$$\|P^n(x, \cdot) - \pi(\cdot)\|_{TV} := \sup_{A \in \mathcal{B}(\mathcal{X})} |P^n(x, A) - \pi(A)| \leq M(x)G(n)\,, \tag{2}$$

where, since the chain is assumed to be ergodic, $G(n) \to 0$ as $n \to \infty$ and $0 < \mathbf{E}_\pi[M(X)] < \infty$; here we use notation $\mathbf{E}_F$ to denote expectations when $X_1 \sim F$. If $G(n) = n^{-k}$ for some $k \geq 1$, the Markov chain is polynomially ergodic of order k and if $G(n) = t^n$ for some $0 \leq t < 1$, the Markov chain is geometrically ergodic.

[1]We use shorthand "π-Harris ergodic" to mean a Markov chain that is π-irreducible, aperiodic, and Harris recurrent. Please see [31, 36] for definitions. Close to all MCMC algorithms employed in practical problems are π-Harris ergodic, as this implies convergence of sample estimators to population quantities.

Markov chains employed in MCMC are typically π-Harris ergodic and thus exhibit an l-step minorization [2]. An l-step minorization for $l \geq 1$ is said to hold if there exists $h : \mathcal{X} \to [0,1]$ with $0 < \mathbf{E}_\pi[h(X)] < \infty$ and a probability measure $Q(\cdot)$ such that for all $x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})$

$$P^l(x, A) \geq h(x)Q(A)\,. \tag{3}$$

Equation (3) allows the following representation of the l-step transition kernel,

$$P^l(x, A) = h(x)Q(A) + (1 - h(x))R(x, A), \tag{4}$$

where $R(x, \cdot)$ is the residual distribution. Consider the augmented Markov chain, $\{(X_t^*, \delta_t)\}_{t\geq 1}$ where δ's are binary variables such that $X_1^* \sim Q$ and $\delta_1|X_1^* \sim$ Bernoulli$(h(X_1^*))$, and for $i \geq 2$, if $\delta_i = 1$, $X_{i+l}^* \sim Q$, else $X_{i+l}^* \sim R(X_i^*, \cdot)$. By virtue of (4), $X_t^* \overset{d}{=} X_t$. Time index i such that $\delta_i = 1$ is known as a regeneration time. Denote the k^{th} regeneration time as T_k, with $T_0 = 0$. We will use the *wide-sense* regenerative properties of X_t to obtain the SIP result.

Definition 1 ([18]). *A Markov chain is wide-sense regenerative if for each $k \geq 0$, $(X_{T_k+s} : s \geq 0)$ is independent of T_k.*

Let $\tau_k := T_k - T_{k-1}$ be the time to the k^{th} regeneration from the $(k-1)^{\text{th}}$ regeneration. Denote $\mu := \mathbf{E}_Q(\tau_1)$. For $k \geq 1$, define the sum of a tour as $Z_k := \sum_{t=T_{k-1}+1}^{T_k} f(X_t)$, and let $\eta := \mathbf{E}_Q(Z_1)$. When $l = 1$, the Markov chain is classically regenerative and $(Z_k, \tau_k)_{k\geq 1}$ are iid. For the one-dimensional case, this feature is exploited by [8] to arrive at an SIP using classical KMT results. These results have been adapted to MCMC by [24]. However, for many MCMC samplers, $l = 1$ is a limiting assumption in that it can be challenging to verify. Since all Harris ergodic Markov chains satisfy an l-step minorization for some l [see 2, for e.g.], the assumption of an l-step minorization is no longer limiting. However, for general l, [18] explain that $(Z_k, \tau_k)_{k\geq 1}$ is a one-dependent stationary process [2], and thus the classical KMT results can no longer be used to establish an SIP.

2.2. Main result

We now present our main results establishing a multivariate SIP. The following covariance matrix will appear in the limiting covariance expression of the SIP:

$$\begin{aligned}\Sigma_Z :=& \mathbf{Var}_Q\left(Z_1 - \frac{\tau_1}{\mu}\eta\right) + \mathbf{Cov}_Q\left(Z_1 - \frac{\tau_1}{\mu}\eta, Z_2 - \frac{\tau_2}{\mu}\eta\right) \\ &+ \mathbf{Cov}_Q\left(Z_2 - \frac{\tau_2}{\mu}\eta, Z_1 - \frac{\tau_1}{\mu}\eta\right).\end{aligned} \tag{5}$$

The result below establishes an SIP rate under general conditions on a π-Harris ergodic Markov chain.

[2] A process $\{Y_t\}_{t\geq 1}$ is 1-dependent if for all k, the joint variables $(Y_n)_{n\leq k}$ are independent of the joint variables $(Y_n)_{n\geq k+2}$.

Theorem 1. *Let $\{X_t\}_{t\geq 1}$ be a π-Harris ergodic Markov chain and thus (3) holds. Suppose*

(a) $\mathbf{E}_Q(\tau_1^p) < \infty$ *for some $p > 1$ and,*
(b) for some $\delta > 0$ and some $\delta^* > 0$

$$\mathbf{E}_\pi \left[\left(\sum_{t=1}^{\tau_1} \left\| f(X_t) - \frac{\eta}{\mu} \right\| \right)^{2+\delta+\delta^*} \right] < \infty \,. \tag{6}$$

Then, on a suitably rich probability space, one can construct $\{f(X_t)\}_{t\geq 1}$ together with a d-dimensional standard Wiener process $\{W(t) : t \geq 0\}$ such that for $\beta = \max\{1/(2+\delta), 1/(2p), 1/4\}$, as $n \to \infty$, with probability 1

$$\left\| \sum_{t=1}^{n} f(X_t) - n\mathbf{E}_\pi[f(X)] - \frac{\Sigma_Z^{1/2}}{\sqrt{\mu}} W(n) \right\| = \mathcal{O}(n^\beta \log(n)). \tag{7}$$

Proof. See Appendix B. □

Remark 1. Theorem 1 is the first result we know that matches the SIP rate of [8] for general multivariate functionals. Additionally, similar to [13], we remove the assumption of a 1-step minorization. Under the same assumptions, [13, 45] obtain the rate with $\beta = \max\{1/(2+\delta), 1/p, 1/4\}$, hence our rates are tighter.

For practical application to MCMC, it is important to assess when and for what values of p is $\mathbf{E}_Q(\tau_1^p) < \infty$. For a 1-step minorization, [22] show that τ_1 has a moment-generating function when $\{X_t\}_{t\geq 1}$ is geometrically ergodic. The next two lemmas are critical to obtaining moment conditions over τ for polynomially and geometrically ergodic Markov chains. Additionally, Lemma 3 and Lemma 4 aid in proving the moment existence of regenerative sums. Proofs of the following lemmas are in Appendix C.

Lemma 1. Let $\{X_t\}_{t\geq 1}$ be a π-stationary polynomially ergodic Markov chain of order $\xi > (2+\delta)(1+(2+\delta)/\delta^*)$ for some $\delta > 0$ and $\delta^* > 0$ and $\mathbf{E}_\pi(M) < \infty$. Then for all $p \in (0, \xi)$, $\mathbf{E}_Q[\tau_1^p] < \infty$.

Lemma 2. Let $\{X_t\}_{t\geq 1}$ be a π-stationary geometrically ergodic Markov chain. Then $\mathbf{E}_Q[\tau_1^p] < \infty$ for any $p > 1$.

Lemma 3. Let $\{X_t\}_{t\geq 1}$ be a π-Harris ergodic Markov chain so that (3) holds. Let $\mathbf{E}_\pi\left(\|f(X)\|^{r+\delta^*}\right) < \infty$ for some $r > 1$ and $\delta^* > 0$ and $\mathbf{E}_\pi(\tau_1^\phi) < \infty$ for $\phi > r(r+\delta^*)/\delta^*$. Then,

$$\mathbf{E}_\pi \left[\left(\sum_{i=1}^{\tau_1} \|f(X_i)\| \right)^r \right] < \infty \,.$$

Lemma 4. Let $\{X_t\}_{t\geq 1}$ be a π-stationary geometrically ergodic Markov chain so that (3) holds. Further, let $\mathbf{E}_\pi\left[(\|f(X)\|)^{r+\delta^*}\right] < \infty$ for some $r > 1$ and $\delta^* > 0$. Then

$$\mathbf{E}_\pi \left[\left(\sum_{i=1}^{\tau_1} \|f(X_i)\| \right)^r \right] < \infty.$$

The above lemmas allow the following result on the existence of an SIP for MCMC under verifiable conditions.

Theorem 2. *Let $\{X_t\}_{t\geq 1}$ be a π-Harris ergodic Markov chain and thus (3) holds. Suppose the chain is either*

(a) polynomially ergodic of order $\xi > (2+\delta)(1+(2+\delta)/\delta^*)$ for some $\delta > 0$ and some $\delta^* > 0$, $\mathbf{E}_\pi\left(\|f(X)\|^{2+\delta+\delta^*}\right) < \infty$, and $\mathbf{E}_\pi(M) < \infty$; or,

(b) geometrically ergodic and $\mathbf{E}_\pi\left(\|f(X)\|^{2+\delta+\delta^*}\right) < \infty$ for some $\delta > 0$ and some $\delta^* > 0$,

then (7) holds with $\beta = \max\{1/(2+\delta), 1/4\}$.

Proof. See Appendix C. □

Theorem 2 marks a three-fold improvement over the existing results of [24] in MCMC; (i) the assumption of a 1-step minorization is completely removed, (ii) an explicit SIP rate for polynomially ergodic Markov chains, and (iii) the critical extension to multivariate functionals.

Remark 2. The assumptions in Theorem 2 are sufficient and are not known to be necessary. Further, when enough moments exists, we obtain a rate of $\kappa(n) = n^{1/4}$. This is now the best known rate for general Markov chains in the multivariate setup, but this is not sharp. The lack of sharpness, [32] describe, is a consequence of the proof techniques available for the multivariate case, and an improvement on these rates remains an open problem in the area.

Remark 3. Theorem 2 presents reasonably weak and verifiable conditions for the existence of a multivariate SIP. To see this, we first note that since $\beta < 1/2$, the SIP implies a weak invariance principle, and thus, a CLT. [23] summarize that for geometrically ergodic Markov chains, the conditions required for a CLT are the same as Theorem 2b. However, for polynomially ergodic chains, an order of $\xi > (1 + 2/(\delta + \delta^*))$ is sufficient for a CLT, whereas our SIP results require slightly larger order of $\xi > (2+\delta)(1+(2+\delta)/\delta^*)$. Finally, note that since the SIP result implies a CLT for $\hat{f}_n$, $\Sigma_f = \Sigma_Z/\mu$.

Remark 4. For any π-Harris ergodic Markov chain, an l-step minorization always holds for some $l \geq 1$, and the chain is wide-sense regenerative. Utilizing the one-dependent nature of the process $\{(Z_t, \tau_t)\}_{t\geq 1}$, estimators of $\Sigma_f = \Sigma_Z/\mu$ can be obtained. This, however, requires detecting wide-sense regenerations from a simulated Markov chain, which can be fairly challenging. We expand on these points in Section 3.2.

3. Application to MCMC variance estimation

We discuss the direct impact of our main results in MCMC variance estimation. First, the explicit rates obtained in Theorem 2 now allows theoretically valid tuning of the popular multivariate batch-means estimator. Further, the form of the limiting variance in Theorem 2 allows for a wide-sense regeneration based estimator of the Monte Carlo variance

3.1. Batch-means estimator

Having generated the process $\{X_t\}_{t\geq 1}$ through an MCMC algorithm, the samples are employed to estimate $\mathbf{E}_\pi[f(X)]$ via the Monte Carlo average since,

$$\hat{f}_n = \frac{1}{n}\sum_{t=1}^{n} f(X_t) \overset{a.s.}{\to} \mathbf{E}_\pi[f(X)] \qquad \text{as } n \to \infty\,.$$

When a Markov chain CLT holds for $\hat{f}_n$, there exists a positive-definite matrix Σ_f such that as $n \to \infty$,

$$\sqrt{n}\left(\hat{f}_n - \mathbf{E}_\pi[f(X)]\right) \overset{\mathrm{d}}{\to} \mathrm{N}(0, \Sigma_f)\,, \tag{8}$$

where

$$\Sigma_f = \mathbf{Var}_\pi[f(X)] + \sum_{s=1}^{\infty}\left[\mathbf{Cov}_\pi(f(X_1), f(X_{1+s})) + \mathbf{Cov}_\pi(f(X_1), f(X_{1+s}))^\top\right]\,.$$

Estimation of Σ_f is widely discussed, both in the univariate case [3, 5, 11, 17, 12, 24, 16], and the multivariate case [27, 10, 40, 43, 44, 42]. Estimators of Σ_f are employed in deciding when to stop the simulation. Thus, the MCMC simulation stops at a random time and [19] show that validity of the subsequent estimators require strong consistency of estimators of Σ_f. Much effort has thus gone into ensuring that estimators of Σ_f are strongly consistent. One particular estimator that stands out due its computational efficiency and theoretical underpinnings, is the batch-means estimator defined in (1). An existence of a multivariate SIP is often assumed for the strong consistency of the batch-means estimators. [43] showed that if the Markov chain is polynomially ergodic, then for some (unknown) $0 < \lambda < 1/2$ a multivariate SIP holds with rate $n^{1/2-\lambda}$. However, strong consistency of $\hat{\Sigma}_{\mathrm{BM}}$ also depends on choosing appropriate rates for the batch-size, b_n, for which it is necessary that λ be known. Theorem 2 overcomes this issue considerably as we will now elucidate.

Assumption 1. *The batch size b_n is such that*

(a) $b_n \to \infty$ and $n/b_n \to \infty$ as $n \to \infty$ where, b_n and n/b_n are monotonically increasing,
(b) there exists a constant $c \geq 1$ such that $\sum_n (b_n n^{-1})^c < \infty$.

Often $b_n = \lfloor n^\nu \rfloor$ for some $\nu > 0$ so that Assumption 1 is trivially satisfied. Common choices in the literature are $b_n = \lfloor n^{1/3} \rfloor$ and $b_n = \lfloor n^{1/2} \rfloor$. The following theorem from [44] presents the conditions for strong consistency of the batch-means estimator.

Theorem 3 ([44])**.** *Suppose P and f exhibit a multivariate SIP with rate $\kappa(n)$. If b_n satisfies Assumption 1 and $b_n^{-1}\log(n)\kappa^2(n) \to 0$ as $n \to \infty$, then $\hat{\Sigma}_{BM} \to \Sigma$ with probability 1 as $n \to \infty$.*

The above theorem highlights the dependence of the the batch size, b_n, on the SIP rate; typically, slower mixing Markov chains will have a slower rate in $\kappa(n)$ requiring larger b_n and vice-versa for fast mixing chains. The following corollary is a consequence of Theorems 2 and 3.

Corollary 1. Let $\{X_t\}_{t\geq 1}$ be a π-Harris ergodic Markov chain and thus (3) holds. Suppose the chain is either

(a) polynomially ergodic of order $\xi > (2+\delta)(1+(2+\delta)/\delta^*)$ for some $\delta > 0$ and some $\delta^* > 0$, $\mathbf{E}_\pi\left(\|f(X)\|^{2+\delta+\delta^*}\right) < \infty$, and $\mathbf{E}_\pi(M) < \infty$; or,
(b) geometrically ergodic and $\mathbf{E}_\pi\left(\|f(X)\|^{2+\delta+\delta^*}\right) < \infty$ for some $\delta > 0$ and some $\delta^* > 0$,

then the batch means estimator is strongly consistent for $b_n = \lfloor n^\nu \rfloor$ for $\nu > \max\{2/(2+\delta), 1/2\}$.

Proof. See Section D. □

For strong consistency in the univariate case, [24] assumed the same conditions on the batch size; for geometrically ergodic Markov chains under a 1-step minorization, [24] showed that a univariate SIP holds with $\kappa(n) = n^\beta \log n$, where β is the same as Theorem 2. This implies that the batch size should be chosen such that $\nu > \max\{2/(2+\delta), 1/2\}$. For geometrically ergodic Markov chains, this is the best known batch size condition.

For the multivariate case, the only known MCMC SIP result was that of [43] with $\kappa(n) = n^{1/2-\lambda}$ for some unknown $0 < \lambda < 1/2$. This implies a choice of $\nu > 1-2\lambda$ but since λ is unknown, the condition, $b_n^{-1}\log(n)\kappa^2(n)$ cannot be verified. Now, by Corollary 1, we match the univariate condition of $\nu > \max\{2/(2+\delta), 1/2\}$. We note, as did [11, 12], that this condition excludes common choices of b_n, like $b_n = \lfloor n^{1/3} \rfloor$. In Section 3, we compare the performance of $b_n = \lfloor n^{1/2+.0.0001} \rfloor$ – a choice that satisfies the conditions of Corollary 1 – versus $b_n = \lfloor n^{1/3} \rfloor$.

3.2. Regenerative estimator

Given that $(Z_k, \tau_k)_{k\geq 1}$ is a one-dependent stationary process and $\Sigma_f = \Sigma_Z/\mu$, if the regeneration times could be identified, we can construct estimators of Σ_f using (5). Although regenerations, particularly wide-sense regenerations, are notoriously difficult to identify, it is natural to consider a wide-sense regenerative estimator of Σ_f, even if it is for a theoretical exposition. In Section 4, we identify examples and situations in which wide-sense regenerations can be identified. For 1-step minorization, [22, 40] present regenerative estimators, and for univariate wide-sense regenerative processes, [21] provide an estimator of $\mathbf{E}_\pi[f(X)]$. Here, we present a regenerative estimator of Σ_f.

For an observed Markov chain of length n, denote the number of regenerations by R. Using the fact that $\Sigma_f = \Sigma_Z/\mu$, we obtain a wide-sense regenerative plug-in estimator of Σ_f. Define $\bar{Z} := R^{-1}\sum_{i=1}^{R}\left(Z_i - \tau_i\hat{f}\right)$. We can estimate μ and

Σ_Z with,

$$\hat{\mu} := \frac{1}{R}\sum_{i=1}^{R}\tau_i\,, \qquad \text{and} \tag{9}$$

$$\begin{aligned}\hat{\Sigma}_Z &:= \frac{1}{R}\sum_{i=1}^{R}(Z_i - \tau_i\hat{f})(Z_i - \tau_i\hat{f})^\top \\ &\quad + \frac{1}{R}\sum_{i=1}^{R-1}(Z_i - \tau_i\hat{f})(Z_{i+1} - \tau_{i+1}\hat{f})^\top \\ &\quad + \frac{1}{R}\sum_{i=1}^{R-1}(Z_{i+1} - \tau_{i+1}\hat{f})(Z_i - \tau_i\hat{f})^\top\,. \end{aligned} \tag{10}$$

Using (9) and (10), the estimator for Σ_f is $\hat{\Sigma}_f = \hat{\Sigma}_Z/\hat{\mu}$. Since as $n \to \infty$, $R \to \infty$, using strong law of large numbers for one-dependent processes [see for e.g. 39], $\hat{\Sigma}_f$ is strongly consistent for Σ_f.

Employing this estimator in practice has two significant challenges: (i) detecting wide-sense regenerations can be difficult since one requires knowledge of l and the minorization kernel Q, and (ii) when minorizations can be established, the bounds in the minorization constant are fairly weak, yielding prohibitively large regeneration times. Even still, there are certain situations where employing these estimators is possible, and we identify two such examples in the sequel.

4. Illustrative examples

With the help of two examples, we demonstrate the utility of our discussions in Section 3. Our primary objective in these examples is to demonstrate that our results allow us to choose batch-sizes that guarantee strong consistency of the multivariate batch-means estimators. In the first example, we also compare the batch-means and regenerative estimators. Reproducible `R` code for the examples is available at https://github.com/Arkagit/Multivariate-SIP-for-MCMC.

4.1. Bivariate normal Gibbs sampler

For known constants μ_1, μ_2, σ_1, σ_2 and ρ such that $\sigma_1\sigma_2 > \rho^2$ we consider a bivariate normal target distribution

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho \\ \rho & \sigma_2^2 \end{pmatrix}\right).$$

In this toy problem, we will employ a deterministic scan Gibbs sampler that alternates updates between the X and Y components using the conditional distributions:

$$(X \mid Y = y) \sim N\left(\mu_1 + \frac{\rho}{\sigma_2^2}(y - \mu_2), \sigma_1^2 - \frac{\rho^2}{\sigma_2^2}\right)$$

$$(Y \mid X = x) \sim N\left(\mu_2 + \frac{\rho}{\sigma_1^2}(x - \mu_1), \sigma_2^2 - \frac{\rho^2}{\sigma_1^2}\right).$$

The consequent 1-step Markov transition density for any i^{th} step of the deterministic scan Gibbs sampler is

$$k_{\text{DS}}\left(x_{i+1}, y_{i+1} \mid x_i, y_i\right) = \pi(y_{i+1} \mid x_{i+1})\, \pi(x_{i+1} \mid y_i).$$

For this deterministic scan Gibbs sampler, the asymptotic covariance matrix in (8) is known [see 20, Theorem 6], and is given by

$$\Sigma_{\text{BNG}} = \frac{1}{\sigma_1^2\sigma_2^2 - \rho^2}\begin{pmatrix} \sigma_1^2\left(\sigma_1^2\sigma_2^2 + \rho^2\right) & 2\sigma_1^2\sigma_2^2\rho \\ 2\sigma_1^2\sigma_2^2\rho & \sigma_2^2\left(\sigma_1^2\sigma_2^2 + \rho^2\right) \end{pmatrix}.$$

Let $y^* \in \mathbb{R}$ be a "distinguished point" and let $D \subseteq \mathbb{R}$ be a "small set" as described by [33]; by using the distinguished point technique of [33], the following 1-step minorization can be established

$$\begin{aligned} k_{\text{DS}}\left(x_{i+1}, y_{i+1} \mid x_i, y_i\right) &\geq \inf_{x \in D}\left\{\frac{\pi(x \mid y_i)}{\pi(x \mid \mathbf{y}^*)}\right\}\pi(y_{i+1} \mid x_{i+1})\ \pi(x_{i+1} \mid \mathbf{y}^*)\ I_D(x_{i+1}) \\ &=: s(y_i)\ Q(x_{i+1}, y_{i+1}), \end{aligned}$$

where the forms of $s(y_i)$ and $Q(x_{i+1}, y_{i+1})$ are known explicitly and presented in Appendix E.1. Consequently, we can show that a 2-step minorization holds as well, since

$$\begin{aligned} & k^2_{\text{DS}}\left(x_{i+2}, y_{i+2} \mid x_i, y_i\right) \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} k_{\text{DS}}\left(x_{i+2}, y_{i+2} \mid x_{i+1}, y_{i+1}\right)\ k_{DS}\left(x_{i+1}, y_{i+1} \mid x_i, y_i\right)\ dy_{i+1}\ dx_{i+1} \\ &\geq s(y_i)\ Q(x_{i+2}, y_{i+2})\ \int_{\mathbb{R}}\int_{\mathbb{R}} s(y_{i+1})Q(x_{i+1}, y_{i+1})\ dy_{i+1}\ dx_{i+1} \\ &= s^*(y_i)\ Q(x_{i+2}, y_{i+2}). \end{aligned}$$

Here, $s^*(y) := s(y)\mathbf{E}_Q(s(y))$ and can be estimated using Monte Carlo.

We fix our target distribution to have $\sigma_1^2 = 1$, $\sigma_2^2 = 5$, $\rho = 1.5$, $\mu_1 = \mu_2 = 0$. To identify both 1-step and 2-step regenerations using the distinguished point approach, we set $y^* = 0$ and $D = \{(x, y) : x \in [-1, 1], y \in [-1, 1]\}$. For increasing Monte Carlo sample size, n, we compare the batch-means estimator for $b_n = \lfloor n^{1/2+0.0001} \rfloor$ and $b_n = \lfloor n^{1/3} \rfloor$ along with regenerative estimators for 1-step and 2-step regenerations. We assess the performance of the estimators by measuring the Frobenius norm distance $\|\hat{\Sigma} - \Sigma_{\text{BNG}}\|_F$ for all estimators, and by studying running estimates of the multivariate effective sample size. The multivariate effective sample size for estimating $\mathbf{E}_\pi(f)$ was defined by [44] as

$$\text{ESS} = n\left(\frac{\det\left(\text{Var}_\pi(f)\right)}{\det\left(\Sigma_f\right)}\right)^{1/p}.$$

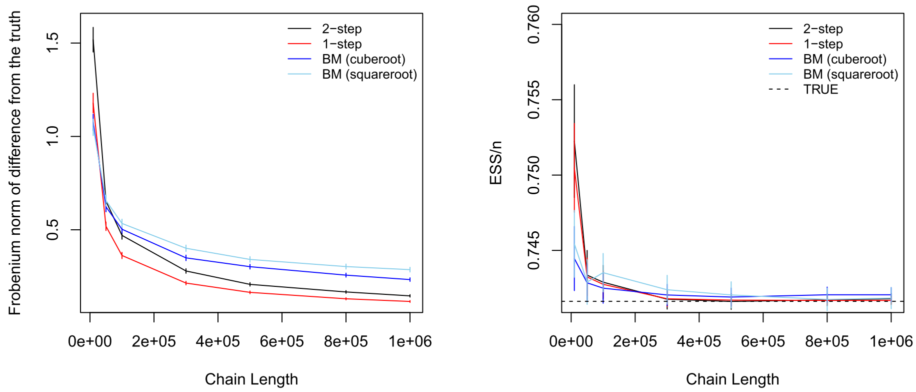

FIG 1. *Left: For varying sample sizes, evolution of the Frobenius norm distance of estimators of Σ_{BVG} from the truth. Right: Running estimates of estimated ESS /n using different estimators.*

In Figure 1, we present the Frobenius norm distance $\|\hat{\Sigma} - \Sigma_{\text{BNG}}\|_F$ and estimated ESS $/n$, averaged over 1000 replications. Since the target is a bivariate normal, moment conditions of Theorem 3 are easily satisfied and batch-size $b_n = \lfloor n^{1/2+0.0001} \rfloor$ yields strong consistency of the batch-means estimator. On the other hand, strong consistency under the popular $b_n = \lfloor n^{1/3} \rfloor$ is unknown. Figure 1 shows that these two choices yield similar performing estimators, with $b_n = \lfloor n^{1/2+0.0001} \rfloor$ ensuring theoretical validity. The 1-step and 2-step regenerative estimators perform better than the batch-means estimators, but of course did require a lot more effort to set up.

4.2. Bayesian probit regression model

For a Bayesian probit regression model, [38] studied 1-step regenerations for some deterministic scan Gibbs samplers. We adapt their setup to establish both 1-step and 2-step minorizations. We found that no regenerations were identified for even a chain of length 10^6, and so in this example we only implement the batch-means estimators.

For $j = 1, 2, \ldots, m$, [38] consider the model $Y_j | \beta \overset{\text{ind}}{\sim} \text{Bernoulli}(\Phi(x_j^T \beta))$ where $x_j \in \mathbb{R}^p$ are given and $\beta \in \mathbb{R}^p$ is the vector of coefficients. Set $X = (x_1^\top, x_2^\top, \ldots, x_m^\top)^\top$ and consider a flat prior for β yielding the posterior density

$$\pi(\beta \mid \mathbf{y}) = \prod_{j=1}^{m} \{\Phi(x_j^\top \beta)\}^{y_j} \{1 - \Phi(x_j^\top \beta)\}^{1-y_j}.$$

Although the complete analytical form of $\pi(\beta \,|\, \mathbf{y})$ is unavailable, [1] provide a deterministic Gibbs sampler using an auxiliary variable, denoted by $\mathbf{z} = (z^1, z^2, \ldots, z^m)^\top$. We refer to this sampler as the AC sampler and implement a random scan version of the AC sampler:

- with probability p, update β: draw $\beta'|\mathbf{z} \sim \mathrm{N}_p\left((X^\top X)^{-1}X^\top \mathbf{z}, (X^\top X)^{-1}\right)$.
- with probability $(1-p)$, update $\mathbf{z}$: for all $j = 1, 2, \ldots, m$, draw $(z^j)'|\beta \overset{\text{iid}}{\sim}$ Truncated Normal$(x_j^\top \beta, 1, y_j)$, where if $y_j = 0$, $(-\infty, 0]$ is the truncation range and if $y_j = 1$, $(0, \infty)$ is the truncation range.

The 1-step transition density for the random scan AC sampler is

$$
\begin{aligned}
&k_{\mathrm{RS}}(\beta_{i+1}, \mathbf{z}_{i+1} \mid \beta_i, \mathbf{z}_i)\\
&= p\pi\left(\beta_{i+1} \mid \mathbf{z}_i\right)\delta_{\mathbf{z}_i}(\mathbf{z}_{i+1}) + (1-p)\pi\left(\mathbf{z}_{i+1} \mid \beta_{i+1}\right)\delta_{\beta_i}(\beta_{i+1}).
\end{aligned}
$$

Consequently, the 2-step random scan transition density is

$$
\begin{aligned}
k^2_{\mathrm{RS}}\left(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_i, \mathbf{z}_i\right) = {} & p^2\pi\left(\beta_{i+2} \mid \mathbf{z}_{i+2}\right)\delta_{\mathbf{z}_i}(\mathbf{z}_{i+2})\\
&+ p(1-p)\pi\left(\beta_{i+2} \mid \mathbf{z}_{i+2}\right)\ \pi\left(\mathbf{z}_{i+2} \mid \beta_i\right)\\
&+ p(1-p)\pi\left(\mathbf{z}_{i+2} \mid \beta_{i+2}\right)\pi\left(\beta_{i+2} \mid \mathbf{z}_i\right)\\
&+ (1-p)^2\pi\left(\mathbf{z}_{i+2} \mid \beta_i\right)\delta_{\beta_i}(\beta_{i+2}). \qquad (11)
\end{aligned}
$$

Since a 1-step minorization of the deterministic scan AC sampler is available, a 2-step minorization of the random scan Gibbs sampler is easily obtained:

$$
\begin{aligned}
k^2_{\mathrm{RS}}\left(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_i, \mathbf{z}_i\right) &\geq p(1-p)s(\mathbf{z}_i)\ Q(\beta_{i+2}, \mathbf{z}_{i+2})\\
&= s'(\mathbf{z}_i)\ Q(\beta_{i+2}, \mathbf{z}_{i+2}), \qquad (12)
\end{aligned}
$$

where the specific form of s' and Q is given in Appendix E.2.

We use the above Bayesian probit regression model to analyze the popular Titanic dataset. The response is an indicator on whether a passenger survived the tragedy or not. The covariates are the sex of the passenger, the age, the number of siblings and spouses accompanying the passenger, number of children and parents accompanying the passenger, and the fare of their ticket. We fit a model with intercept, yielding a $p = 6$ dimensional posterior, and aim to estimate the posterior mean.

We run the random scan AC sampler for varying lengths, and compute the batch-means estimator with b_n being both $\lfloor n^{1/2+0.0001} \rfloor$ and $\lfloor n^{1/3} \rfloor$. Unfortunately, we did not observe any 2-step regenerations, and thus are unable to compare the regenerative estimators in this example. Further, since the true Σ_f is unknown, we only present the running estimates of the effective sample size. Averaged over 500 replications, we plot the estimated ESS $/n$ for both batch-means estimators in Figure 2. Here it is evident that in addition to guaranteeing theoretical convergence and consistency, batch-size $b_n = \lfloor n^{1/2+0.0001} \rfloor$ performs significantly better than the cuberoot batch-size, reaching to a point of stability faster than the cuberoot batch-size.

Appendix A: Some preliminary results

Lemma 5. Let $\{X_t\}_{t\geq 1}$ be a π-Harris ergodic Markov chain. Recall f, η, and μ from Section 2.1. Then,

$$
\mathbf{E}_\pi[f(X)] = \frac{\eta}{\mu}\,.
$$

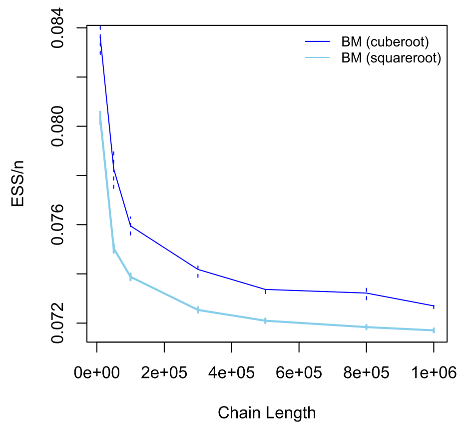

FIG 2. *Monte Carlo sample size (log scale) versus estimated ESS /n.*

Proof. By [18], $\{(Z_k, \tau_k) : k \geq 1\}$ form a stationary 1-dependent process. By a strong law of large numbers for 1-dependent processes,

$$\frac{1}{R}\sum_{i=1}^{R} Z_i \stackrel{a.s.}{\to} \mathbf{E}_Q(Z_1) \text{ as } R \to \infty \quad \text{and,} \tag{13}$$

$$\frac{1}{R}\sum_{i=1}^{R} \tau_i \stackrel{a.s.}{\to} \mathbf{E}_Q(\tau_1) \text{ as } R \to \infty. \tag{14}$$

By a strong law for ergodic Markov chains and from (14),

$$\frac{1}{T_R}\sum_{t=1}^{T_R} f(X_t) = \frac{1/R\sum_{i=1}^{R} Z_i}{1/R\sum_{i=1}^{R}\tau_i} \stackrel{a.s.}{\to} \mathbf{E}_\pi[f(X)] \text{ as } R \to \infty$$

$$\Rightarrow \left(1/R\sum_{i=1}^{R}\tau_i\right)\frac{1/R\sum_{i=1}^{R} Z_i}{1/R\sum_{i=1}^{R}\tau_i} \stackrel{a.s.}{\to} \mathbf{E}_Q(\tau_1)\mathbf{E}_\pi[f(X)] \text{ as } R \to \infty$$

$$\Rightarrow \left(1/R\sum_{i=1}^{R} Z_i\right) \stackrel{a.s.}{\to} \mathbf{E}_Q(\tau_1)\mathbf{E}_\pi[f(X)] \text{ as } R \to \infty. \tag{15}$$

Thus, by (13) and (15),

$$\mathbf{E}_Q(Z_1) = \mathbf{E}_Q(\tau_1)\mathbf{E}_\pi[f(X)] \Rightarrow \mathbf{E}_\pi[f(X)] = \frac{\eta}{\mu}. \tag{16}$$

□

The following lemma will be employed for the proof of Theorem 2 and is an extension of [22, Lemma 1] to the multivariate and l-step minorization case.

Lemma 6. Let $\{X_t\}_{t\geq 1}$ be a π-Harris ergodic Markov chain so that (3) holds. Then, for any measurable function $\Psi : \mathcal{X}^\infty \to \mathbb{R}^d$

$$\mathbf{E}_\pi \|\Psi(X_1, X_2, X_3, \ldots)\| \geq \mathbf{E}_\pi[h(X)]\mathbf{E}_Q\|\Psi(X_1, X_2, X_3, \ldots)\| \,. \tag{17}$$

Proof. For all $A \in \mathcal{B}(\mathcal{X})$ and $x \in \mathcal{X}$,

$$\pi(A) = \left(\pi P^l\right)(A) = \int_{\mathcal{X}} \pi(dx)P^l(x, A) \geq Q(A)\int_{\mathcal{X}} h(x)\pi(x) = Q(A)\mathbf{E}_\pi[h(X)]. \tag{18}$$

By taking conditional expectation over X_1

$$\mathbf{E}_\pi \|\Psi(X_1, X_2, X_3, \ldots)\| = \mathbf{E}_\pi \left(\mathbf{E}\{\|\Psi(X_1, X_2, X_3, \ldots)\| \mid X_1\}\right). \tag{19}$$

Since $\mathbf{E}\{\|\Psi(X_1, X_2, X_3, \ldots)\| \, |X_1\}$ is a positive function of X_1, by (18) and (19)

$$\begin{aligned}
\mathbf{E}_\pi \|\Psi(X_1, X_2, X_3, \ldots)\| &= \int_{\mathcal{X}} \mathbf{E}\{\|\Psi(X_1, X_2, X_3, \ldots)\| \mid X_1 = x\}\pi(dx) \\
&\geq \mathbf{E}_\pi[h(X)] \int_{\mathcal{X}} \mathbf{E}\left(\|\Psi(X_1, \ldots)\| \mid X_1 = x\right) Q(dx) \\
&= \mathbf{E}_\pi[h(X)]\mathbf{E}_Q\|\Psi(X_1, \ldots)\| \\
\Rightarrow \mathbf{E}_\pi\|\Psi(X_1, \ldots)\| &\geq \mathbf{E}_\pi[h(X)]\mathbf{E}_Q\|\Psi(X_1, \ldots)\|.
\end{aligned}$$

□

Appendix B: Proof of Theorem 1

We will be using [29, Theorem 2.1] for our SIP result. The original statement of the theorem is presented in generality, and here we present a version is that amenable to our setup.

Theorem 4. *[see 29, Theorem 2.1 for more details] Let $\{\epsilon_n : n \geq 0\}$ be iid random elements and let $\{A_k\}_{k\geq 1}$ be such that $A_k = g(\epsilon_0, \epsilon_1, \ldots, \epsilon_k)$ for some measurable function g. Let $\mathbf{E}(A_k) = 0$ and for $2 < r < 4$, let $\mathbf{E}\|A_k\|^{r+\delta} < \infty$ for $\delta > 0$. Further, as $n \to \infty$, let $\mathbf{Cov}\left(\sum_{k=1}^n A_k\right)/n \to D$, where D is a positive definite matrix. Define $A_k^* = g(\epsilon_0^*, \epsilon_1, \ldots, \epsilon_k)$, where ϵ_0^* is an iid copy ϵ_0. If*

$$\Theta_{n;2+\delta} := \sum_{k=n}^{\infty} \|A_k - A_k^*\|_{(2+\delta)} = \mathcal{O}\left(n^{-(r-2)/(2(4-r))-\delta}\right) \text{ for all } n \geq 1,$$

then on a suitably rich probability space one can construct a Wiener process $\{W(t) : t > 0\}$, such that

$$\left\|\sum_{k=1}^n A_k - \mathbf{E}\left(\sum_{k=1}^n A_k\right) - D^{1/2}W(n)\right\| \overset{a.s.}{=} \mathcal{O}(n^{1/r}).$$

Proof of Theorem 1. We will start by showing some moment properties of the sequence of regeneration times in order to prove strong convergence. Denote the number of regenerations in a sample of size n by:

$$\xi(n) := \sup\{k \geq 1 : T_k \leq n\} = \inf\{k \geq 1 : T_{k+1} > n\}.$$

By [18], $\{(Z_k, \tau_k) : k \geq 1\}$ is a stationary 1-dependent process. Further, by the conditions in the theorem, $\mathbf{E}_Q(\tau_1^p) < \infty$ for $p > 1$. Define $p' = \min\{2, p\}$ for $p > 1$. Thus, $\mathbf{E}_Q(\tau_1^{p'}) < \infty$. By a Marcinkiewicz-Zygmund strong law of large numbers for 1-dependent process [see 39, for e.g.], for $1 < p' \leq 2$,

$$\left| \sum_{i=1}^{\xi(n)} \tau_i - \xi(n)\mu \right| = |T_{\xi(n)} - \xi(n)\mu| \overset{a.s.}{=} o(\xi(n)^{1/p'}) . \tag{20}$$

With $n < T_{\xi(n)+1}$, subtracting $\xi(n)\mu$ from both sides

$$n - \xi(n)\mu < T_{\xi(n)+1} - (\xi(n) + 1)\mu + \mu.$$

Now, using (20), $T_{\xi(n)+1} - (\xi(n)+1)\mu \overset{a.s.}{=} \mathcal{O}((\xi(n)+1)^{1/p'})$ and hence $T_{\xi(n)+1} - (\xi(n) + 1)\mu + \mu \overset{a.s.}{=} \mathcal{O}((\xi(n) + 1)^{1/p'})$. Thus

$$\begin{aligned} & n - \xi(n)\mu \overset{a.s.}{=} \mathcal{O}((\xi(n) + 1)^{1/p'}) \\ & \Rightarrow n - \xi(n)\mu \overset{a.s.}{=} \mathcal{O}(n^{1/p'}) \Rightarrow \xi(n) \overset{a.s.}{=} n/\mu + \mathcal{O}(n^{1/p'}) \qquad (21) \\ & \Rightarrow |\xi(n) - n/\mu| \overset{a.s.}{=} \mathcal{O}(n^{1/p'}). \qquad (22) \end{aligned}$$

Further, by the assumption in (6) and Lemma 6,

$$\begin{aligned} & \mathbf{E}_\pi \left[\left(\sum_{t=1}^{\tau_1} \left\| f(X_t) - \frac{\eta}{\mu} \right\| \right)^{2+\delta} \right] < \infty \\ \Rightarrow & \mathbf{E}_Q \left[\left(\sum_{t=1}^{\tau_1} \left\| f(X_t) - \frac{\eta}{\mu} \right\| \right)^{2+\delta} \right] < \infty \\ \Rightarrow & \mathbf{E}_Q \left[\left(\left\| Z_1 - \tau_1 \frac{\eta}{\mu} \right\| \right)^{2+\delta} \right] < \infty \\ \Rightarrow & \left\| \mathbf{E}_Q \left(Z_1 - \tau_1 \frac{\eta}{\mu} \right) \right\|^{2+\delta} < \infty , \end{aligned} \tag{23}$$

where the last implication follows from Jensen's inequality.

Define, $S_k := (\sum_{i=T_{k-1}+l}^{T_k} X_i, \sum_{i=T_k+1}^{T_k+l-1} X_i, T_k - T_{k-1} - 2)^\top$ for all $k \geq 2$ where $S_1 := (\sum_{i=T_0+1}^{T_1} X_i, \sum_{i=T_1+1}^{T_1+l-1} X_i, T_1 - T_0 + l - 2)^\top$ and $S_0 \sim S_1$ independently. By construction S_k's are iid vectors [see 18, Section 2]. Consequently, $g(S_0, S_1) = (Z_1 - \tau_1 \frac{\eta}{\mu})$ for some measurable function $g(\cdot)$. Again $X_{T_1+l} \sim Q$ and independent to all previous elements in the chain; define $g(S_0, S_1, S_2) = (Z_2 - \tau_2 \frac{\eta}{\mu})$. Hence, for all $k \geq 1$ we can say $g(S_0, S_1, \ldots, S_{k-1}, S_k) = (Z_k - \tau_k \frac{\eta}{\mu})$

Define, $(Z_k^* - \tau_k^* \frac{\eta}{\mu}) = g(S_0^*, S_1, \ldots, S_{k-1}, S_k)$, where S_0^* is an iid copy of S_0. See that $(Z_k^* - \tau_k^* \frac{\eta}{\mu}) = (Z_k - \tau_k \frac{\eta}{\mu})$ for all $k \geq 1$. So, $\Theta_{n;2+\delta} := \sum_{k=n}^\infty \|(Z_k - \tau_k \frac{\eta}{\mu}) - (Z_k^* - \tau_k^* \frac{\eta}{\mu})\|_{2+\delta} = 0$ for all $n \geq 1$. Also from (5)

$$\lim_{\xi(n)\to\infty} \frac{\mathbf{Var}\left(\sum_{k=1}^{\xi(n)} \left(Z_k - \tau_k \frac{\eta}{\mu}\right)\right)}{\xi(n)} = \Sigma_Z ,$$

which is a positive definite matrix. By Theorem 4, for the stationary 1-dependent process $\{(Z_k - \tau_k \frac{\eta}{\mu}) : k \geq 1\}$ and $\{W(t) : t > 0\}$, a d-dimensional standard Wiener process,

$$\left\| \sum_{k=1}^{\xi(n)} Z_k - T_{\xi(n)} \frac{\eta}{\mu} - \Sigma_Z^{1/2} W(\xi(n)) \right\| \overset{a.s.}{=} \mathcal{O}(\xi(n)^{1/(2+\delta)})$$

$$\Rightarrow \left\| \sum_{k=1}^{\xi(n)} Z_k - T_{\xi(n)} \frac{\eta}{\mu} - \Sigma_Z^{1/2} W(\xi(n)) \right\| \overset{a.s.}{=} \mathcal{O}(n^{1/(2+\delta)}). \tag{24}$$

By (21) and [9, Proposition 1.2.1],

$$\|W(\xi(n)) - W(n/\mu)\| \overset{a.s.}{=} \mathcal{O}(b_n)\,,$$

where for positive constants c and c',

$$\begin{aligned} b_n &= \left(2cn^{1/p'}\left(\log\left(\frac{n/\mu}{n^{1/p'}}\right) + \log\log\left(n/\mu\right)\right)\right)^{1/2} \\ &= \left(2cn^{1/p'}\left(\log\left(\frac{n^{1-1/p'}}{\mu}\right) + \log\log\left(n/\mu\right)\right)\right)^{1/2} \\ &< c'n^{1/(2p')}\left(\log n\right)\,. \end{aligned}$$

Consequently,

$$\|W(\xi(n)) - W(n/\mu)\| \overset{a.s.}{=} \mathcal{O}(n^{1/(2p')}\log n). \tag{25}$$

Using triangle inequality and since $T_{\xi(n)} < n < T_{\xi(n)+1}$,

$$\begin{aligned} Y_{\xi(n)} &:= \sum_{i=T_{\xi(n)}+1}^{T_{\xi(n)+1}} \left\| f(X_i) - \frac{\eta}{\mu} \right\| \\ &> \left\| \sum_{i=T_{\xi(n)}+1}^{n} \left(f(X_i) - \frac{\eta}{\mu} \right) \right\| \\ &= \left\| \sum_{i=1}^{n} \left(f(X_i) - \frac{\eta}{\mu} \right) - \sum_{k=1}^{\xi(n)} \left(Z_k - \tau_k \frac{\eta}{\mu} \right) \right\|. \end{aligned} \tag{26}$$

By construction, $\{Y_k\}_{k\geq 1}$ are positive and identical random variables generated from the sum of absolute values of correlated units sampled through wide-sense regeneration. By the integral transformation inequality and the assumption in equation-(6)

$$\begin{aligned} \sum_{i=1}^{\infty} \Pr\left({Y_i}^{2+\delta} > i\right) &= \sum_{i=1}^{\infty} \Pr\left({Y_1}^{2+\delta} > i\right) \\ &< \int_1^{\infty} \Pr\left({Y_1}^{2+\delta} > x\right) dx \end{aligned}$$

$$
\begin{aligned}
&< \int_0^\infty \Pr\left({Y_1}^{2+\delta} > x\right) dx \\
&= \mathbf{E}_Q[{Y_1}^{2+\delta}] \\
&< \infty.
\end{aligned}
$$

Consequently,

$$
\sum_{i=1}^{\infty} \Pr\left({Y_i}^{2+\delta} > i\right) < \infty \Rightarrow \sum_{i=1}^{\infty} \Pr\left(Y_i > i^{\frac{1}{2+\delta}}\right) < \infty. \tag{27}
$$

Thus by Borel-Cantelli lemma

$$
Y_n \overset{a.s.}{=} \mathcal{O}\left(n^{\frac{1}{2+\delta}}\right) \qquad \text{as } n \to \infty. \tag{28}
$$

From (26) and (28) as $n \to \infty$

$$
\begin{aligned}
&Y_{\xi(n)} \overset{a.s.}{=} \mathcal{O}\left(\xi(n)^{\frac{1}{2+\delta}}\right) \\
\Rightarrow &\left\| \sum_{i=1}^{n} \left(f(X_i) - \frac{\eta}{\mu}\right) - \sum_{k=1}^{\xi(n)} \left(Z_k - \tau_k \frac{\eta}{\mu}\right) \right\| \overset{a.s.}{=} \mathcal{O}\left(n^{\frac{1}{2+\delta}}\right).
\end{aligned} \tag{29}
$$

Using the triangle inequality and (24), (25), and (29)

$$
\begin{aligned}
\left\| \sum_{i=1}^{n} f(X_i) - n\frac{\eta}{\mu} - \frac{\Sigma_Z^{1/2}}{\sqrt{\mu}} W(n) \right\| &< \left\| \sum_{i=1}^{n} \left(f(X_i) - \frac{\eta}{\mu}\right) - \sum_{k=1}^{\xi(n)} \left(Z_k - \tau_k \frac{\eta}{\mu}\right) \right\| \\
&+ \left\| \sum_{k=1}^{\xi(n)} Z_k - T_{\xi(n)} \frac{\eta}{\mu} - \Sigma_Z^{1/2} W(\xi(n)) \right\| \\
&+ \left\| \Sigma_Z^{1/2} \left(W(\xi(n)) - W\left(\frac{n}{\mu}\right)\right) \right\|,
\end{aligned}
$$

This implies,

$$
\left\| \sum_{i=1}^{n} f(X_i) - n\frac{\eta}{\mu} - \frac{\Sigma_Z^{1/2}}{\sqrt{\mu}} W(n) \right\| = \mathcal{O}(n^{1/(2+\delta)}) + \mathcal{O}(n^{1/(2+\delta)}) + \mathcal{O}(n^{1/p} \log n).
$$

Thus, with $\beta = \max\{1/(2+\delta), 1/(2p')\} = \max\{1/(2+\delta), 1/(2p), 1/4\}$ and by Lemma-5 as $n \to \infty$

$$
\begin{aligned}
&\left\| \sum_{i=1}^{n} f(X_i) - n\frac{\eta}{\mu} - \frac{\Sigma_Z^{1/2}}{\sqrt{\mu}} W(n) \right\| \overset{a.s.}{=} \mathcal{O}(n^\beta \log n) \qquad (30) \\
\Rightarrow &\left\| \sum_{i=1}^{n} f(X_i) - n\mathbf{E}_\pi[f(X)] - \frac{\Sigma_Z^{1/2}}{\sqrt{\mu}} W(n) \right\| \overset{a.s.}{=} \mathcal{O}(n^\beta \log n).
\end{aligned}
$$

□

Appendix C: Proof of Theorem 2

Definition 2. *Let $\{X_t\}_{t\geq 1}$ be a stationary stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathcal{F}_p^q := \sigma(X_p, \dots, X_q)$ for some $q > p$. The process is said to be α-mixing or strongly mixing if $\alpha(n) \to 0$ as $n \to \infty$ where*

$$\alpha(n) := \sup_{k\geq 1} \sup_{A\in\mathcal{F}_1^k; B\in\mathcal{F}_{k+n}^\infty} |\Pr(A\cap B) - \Pr(A)\Pr(B)|.$$

Proof of Lemma 1. $\{X_t\}_{t\geq 1}$ is a polynomially ergodic sequence of random variables of order ξ for $\xi > (2+\delta)(1+(2+\delta)/\delta^*)$. So in (2), $G(n) = n^{-\xi}$. Further, from [23] we have that $\alpha(n) \leq n^{-\xi}$ for $n \geq 1$ (see Definition 2 for $\alpha(n)$). Consequently, for $p < \xi$

$$\sum_{n=1}^{\infty} n^{p-1}\alpha(n) < \sum_{n=1}^{\infty} n^{p-1} n^{-\xi} < \infty.$$

By [39, Proposition 3.1]

$$\mathbf{E}_\pi[\tau_1^p] < \infty \quad \text{for } p \in (0, \xi). \tag{31}$$

From Lemma 6 we have $\mathbf{E}_Q[\tau_1^p] < \infty$ for $p \in (0, \xi)$. □

Proof of Lemma 2. Since $\{X_t\}_{t\geq 1}$ is geometrically ergodic, $G(n) = t^n$ for some $0 < t < 1$ (see Definition 2 for $\alpha(n)$). From [23], $\alpha(n) \leq t^n$ for $n \geq 1$. Consequently for $p > 1$,

$$\sum_{n=1}^{\infty} n^{p-1}\alpha(n) < \sum_{n=1}^{\infty} n^{p-1} t^n.$$

By a ratio test, for all $p > 1$

$$\lim_{n\to\infty} \frac{(n+1)^{p-1}t^{n+1}}{n^{p-1}t^n} = \lim_{n\to\infty} (1+1/n)^{p-1} t = t < 1.$$

So,

$$\sum_{n=1}^{\infty} n^{p-1}\alpha(n) < \infty \quad \text{for } p > 1.$$

By [39, Proposition 3.1]

$$\mathbf{E}_\pi[\tau_1^p] < \infty \quad \text{for } p > 1. \tag{32}$$

From Lemma 6, $\mathbf{E}_Q[\tau_1^p] < \infty$ for $p > 1$. □

Proof of Lemma 3. For $r > 1$ and using triangle inequality for L^r-distances on π, Hölder's inequality, Markov's inequality, and, infinite sum of r-series,

$$\left(\mathbf{E}_\pi\left[\left(\sum_{i=1}^{\tau_1} \|f(X_i)\|\right)^r\right]\right)^{1/r}$$

$$
\begin{aligned}
&= \left(\mathbf{E}_\pi\left[\left(\sum_{i=1}^{\infty}\mathbb{I}(i\le\tau_1)\|f(X_i)\|\right)^r\right]\right)^{1/r}\\
&\le \sum_{i=1}^{\infty}\left(\mathbf{E}_\pi\left(\mathbb{I}(i\le\tau_1)\|f(X_i)\|^r\right)\right)^{1/r}\\
&\le \sum_{i=1}^{\infty}\left(\left(\mathbf{E}_\pi\mathbb{I}(i\le\tau_1)\right)^{\delta^*/(r+\delta^*)}\left(\mathbf{E}_\pi\left(\|f(X_i)\|^{r+\delta^*}\right)\right)^{r/(r+\delta^*)}\right)^{1/r}\\
&= \left(\mathbf{E}_\pi\left(\|f(X)\|^{r+\delta^*}\right)\right)^{1/(r+\delta^*)}\sum_{i=1}^{\infty}\left(\Pr_\pi(\tau_1\ge i)\right)^{\delta^*/r(r+\delta^*)}\\
&\le \left(\mathbf{E}_\pi\left(\|f(X)\|^{r+\delta^*}\right)\right)^{1/(r+\delta^*)}\sum_{i=1}^{\infty}\left(\Pr_\pi(\tau_1^\phi\ge i^\phi)\right)^{\delta^*/r(r+\delta^*)}\\
&\le \left(\mathbf{E}_\pi\left(\|f(X)\|^{r+\delta^*}\right)\right)^{1/(r+\delta^*)}\sum_{i=1}^{\infty}\left(\frac{\mathbf{E}_\pi(\tau_1^\phi)}{i^\phi}\right)^{\delta^*/r(r+\delta^*)}\\
&= \left(\mathbf{E}_\pi\left(\|f(X)\|^{r+\delta^*}\right)\right)^{1/(r+\delta^*)}\left(\mathbf{E}_\pi(\tau_1^\phi)\right)^{\delta^*/r(r+\delta^*)}\sum_{i=1}^{\infty}\left(\frac{1}{i^\phi}\right)^{\delta^*/r(r+\delta^*)} < \infty\,.
\end{aligned}
$$

□

Proof of Lemma 4. Since $\{X_t\}_{t\ge1}$ is geometrically ergodic, by (32), $\mathbf{E}_\pi[\tau_1^q]<\infty$ for $q>1$. Proceeding similarly as Lemma 3, $\mathbf{E}_\pi\left(\left(\sum_{i=1}^{\tau_1}\|f(X_i)\|\right)^r\right)<\infty$. □

Proof of Theorem 2. (*a*) $\{X_t\}_{t\ge1}$ is polynomially ergodic of order ξ for $\xi>(2+\delta)(1+(2+\delta)/\delta^*)$. Thus (31) holds. Then by Lemma 1,

$$
\mathbf{E}_Q[\tau_1^p]<\infty \qquad \text{for } p\in(0,\xi). \tag{33}
$$

By the assumption in the theorem, $\mathbf{E}_\pi\left(\|f(X)\|^{(2+\delta+\delta^*/2)+\delta^*/2}\right)<\infty$ for some $\delta>0$ and $\delta^*>0$. By (31) and Lemma 3

$$
\begin{aligned}
&\mathbf{E}_\pi\left[\left(\sum_{i=1}^{\tau_1}\|f(X_i)\|\right)^{2+\delta+\delta^*/2}\right]<\infty\\
\Rightarrow\ &\mathbf{E}_\pi\left[\left(\sum_{i=1}^{\tau_1}\left\|f(X_i)-\frac{\eta}{\mu}\right\|\right)^{2+\delta+\delta^*/2}\right]<\infty.
\end{aligned} \tag{34}
$$

By Theorem 1, (33), and (34),

$$
\left\|\sum_{i=1}^{n}f(X_i)-n\mathbf{E}_\pi[f(X)]-\frac{\Sigma_Z^{1/2}}{\sqrt{\mu}}W(n)\right\| \overset{a.s.}{=} \mathcal{O}(n^\beta\log(n)) \tag{35}
$$

as $n\to\infty$ where $\beta=\max\{1/(2+\delta),1/(2p),1/4\}$ for all $p\in(0,\xi)$. Since $\xi>2$ always, $\beta=\max\{1/(2+\delta),1/4\}$.

(b) Let $\{X_t\}_{t\geq 1}$ be geometrically ergodic. So (32) holds. By Lemma 2

$$\mathbf{E}_Q[\tau_1^p] < \infty \ \forall p > 1. \tag{36}$$

Since $\mathbf{E}_\pi\left(\|f(X)\|^{(2+\delta+\delta^*/2)+\delta^*/2}\right) < \infty$ for some $\delta > 0$ and $\delta^* > 0$, by (32) and Lemma 4

$$\mathbf{E}_\pi\left[\left(\sum_{i=1}^{\tau_1}\|f(X_i)\|\right)^{2+\delta+\delta^*/2}\right] < \infty$$

$$\Rightarrow \mathbf{E}_\pi\left[\left(\sum_{i=1}^{\tau_1}\left\|f(X_i) - \frac{\eta}{\mu}\right\|\right)^{2+\delta+\delta^*/2}\right] < \infty. \tag{37}$$

By Theorem 1, (36), and (37), with $\beta = \max\{1/(2+\delta), 1/4\}$ as $n \to \infty$

$$\left\|\sum_{i=1}^{n} f(X_i) - n\mathbf{E}_\pi[f(X)] - \frac{\Sigma_Z^{1/2}}{\sqrt{\mu}}W(n)\right\| \overset{a.s.}{=} \mathcal{O}(n^\beta \log(n)). \tag{38}$$

□

Appendix D: Proof of Corollary 1

From Theorem 2, $\beta = \max\{1/(2+\delta), 1/4\}$. Hence, if $b_n^{-1}\log(n)\kappa^2(n) \to 0$ as $n \to \infty$, then $\log^3(n)n^{\{\max\{2/(2+\delta),1/2\}-\nu\}\}} \to 0$ as $n \to \infty$. Hence, a choice of b_n such that $\nu > \max\{2/(2+\delta), 1/2\}$ will ensure the strong consistency of batch means estimator.

Appendix E: Theory for illustrative examples

E.1. Bivariate normal Gibbs sampler

A 1-step minorization is obtained from k_{DS} by using a "distinguished point technique" as described in [see 33]. Define a distinguished point y^* in a small set $D = \{(u, v) \mid u \in [\mu_1 - h, \mu_1 + h], v \in [\mu_2 - h, \mu_2 + h]; h > 0\}$ and π denote the bivariate Gaussian target density. Then,

$$\begin{aligned}
k_{\text{DS}}(x_{i+1}, y_{i+1} \mid x_i, y_i) &= \frac{\pi(x_{i+1} \mid y_i)}{\pi(x_{i+1} \mid y^*)}\pi(y_{i+1} \mid x_{i+1})\pi(x_{i+1} \mid y^*) \\
&\geq \inf_{x\in D}\left\{\frac{\pi(x \mid y_i)}{\pi(x \mid y^*)}\right\}\pi(y_{i+1} \mid x_{i+1})\ \pi(x_{i+1} \mid y^*)\ I_D(x_{i+1}) \\
&= s(y_i)\ Q(x_{i+1}, y_{i+1}),
\end{aligned}$$

where $Q(x_{i+1}, y_{i+1}) := \epsilon^{-1}\pi(y_{i+1} \mid x_{i+1})\ \pi(x_{i+1} \mid y^*)\ I_D(x_{i+1})$ such that for $\Phi_{(\mu;\ \sigma^2)}(\cdot)$ to be the probability measure for Normal(μ, σ^2)

$$\epsilon := \int_{\mathbb{R}}\int_{\mathbb{R}} \pi(y_{i+1} \mid x_{i+1})\ \pi(x_{i+1} \mid y^*)\ I_D(x_{i+1})\ dy_{i+1}\ dx_{i+1}$$

$$
\begin{aligned}
&= \int_D \left\{ \int_{\mathbb{R}} \pi(y_{i+1} \mid x_{i+1})\ dy_{i+1} \right\}\ \pi(x_{i+1} \mid y^*)\ dx_{i+1} \\
&= \int_D \pi(x_{i+1} \mid y^*)\ dx_{i+1} \\
&= \Phi_{\left(\mu_1 + \frac{\rho}{\sigma_2^2}(y^* - \mu_2);\ \sigma_1^2 - \frac{\rho^2}{\sigma_2^2}\right)} (D)\,.
\end{aligned}
$$

Similarly, the minorization constant will be

$$
\begin{aligned}
s(y_i) &:= \epsilon \inf_{x \in D} \left\{ \frac{\pi(x \mid y_i)}{\pi(x \mid y^*)} \right\} \\
&= \epsilon \inf_{x \in D} \left[\exp\left\{ -\frac{(x - \mu_1 - \frac{\rho}{\sigma_2^2}(y_i - \mu_2))^2 - (x - \mu_1 - \frac{\rho}{\sigma_2^2}(y^* - \mu_2))^2}{2\left(\sigma_1^2 - \frac{\rho^2}{\sigma_2^2}\right)} \right\} \right] \\
&= \epsilon \exp\left[-\frac{\rho^2\left\{(y_i - \mu_2)^2 - (y^* - \mu_2)^2\right\}}{2\sigma_2^2(\sigma_1^2\sigma_2^2 - \rho^2)} \right] \\
&\quad \times \inf_{x \in D} \left[\exp\left\{ \frac{\rho\sigma_2^2(x - \mu_1)(y_i - y^*)}{\sigma_1^2\sigma_2^2 - \rho^2} \right\} \right].
\end{aligned}
$$

Clearly, the value of x required for minimizing the above function is dependent on ρ and $(y - y^*)$. A detailed table of $(x - \mu_1)$ values are given in the following table.

ρ	$(y - y^*)$	$(x - \mu_1)$
$\rho > 0$	$(y - y^*) \geq 0$	$-h$
$\rho > 0$	$(y - y^*) < 0$	h
$\rho \leq 0$	$(y - y^*) \geq 0$	h
$\rho \leq 0$	$(y - y^*) < 0$	$-h$

The 2-step minorization kernel for the deterministic scan Gibbs sampler can be achieved as

$$
\begin{aligned}
k_{DS}^2\left(x_{i+2}, y_{i+2} \mid x_i, y_i\right) &:= \int_{\mathbb{R}} \int_{\mathbb{R}} k_{DS}\left(x_{i+2}, y_{i+2} \mid x_{i+1}, y_{i+1}\right) \times \\
&\qquad\qquad k_{DS}\left(x_{i+1}, y_{i+1} \mid x_i, y_i\right)\ dy_{i+1}\ dx_{i+1} \\
&\geq s(y_i)\ Q(x_{i+2}, y_{i+2}) \times \\
&\qquad \int_{\mathbb{R}} \int_{\mathbb{R}} s(y_{i+1}) Q(x_{i+1}, y_{i+1})\ dy_{i+1}\ dx_{i+1} \\
&= s^*(y_i)\ Q(x_{i+2}, y_{i+2}),
\end{aligned}
$$

where $s^*(y_i) := s(y_i)\mathbf{E}_Q(s(y))$. Determining the form of s^* and hence the form of $\mathbf{E}_Q(s(y))$ is crucial for determining regeneration times.

E.2. Probit regression model

Establishing a 1-step minorization with the corresponding small set calculations has been done for deterministic scan Gibbs samplers (see [33], [38]). Using a random scan version of the Gibbs sampler of [1], we present a framework for identifying wide-sense regenerations.

The deterministic scan AC sampler is geometrically ergodic [see 38, Theorem 1] and regenerations can be identified using "distinguished point" technique of [33]. For a distinguished point $\mathbf{z}^* \in \mathbb{R}^n$ and a rectangular small set $D^* = [c_1, d_1] \times [c_2, d_2] \times \dots [c_p, d_p]$ the minorization kernel is

$$Q(\beta_{i+1}, \mathbf{z}_{i+1}) := \frac{1}{\epsilon} \pi(\beta_{i+1} \mid \mathbf{z}^*, \mathbf{y}) \pi(\mathbf{z}_{i+1} \mid \beta_{i+1}, \mathbf{y}) I_{D^*}(\beta_{i+1}),$$

where $\epsilon = \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} \pi(\beta_{i+1} \mid \mathbf{z}^*, \mathbf{y})\ \pi(\mathbf{z}_{i+1} \mid \beta_{i+1}, \mathbf{y}) I_{D^*}(\beta_{i+1})\ d\mathbf{z}_{i+1} d\beta_{i+1}$. For t_j being the j^{th} term of $t = (\mathbf{z}_i - \mathbf{z}_*)^T X$, the minorization constant will be

$$s(\mathbf{z}_i) := \frac{\epsilon\ \exp\left(\sum_{j=1}^{p} (c_j t_j I_{\mathbb{R}^+}(t_j) + d_j t_j I_{\mathbb{R}^-}(t_j))\right)}{\exp(0.5(\mathbf{z}_i)^T X^T (X^T X)^{-1} X(\mathbf{z}_i) - 0.5(\mathbf{z}^*)^T X^T (X^T X)^{-1} X(\mathbf{z}^*))}.$$

Consequently, the minorization holds with

$$k_{\text{DS}}(\beta_{i+1}, \mathbf{z}_{i+1} \mid \beta_i, \mathbf{z}_i) \geq s(\mathbf{z}_i)\ Q(\beta_{i+1}, \mathbf{z}_{i+1}). \tag{39}$$

For Δ being the Dirac measure, the one-step random scan AC-sampler kernel is defined as

$$\begin{aligned} k_{\text{RS}}(\beta_{i+1}, \mathbf{z}_{i+1} \mid \beta_i, \mathbf{z}_i) =& p\pi\left(\beta_{i+1} \mid \mathbf{z}_i\right) \Delta_{\mathbf{z}_i}(z_{i+1}) \\ &+ (1-p)\pi\left(z_{i+1} \mid \beta_{i+1}\right) \Delta_{\beta_i}(\beta_{i+1}). \end{aligned}$$

Similarly, the 2-step random scan Markov transition kernel is

$$\begin{aligned} &k^2_{\text{RS}}(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_i, \mathbf{z}_i) \\ &= \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} k_{\text{RS}}(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_{i+1}, \mathbf{z}_{i+1}) k_{\text{RS}}(\beta_{i+1}, \mathbf{z}_{i+1} \mid \beta_i, \mathbf{z}_i)\ d\mathbf{z}_{i+1} d\beta_{i+1} \\ &= \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} \left(p\pi\left(\beta_{i+2} \mid \mathbf{z}_{i+1}\right) \Delta_{\mathbf{z}_{i+1}}(\mathbf{z}_{i+2}) + (1-p)\pi\left(\mathbf{z}_{i+2} \mid \beta_{i+1}\right) \Delta_{\beta_{i+1}}(\beta_{i+2})\right) \\ &\qquad \left(p\pi\left(\beta_{i+1} \mid \mathbf{z}_i\right) \Delta_{\mathbf{z}_i}(\mathbf{z}_{i+1}) + (1-p)\pi\left(\mathbf{z}_{i+1} \mid \beta_i\right) \Delta_{\beta_i}(\beta_{i+1})\right)\ d\mathbf{z}_{i+1} d\beta_{i+1} \\ &= \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} p^2 \pi\left(\beta_{i+2} \mid \mathbf{z}_{i+1}\right) \pi\left(\beta_{i+1} \mid \mathbf{z}_i\right) \Delta_{\mathbf{z}_i}(\mathbf{z}_{i+1}) \Delta_{\mathbf{z}_{i+1}}(\mathbf{z}_{i+2})\ d\mathbf{z}_{i+1} d\beta_{i+1} \\ &\quad + \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} p(1-p)\pi\left(\beta_{i+2} \mid \mathbf{z}_{i+1}\right) \pi\left(\mathbf{z}_{i+1} \mid \beta_i\right) \Delta_{\beta_i}(\beta_{i+1}) \Delta_{\mathbf{z}_{i+1}}(\mathbf{z}_{i+2})\ d\mathbf{z}_{i+1} d\beta_{i+1} \\ &\quad + \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} p(1-p)\pi\left(\mathbf{z}_{i+2} \mid \beta_{i+1}\right) \pi\left(\beta_{i+1} \mid \mathbf{z}_i\right) \Delta_{\mathbf{z}_i}(\mathbf{z}_{i+1}) \Delta_{\beta_{i+1}}(\beta_{i+2})\ d\mathbf{z}_{i+1} d\beta_{i+1} \\ &\quad + \int_{\mathbb{R}^p} \int_{\mathbb{R}^n} (1-p)^2 \pi\left(\mathbf{z}_{i+2} \mid \beta_{i+1}\right) \pi\left(\mathbf{z}_{i+1} \mid \beta_i\right) \Delta_{\beta_i}(\beta_{i+1}) \Delta_{\beta_{i+1}}(\beta_{i+2})\ d\mathbf{z}_{i+1} d\beta_{i+1}. \end{aligned}$$

So,

$$\begin{aligned} k^2_{\text{RS}}\left(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_i, \mathbf{z}_i\right) = {} & p^2 \pi\left(\beta_{i+2} \mid \mathbf{z}_{i+2}\right) \Delta_{\mathbf{z}_i}(\mathbf{z}_{i+2}) \\ & + p(1-p)\pi\left(\beta_{i+2} \mid \mathbf{z}_{i+2}\right) \ \pi\left(\mathbf{z}_{i+2} \mid \beta_i\right) \\ & + p(1-p)\pi\left(\mathbf{z}_{i+2} \mid \beta_{i+2}\right)\pi\left(\beta_{i+2} \mid \mathbf{z}_i\right) \\ & + (1-p)^2 \pi\left(\mathbf{z}_{i+2} \mid \beta_i\right) \Delta_{\beta_i}(\beta_{i+2}). \end{aligned} \tag{40}$$

Using the 1-step minorization of the deterministic scan Gibbs sampler and with $s'(\mathbf{z}_i) := p(1-p)s(\mathbf{z}_i)$, we get

$$\begin{aligned} & k^2_{RS}\left(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_i, \mathbf{z}_i\right) \\ & = p^2 \pi\left(\beta_{i+2} \mid \mathbf{z}_{i+2}\right) \Delta_{\mathbf{z}_i}(\mathbf{z}_{i+2}) + p(1-p)\pi\left(\beta_{i+2} \mid \mathbf{z}_{i+2}\right) \ \pi\left(\mathbf{z}_{i+2} \mid \beta_i\right) \\ & \quad + p(1-p)\pi\left(\mathbf{z}_{i+2} \mid \beta_{i+2}\right)\pi\left(\beta_{i+2} \mid \mathbf{z}_i\right) + (1-p)^2 \pi\left(\mathbf{z}_{i+2} \mid \beta_i\right) \Delta_{\beta_i}(\beta_{i+2}) \\ & \geq p(1-p)\pi\left(\mathbf{z}_{i+2} \mid \beta_{i+2}\right)\pi\left(\beta_{i+2} \mid \mathbf{z}_i\right) \\ & \geq p(1-p)s(\mathbf{z}_i) \ Q(\beta_{i+2}, \mathbf{z}_{i+2}) \\ & = s'(\mathbf{z}_i) \ Q(\beta_{i+2}, \mathbf{z}_{i+2}). \end{aligned} \tag{41}$$

Following the ideas in [33] and [38], we can obtain the probability of a regeneration from the observed chain in the following way:

$$\begin{aligned} \eta_i & = \Pr\left(\delta_i = 1 \mid (\beta_i, \mathbf{z}_i), (\beta_{i+2}, \mathbf{z}_{i+2})\right); \\ & = \frac{\Pr(X_{i+2} = y \mid X_i = x, \delta_i = 1)\Pr(\delta_i = 1 \mid X_i = x)\Pr(X_i = x)}{\Pr(X_{i+2} = y \mid X_i = x)\Pr(X_i = x)}; \\ & = \frac{\Pr(X_{i+2} = y \mid X_i = x, \delta_i = 1)\Pr(\delta_i = 1 \mid X_i = x)}{\Pr(X_{i+2} = y \mid X_i = x)}; \\ & = \frac{s'(\mathbf{z}_i) \ Q(\beta_{i+2}, \mathbf{z}_{i+2})}{k^2_{\text{RS}}\left(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_i, \mathbf{z}_i\right)}; \\ & = p(1-p)\frac{\exp(-0.5(\mathbf{z}_i)^T X^T (X^T X)^{-1} X(\mathbf{z}_i)))}{\exp(-0.5(\mathbf{z}_*)^T X^T (X^T X)^{-1} X(\mathbf{z}_*)))}; \\ & \times \exp\left\{ \sum_{i=1}^{p} (c_i t_i I_{\mathbb{R}^+}(t_i) + d_i t_i I_{\mathbb{R}^-}(t_i)) \right\}; \\ & \times \pi(\beta_{i+2} \mid \mathbf{z}^*, \mathbf{y}) \ \pi(\mathbf{z}_{i+2} \mid \beta_{i+2}, \mathbf{y}) I_{D^*}(\beta_{i+2}); \\ & \times \frac{1}{k^2_{\text{RS}}\left(\beta_{i+2}, \mathbf{z}_{i+2} \mid \beta_i, \mathbf{z}_i\right)}. \end{aligned}$$

The η_i can be calculated analytically from a run of the random scan Gibbs sampler. By drawing Bernoulli samples with success probability η_i we identify the regenerations if the outcome of a trial is 1.

Acknowledgments

The authors are grateful to Galin Jones, James Flegal, Jing Dong and Sanket Agrawal for their helpful conversations. Dootika Vats is supported by SERB (SPG/2021/001322).

References

[1] Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88** 669–679. MR1224394

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

[3] Berg, S. and Song, H. (2023). Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain. *The Annals of Statistics* **51** 2440–2470. MR4682704

[4] Berkes, I. and Philipp, W. (1979). Approximation thorems for independent and weakly dependent random vectors. *The Annals of Probability* 29–54. MR0515811

[5] Chakraborty, S., Bhattacharya, S. K. and Khare, K. (2022). Estimating accuracy of the MCMC variance estimator: Asymptotic normality for batch means estimators. *Statistics & Probability Letters* **183** 109337. MR4363701

[6] Chan, K. W. (2022). Mean-structure and autocorrelation consistent covariance matrix estimation. *Journal of Business & Economic Statistics* **40** 201–215. MR4356567

[7] Chen, D.-F. R. and Seila, A. F. (1987). Multivariate inference in stationary simulation using batch means. In *Proceedings of the 19th conference on Winter simulation* 302–304.

[8] Csáki, E. and Csörgő, M. (1995). On additive functionals of Markov chains. *Journal of Theoretical Probability* **8** 905–919. MR1353559

[9] Csörgo, M. and Révész, P. (2014). *Strong approximations in probability and statistics*. Academic press. MR0666546

[10] Dai, N. and Jones, G. L. (2017). Multivariate initial sequence estimators in Markov chain Monte Carlo. *Journal of Multivariate Analysis* **159** 184–199. MR3668555

[11] Damerdji, H. (1991). Strong consistency and other properties of the spectral variance estimator. *Management Science* **37** 1424–1440.

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

[13] Dong, J. and Glynn, P. (2019). A new approach to sequential stopping for stochastic simulation. *Preprint*.

[14] Eberlein, E. (1986). On strong invariance principles under dependence assumptions. *The Annals of Probability* 260–270. MR0815969

[15] Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *Journal of multivariate analysis* **28** 20–68. MR0996984

[16] Flegal, J. M. and Jones, G. L. (2010). Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics* **38** 1034–1070. MR2604704

[17] Geyer, C. J. (1992). Practical Markov chain Monte Carlo. *Statistical science* 473–483.

[18] Glynn, P. W. (2011). Wide-sense regeneration for Harris recurrent Markov processes: an open problem. *Queueing Systems* **68** 305–311. MR2834201

[19] Glynn, P. W. and Whitt, W. (1992). The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability* **2** 180–198. MR1143399

[20] Gupta, K. and Vats, D. (2020). Estimating Monte Carlo variance from multiple Markov chains. *arXiv preprint arXiv:2007.04229*.

[21] Henderson, S. G. and Glynn, P. W. (2001). Regenerative steady-state simulation of discrete-event systems. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **11** 313–345.

[22] Hobert, J. P., Jones, G. L., Presnell, B. and Rosenthal, J. S. (2002). On the applicability of regenerative simulation in Markov chain Monte Carlo. *Biometrika* **89** 731–743. MR1946508

[23] Jones, G. L. (2004). On the Markov chain central limit theorem. *Probability surveys* **1** 299–320. MR2068475

[24] Jones, G. L., Haran, M., Caffo, B. S. and Neath, R. (2006). Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association* **101** 1537–1547. MR2279478

[25] Komlós, J., Major, P. and Tusnády, G. (1975). An approximation of partial sums of independent RV'-s, and the sample DF. I. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **32** 111–131. MR0375412

[26] Komlós, J., Major, P. and Tusnády, G. (1976). An approximation of partial sums of independent RV's, and the sample DF. II. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **34** 33–58. MR0402883

[27] Kosorok, M. R. (2000). Monte Carlo error estimation for multivariate Markov chains. *Statistics & probability letters* **46** 85–93. MR1731349

[28] Kuelbs, J. and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing B-valued random variables. *The Annals of Probability* 1003–1036. MR0602377

[29] Liu, W. and Lin, Z. (2009). Strong approximation for a class of stationary processes. *Stochastic Processes and their Applications* **119** 249–280. MR2485027

[30] Merlevède, F. and Rio, E. (2015). Strong approximation for additive functionals of geometrically ergodic Markov chains. *Electronic Journal of Probability* **20** 1–27. MR3317156

[31] Meyn, S. P. and Tweedie, R. L. (2012). *Markov chains and stochastic stability*. Springer Science & Business Media. MR2509253

[32] Monrad, D. and Philipp, W. (1991). The problem of embedding vector-valued martingales in a Gaussian process. *Theory of Probability & Its Applications* **35** 374–377. MR1069143

[33] Mykland, P., Tierney, L. and Yu, B. (1995). Regeneration in Markov

chain samplers. *Journal of the American Statistical Association* **90** 233–241. MR1325131

[34] Pengel, A. and Bierkens, J. (2024). Strong invariance principles for ergodic Markov processes. *Electronic Journal of Statistics* **18** 191–246. MR4698638

[35] Philipp, W. and Stout, W. F. (1975). *Almost sure invariance principles for partial sums of weakly dependent random variables* **161**. American Mathematical Soc. MR0433597

[36] Roberts, G. O. and Rosenthal, J. S. (2004). General state space Markov chains and MCMC algorithms. *Probability Surveys* **1** 20 – 71. https://doi.org/10.1214/154957804100000024. MR2095565

[37] Roy, V. (2020). Convergence diagnostics for Markov chain Monte Carlo. *Annual Review of Statistics and Its Application* **7** 387–412. MR4104198

[38] Roy, V. and Hobert, J. P. (2007). Convergence rates and asymptotic standard errors for Markov chain Monte Carlo algorithms for Bayesian probit regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69** 607–623. MR2370071

[39] Samur, J. D. (2004). A regularity condition and a limit theorem for Harris ergodic Markov chains. *Stochastic processes and their applications* **111** 207–235. MR2056537

[40] Seila, A. F. (1982). Multivariate estimation in regenerative simulation. *Operations Research Letters* **1** 153–156. MR0687356

[41] South, L. F., Riabiz, M., Teymur, O. and Oates, C. J. (2022). Post-processing of MCMC. *Annual Review of Statistics and Its Application* **9** 529–555. MR4394919

[42] Vats, D. and Flegal, J. M. (2021). Lugsail lag windows for estimating time-average covariance matrices. *Biometrika* **109** 735-750. https://doi.org/10.1093/biomet/asab049. MR4472845

[43] Vats, D., Flegal, J. M. and Jones, G. L. (2018). Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli* **24** 1860–1909. MR3757517

[44] Vats, D., Flegal, J. M. and Jones, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. *Biometrika* **106** 321–337. MR3949306

[45] Zhu, Y. (2020). Asymptotic Uncertainty Quantification and Its Application in Efficient Sampling and Learning, PhD thesis, Northwestern University. MR4132371