Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain

Stephen Berg* and Hyebin Song*†
Department of Statistics, Pennsylvania State University

November 13, 2023

Abstract

In this paper, we study the problem of estimating the autocovariance sequence resulting from a reversible Markov chain. A motivating application for studying this problem is the estimation of the asymptotic variance in central limit theorems for Markov chains. We propose a novel shape-constrained estimator of the autocovariance sequence, which is based on the key observation that the representability of the autocovariance sequence as a moment sequence imposes certain shape constraints. We examine the theoretical properties of the proposed estimator and provide strong consistency guarantees for our estimator. In particular, for geometrically ergodic reversible Markov chains, we show that our estimator is strongly consistent for the true autocovariance sequence with respect to an ℓ_2 distance, and that our estimator leads to strongly consistent estimates of the asymptotic variance. Finally, we perform empirical studies to illustrate the theoretical properties of the proposed estimator as well as to demonstrate the effectiveness of our estimator in comparison with other current state-of-the-art methods for Markov chain Monte Carlo variance estimation, including batch means, spectral variance estimators, and the initial convex sequence estimator.

Keywords: Markov chain Monte Carlo, Shape-constrained inference, Autocovariance sequence estimation, Asymptotic variance

*Both authors contributed equally.

†Corresponding author: hps5320@psu.edu

1 Introduction

Markov chain Monte Carlo (MCMC) is a routinely used tool for approximating intractable integrals of the form $\mu = \int g(x)\pi(dx)$, where π is an intractable probability measure on a measurable space $(\mathsf{X}, \mathscr{X})$ and $g : \mathsf{X} \to \mathbb{R}$ is a π-integrable function. In MCMC, a Markov chain $X_0, X_1, X_2, ...$ with transition kernel Q and stationary probability measure π is simulated for some finite number of iterations M, possibly after an initial burn-in period, and μ can then be estimated by the empirical average

$$Y_M = M^{-1} \sum_{t=0}^{M-1} g(X_t).$$

In general, $g(X_t)$ from a Markov chain may have nonzero covariance. For a Markov chain transition kernel Q with a unique stationary probability measure π, define the autocovariance sequence $\gamma = \{\gamma(k)\}_{k=-\infty}^{\infty}$

$$\gamma(k) = E_\pi[(g(X_0) - \mu)(g(X_{|k|}) - \mu)], \qquad k \in \mathbb{Z}.$$

In this work, we study the problem of estimating the autocovariance sequence $\gamma \in \mathbb{R}^{\mathbb{Z}}$ from a reversible Markov chain by exploiting shape constraints satisfied by the autocovariance sequence γ. It is a well known result that for a reversible Markov chain, the autocovariance sequence γ admits the following representation [e.g., Rudin, 1991]:

$$\gamma(k) = \int x^{|k|} F(dx) \qquad k \in \mathbb{Z} \tag{1}$$

for a unique positive measure F supported on $[-1, 1]$. For a function on $\mathbb{R}$ or $\mathbb{Z}$, admitting a certain mixture representation has an implication in its global shape [Hausdorff, 1921, Feller, 1939, Steutel, 1969, Jewell, 1982, Balabdaoui and de Fournas-Labrosse, 2020]. For instance, if the support of F in (1) is contained in $[0, 1]$, γ is *completely monotone*, meaning the inequalities $(-1)^n \Delta^n \gamma(j) \geq 0$ are satisfied for all $j, n \in \mathbb{N}$ where $\Delta^n \gamma(j) = \Delta^{n-1}\gamma(j + 1) - \Delta^{n-1}\gamma(j)$ is a difference operator with $\Delta^0 \gamma = \gamma$. While γ is not, in general, completely monotone because the support of F may extend outside of $[0, 1]$, the representation (1) still imposes an infinite number of shape constraints on γ (see Proposition 2). To exploit such structure in γ, in this work, we propose an estimator of the autocovariance sequence based on the ℓ_2 projection of an initial input autocovariance sequence estimate, such as

the ordinary empirical autocovariance sequence, onto the set of sequences admitting a representation as in (1).

1.1 Main application: asymptotic variance estimation for MCMC estimates

There are several motivations for the estimation of the autocovariance sequence. As a main application, we consider the problem of estimating the asymptotic variance in a Markov chain central limit theorem. This problem has practical importance, as the asymptotic variance quantifies uncertainties in the MCMC estimate Y_M. Under mild conditions [Meyn and Tweedie, 2009], a central limit theorem can be established for Y_M such that

$$\sqrt{M}(Y_M - \mu) \xrightarrow{d} N(0, \sigma^2) \tag{2}$$

where

$$\sigma^2 = \sum_{k=-\infty}^{\infty} \gamma(k). \tag{3}$$

The infinite sum in (3) arises from covariance between terms in the sum in the definition of Y_M. From (2), the variance of the empirical mean Y_M from an MCMC simulation as an estimator of μ is quantified, in an asymptotic sense, by the asymptotic variance σ^2. In turn, from (3), σ^2 can be estimated based on an estimate of the autocovariance sequence γ. Fixed width stopping rules for MCMC, as in Jones et al. [2006], Bednorz and Latuszyński [2007], Flegal et al. [2008], Latuszynski [2009], Flegal and Gong [2015], and Vats et al. [2019], depend on an estimate of σ^2.

One natural estimate for $\gamma(k)$ based on the first M iterates $X_0, X_1, ..., X_{M-1}$ is the empirical autocovariance $\tilde{r}_M(k)$, defined by

$$\tilde{r}_M(k) = \begin{cases} \frac{1}{M} \sum_{t=0}^{M-1-|k|} (g(X_t) - Y_M)(g(X_{t+|k|}) - Y_M) & |k| \leq M-1 \\ 0 & |k| > (M-1). \end{cases} \tag{4}$$

It is well known that some natural estimators of σ^2 based on the $\tilde{r}_M(k)$ sequence are inconsistent. For the empirical autocovariances with mean centering based on the empirical mean Y_M as in (4), an elementary calculation shows that $\sum_{k=-(M-1)}^{M-1} \tilde{r}_M(k) = 0$, and the

estimator $\hat{\sigma}^2_{M,\text{emp}} = \sum_{k=-(M-1)}^{M-1} \tilde{r}_M(k) = 0$ is thus inconsistent as an estimator of σ^2. With centering based on the true mean μ rather than Y_M in (4), the corresponding estimator converges in distribution to a scaled χ^2 random variable [Anderson, 1971, Flegal and Jones, 2010], and is thus also inconsistent. These difficulties have led to methods for estimating σ^2 with better statistical properties. These methods include spectral variance estimators [Anderson, 1971, Flegal and Jones, 2010], estimators based on batch means [Priestley, 1981, Flegal and Jones, 2010, Chakraborty et al., 2022], and a class of methods for reversible Markov chains called initial sequence estimators [Geyer, 1992, Kosorok, 2000, Dai and Jones, 2017].

The batch means and spectral variance estimators have known consistency properties. In particular, they are a.s. consistent for σ^2, and have $M^{1/3}$ rate of convergence with an optimal choice of batch or window size [Damerdji, 1991, Flegal and Jones, 2010]. Practically, they involve tuning parameters which are known in advance only up to a constant of proportionality. For instance, the batch means, overlapping batch means, and spectral variance estimators in Flegal and Jones [2010] require the selection of a batch size b_M depending on the Markov chain sample length M. The optimal setting is $b_M = CM^{1/3}$, but the constant of proportionality depends on problem-dependent parameters that will typically be unknown.

Geyer [1992], on the other hand, introduces initial sequence estimators for estimating σ^2. The initial sequence estimators exploit positivity, monotonicity, and convexity constraints satisfied for reversible Markov chains by the sequence $\Gamma = \{\Gamma(k)\}_{k=0}^{\infty}$ defined by

$$\Gamma(k) := \gamma(2k) + \gamma(2k+1) \qquad k = 0, 1, 2, ... \tag{5}$$

More specifically, to impose such constraints, first the initial positive sequence estimator is obtained by truncating the empirical $\hat{\Gamma}_M(k) = \tilde{r}_M(2k) + \tilde{r}_M(2k+1)$ sequence at the first k such that $\hat{\Gamma}_M(k) < 0$, to obtain $\hat{\Gamma}_M^{(\text{pos})} = \{\hat{\Gamma}_M(k)\}_{k=0}^{T-1}$ where $T := \min\{k \in \mathbb{N}; \hat{\Gamma}(k) < 0\}$. The argument given in Geyer [1992] for truncating the sequence at T is that T is the estimated time point when the autocovariance curve falls below the noise level. In addition to the initial positive sequence estimator, Geyer [1992] introduces the initial monotone sequence and initial convex sequence estimators. The initial monotone sequence and initial convex sequence estimators can then be calculated by replacing each $\hat{\Gamma}^{(\text{pos})}(k)$ with the minimum of the preceding ones and with the kth element of the greatest convex minorant

of the initial positive sequence, respectively.

Despite their simplicity, initial sequence estimators have very strong empirical performance and do not require the choice of a tuning parameter value, making them very useful in practice. For example, the widely used Stan software [Stan Development Team, 2019] employs the initial sequence estimators to estimate the effective sample size of Markov chain simulations. However, the statistical guarantees of the initial sequence estimators are somewhat lacking compared to the batch means and spectral variance estimators. To our knowledge, the only consistency results for the initial sequence estimators are that the initial sequence estimates asymptotically do not underestimate σ^2, that is,

$$\liminf_{M\to\infty} \widehat{\sigma^2}_{M,\text{init}} \geq \sigma^2 \ a.s., \tag{6}$$

as in Geyer [1992], Kosorok [2000], Brooks et al. [2011], and Dai and Jones [2017], rather than $\lim_{M\to\infty} \widehat{\sigma^2}_{M,\text{init}} = \sigma^2$ almost surely.

1.2 Review on estimation with shape constraints and connection to autocovariance sequence estimation

The work of Geyer [1992] can be viewed as an example of shape constrained inference, where the sequence $\{\Gamma_k\}_{k\in\mathbb{N}}$ is estimated in such a way that various shape constraints (positivity, monotonicity, and convexity) are enforced. Shape constrained inference has a long history in statistics. One of the standard examples is the isotonic regression, where in the most basic scenario one observes n independent random variables Y_i which are assumed to be noisy observations of some monotone increasing signal, i.e., $E[Y_1] \leq E[Y_2] \leq \ldots E[Y_n]$. The goal is to estimate the underlying n-dimensional signal [Barlow et al., 1972, Robertson, 1988]. However, shape constrained inference is not limited to the estimation of a finite dimensional vector and to monotonicity constraints. In fact, shape constrained inference has also been applied to infinite-dimensional problems where the quantity of interest is an infinite-dimensional vector or a function on $\mathbb{R}$ with different shape constraints. Examples include nonparametric estimation of monotone sequences or functions, the estimation of a convex or log-convex density, etc. [Grenander, 1956, Jankowski and Wellner, 2009, Dümbgen and Rufibach, 2011, Balabdaoui and Durot, 2015, Kuchibhotla et al., 2021].

Among such constraints, k-monotonicity, which is a refinement of the monotonicity property, has been studied by several authors [Balabdaoui and Wellner, 2007, Lefèvre and

Loisel, 2013, Durot et al., 2015, Chee and Wang, 2016, Giguelay, 2017]. A sequence m is called a k-monotone decreasing sequence if its successive differences up to order k are alternatively nonnegative and nonpositive, i.e.,

$$(-1)^n \Delta^n m(j) \geq 0 \text{ for } \; j \in \mathbb{N}, \; n = 0, \ldots, k \tag{7}$$

where $\Delta^n m(j) = \Delta^{n-1} m(j+1) - \Delta^{n-1} m(j)$ is a difference operator with $\Delta^0 m = m$. The case of $k = 0$ corresponds to nonnegativity, so that $(-1)^0 \Delta^0 m(j) = m(j) \geq 0$. The case $k = 1$ corresponds to monotonicity $m(j+1) - m(j) \leq 0$ in addition to nonnegativity, and $k = 2$ corresponds to convexity $m(j+2) - m(j+1) \geq m(j+1) - m(j)$ in addition to nonnegativity and monotonicity.

When $(-1)^n \Delta^n m(j) \geq 0$ for all $j, n \in \mathbb{N}$, the sequence m is called completely monotone. For functions on the real line, analogous versions of complete monotonicity involving derivatives rather than differences have been considered. Complete monotonicity conditions have been investigated by various authors. One prominent feature of prior results is an equivalence between satisfying a complete monotonicity constraint and admitting a mixture representation. For instance, Hausdorff [1921] proved that a sequence m is completely monotone if and only if the sequence m admits a moment representation, namely, if there exists a nonnegative measure F supported on $[0, 1]$ such that $m(k)$ is the kth moment of F, i.e., $m(k) = \int x^k F(dx)$. Similarly, completely monotone functions on $\mathbb{R}^+ \cup \{0\}$ can be represented as a scale mixture of exponentials [Feller, 1939, Jewell, 1982], and a completely monotone probability mass function (pmf) can be represented as a mixture of geometric pmfs [Steutel, 1969]. The latter fact was used in the recent work by Balabdaoui and de Fournas-Labrosse [2020] for the estimation of a completely monotone pmf using nonparametric least squares estimation.

In the context of asymptotic variance estimation, the result of Geyer [1992] on the Γ sequence can be refined using the concept of complete monotonicity. Recall that Geyer [1992] showed that the sequence Γ obtained as the rolling sum of γ with window size 2, i.e., $\Gamma(k) = \gamma(2k) + \gamma(2k+1)$, is 2-monotone. In this paper, we show that the Γ sequence is not only 2-monotone but completely monotone (Proposition 1). This suggests that higher order shape structure could be exploited in the estimation of $\Gamma(k)$ and, consequently, the asymptotic variance. However, while the Γ sequence is completely monotone, the set of completely monotone sequences is not entirely satisfactory to work with for our purpose of

estimating the *entire* autocovariance sequence γ, since γ may not be a completely monotone sequence.

Our contribution and organization of the paper To our knowledge, this is the first work in which the moment representation of the autocovariance sequence (1) is directly exploited in this manner to carry out shape-constrained inference for the estimation of the autocovariance sequence and asymptotic variance. Our work is the first to use shape-constrained inference methods to provide a provably consistent estimator for the asymptotic variance for a Markov chain. The work of Balabdaoui and de Fournas-Labrosse [2020] on estimating a completely monotone pmf is the most similar to ours of which we are aware. However, Balabdaoui and de Fournas-Labrosse [2020] consider a substantially different setting involving the estimation of a completely monotone probability mass function (pmf) from iid samples. In our setting, the dependence between observations necessitates the use of different tools for the statistical analysis. To the best of our knowledge, this is the first work in which shape-constrained inference is used to alter the convergence property of input sequences as well.

The remainder of the paper is organized as follows. In Section 2, we introduce background on Markov chains and prove Proposition 1 on the representation of γ and Γ as moment sequences. In Section 3, we introduce our proposed estimator, the moment least squares estimator, and study some basic properties of the proposed estimator. In Section 4, we provide statistical convergence results for the moment least squares estimator. In particular, we prove the almost sure convergence in the ℓ_2 norm of the estimated autocovariance sequence (Theorem 2), the almost sure vague convergence of the representing measure for the moment least squares estimator to the representing measure for γ (Proposition 10), and the almost sure convergence of the estimated asymptotic variance (Theorem 3). In Section 5, we show the results of our empirical study, in which the moment least squares estimator performs well relative to other state-of-the-art estimators for MCMC asymptotic variance and autocovariance sequence estimation.

2 Problem set-up

We now describe our setup in detail and fix some notation. We consider a ψ-irreducible, aperiodic Markov chain $X = \{X_t\}_{t=0}^{\infty}$ evolving over t on a measurable space $(\mathsf{X}, \mathscr{X})$, where

the state space X is a complete separable metric space and $\mathscr{X}$ is the associated Borel σ-algebra. We let π denote a probability measure defined on $(\mathsf{X}, \mathscr{X})$ with respect to which we would like to compute expectations. We use $g : \mathsf{X} \to \mathbb{R}$ to denote a function for which it is of interest to obtain $\mu = \int g(x)\pi(dx)$. We define a transition kernel as a function $Q : \mathsf{X} \times \mathscr{X} \to [0, 1]$ such that $Q(\cdot, A) : \mathsf{X} \to [0, 1]$ is an $\mathscr{X}$-measurable function for each $A \in \mathscr{X}$ and $Q(x, \cdot) : \mathscr{X} \to [0, 1]$ is a probability measure on $(\mathsf{X}, \mathscr{X})$ for each $x \in \mathsf{X}$. For a probability measure π on $(\mathsf{X}, \mathscr{X})$, a probability kernel Q is said to be π-stationary if $\pi(A) = \int Q(x, A)\pi(dx)$ for all $A \in \mathscr{X}$. An initial measure ν on $\mathcal{X}$ and a transition kernel Q define a Markov chain probability measure P_ν for $X = (X_0, X_1, X_2, \dots)$ on the canonical sequence space $(\Omega, \mathcal{F})$. We write E_ν to denote expectation with respect to P_ν.

For a function $f : \mathsf{X} \to \mathbb{R}$ and a transition kernel Q, we define the linear operator Q by

$$Qf(x) = \int Q(x, dy)f(y) \tag{8}$$

We define $Q^0 f(x) = f(x)$, $Q^1 f(x) = Qf(x)$, and $Q^t f(x) = Q(Q^{t-1} f)(x)$ for $t > 1$, and we define $Q^t(x, A) = Q^t I_A(x)$, where $I_A(\cdot)$ is the indicator function for the set A. We let $L^2(\pi)$ be the space of functions which are square integrable with respect to π, i.e., $L_2(\pi) = \{f : \mathsf{X} \to \mathbb{R}; \int f(x)^2\pi(dx) < \infty\}$. For functions $f, g \in L^2(\pi)$, we define an inner product

$$\langle f, g\rangle_\pi = \int f(x)g(x)\pi(dx). \tag{9}$$

We note that $L^2(\pi)$ is a Hilbert space equipped with the inner product (9). For $f \in L^2(\pi)$, we define $\|f\|_{L^2(\pi)} = \sqrt{\langle f, f\rangle_\pi}$. Also, for an operator Q on $L^2(\pi)$, we define $\|Q\|_{L^2(\pi)} = \sup_{f;\|f\|_{L^2(\pi)} \le 1} \|Qf\|_{L^2(\pi)}$ and we say Q is bounded if $\|Q\|_{L^2(\pi)} < \infty$.

We say that a transition kernel Q satisfies the reversibility property with respect to π if

$$\langle f_1, Qf_2\rangle_\pi = \langle Qf_1, f_2\rangle_\pi \tag{10}$$

for any functions $f_1, f_2 \in L^2(\pi)$, i.e., if Q is a self-adjoint operator. Reversibility with respect to π is a sufficient condition for π-stationarity of Q, since for a reversible transition

kernel Q, we have

$$\pi(A) = \int I_A(x) Q I_X(x) \pi(dx) = \int I_X(x) Q I_A(x) \pi(dx).$$

The spectrum of the operator Q plays a key role in determining the mixing properties of a Markov chain with transition kernel Q. Recall that for an operator T on the Hilbert space $L^2(\pi)$, the spectrum of T is defined as

$$\sigma(T) = \{\lambda \in \mathbb{C}; (T - \lambda I)^{-1} \text{does not exist or is unbounded } \}. \tag{11}$$

For Markov operators Q, we define the spectral gap δ of Q as $\delta = 1 - \sup\{|\lambda|; \lambda \in \sigma(Q_0)\}$ where Q_0 is defined as

$$Q_0 f = Qf - E_\pi[f(X_0)] f_0 \tag{12}$$

and $f_0 \in L^2(\pi)$ is the constant function such that $f_0(x) = 1$ for all $x \in \mathsf{X}$. It is easy to check that Q_0 is self-adjoint and bounded. If Q is reversible, Q has a positive spectral gap ($\delta > 0$) if and only if the chain is geometrically ergodic [Roberts and Rosenthal, 1997, Kontoyiannis and Meyn, 2012]. In addition, $(1-\delta)^k$ is the maximal lag k correlation of any two functions, and therefore for any function f and $Y_{fM} = M^{-1} \sum_{t=0}^{M-1} f(X_t)$, the asymptotic variance σ_f^2 of $\sqrt{M}(Y_{fM} - E_\pi[f(X_0)])$ is bounded above by

$$\sigma_f^2 = \gamma_f(0) + 2 \sum_{k \geq 1} \gamma_f(k) \leq \gamma_f(0) + 2 \sum_{k \geq 1} (1-\delta)^k \gamma_f(0) = \frac{2 - \delta}{\delta} \gamma_f(0)$$

where $\gamma_f(k) = \mathrm{Cov}_\pi(f(X_0), f(X_k))$.

In the remainder, we consider a discrete time Markov chain $X = \{X_t\}_{t=0}^\infty$ with stationary distribution π and π-reversible transition kernel Q with a positive spectral gap. We let g be a square integrable function with respect to π, and use $\gamma(k)$ defined by

$$\gamma(k) = \mathrm{Cov}_\pi\{g(X_0), g(X_{|k|})\} = \langle g, Q_0^{|k|} g \rangle_\pi \quad \text{for } k \in \mathbb{Z}$$

to denote the lag k autocovariance of the stationary time series $\{g(X_t)\}_{t=0}^\infty$ obtained with $X_0 \sim \pi$. We use $\gamma = \{\gamma(k)\}_{k \in \mathbb{Z}}$ to denote the autocovariance sequence on $\mathbb{Z}$. We summarize our assumptions on the Markov chain X as follows for future reference:

(A.1) (Harris ergodicity) X is ψ-irreducible, aperiodic, and positive Harris recurrent.

(A.2) (Reversibility) The transition kernel Q is π-reversible for a probability measure π on $(\mathsf{X}, \mathscr{X})$.

(A.3) (Geometric ergodicity) There exists a real number $\rho < 1$ and a non-negative function M on the state space X such that

$$\|Q^n(x, \cdot) - \pi(\cdot)\|_{\mathrm{TV}} \leq M(x)\rho^n, \text{ for all } x \in \mathsf{X},$$

where $\|\cdot\|_{\mathrm{TV}}$ is the total variation norm.

Throughout the paper, we assume that the function of interest $g : \mathsf{X} \to \mathbb{R}$ is in $L^2(\pi)$, i.e,

(B.1) (Square integrability) $\int g(x)^2\pi(dx) < \infty$.

For the definitions of ψ-irreducibility, aperiodicity, and positive Harris recurrence, see e.g., Meyn and Tweedie [2009]. Reversibility is a key requirement for our estimator because it allows us to use the shape constraints implied by the spectral decomposition of the Markov chain kernel (see Proposition 1). Many practical transition kernels satisfy π-reversibility. Notably, all Metropolis-Hastings transition kernels satisfy reversibility. Additionally, all Gibbs component update kernels are reversible. As noted by a referee, in practice, it is common to combine a set of reversible transition kernels $\{Q_k\}_{k=1}^K$, such as those from Metropolis-Hastings or Gibbs updates, to form a joint transition mechanism Q. The reversibility of the combined mechanism Q depends on the way in which the individual kernels Q_k are combined. For example, in deterministic scan sampling, where each update consists of sequentially applying Q_k, $k = 1, \ldots, K$, the resulting kernel $Q(x, A) = Q_1Q_2 \cdots Q_K(x, A)$ is generally non-reversible. On the other hand, there are schemes for combining reversible kernels Q_k in such a way that the resulting Q is reversible. For example, the random scan transition kernel $Q = K^{-1}\sum_{k=1}^K Q_k$, corresponding to randomly selecting the transition kernel at each iteration, is reversible. Additionally, random permutation scans, in which at each iteration the reversible Q_k are composed in a randomly permuted order, and palindromic scan updates, in which $Q = Q_1 \ldots Q_{K-1}Q_KQ_{K-1} \ldots Q_1$, lead to reversible Markov chains [see, e.g., page 376 of Robert and Casella, 2004]. Finally, we note that in data augmentation Gibbs sampling, the marginal chains are reversible [see, e.g., Liu et al., 1994, Robert and Casella, 2004].

Geometric ergodicity implies exponential convergence of the Markov chain X to its target distribution π. When the state space X is finite, all irreducible and aperiodic Markov chains are geometrically ergodic. While this is no longer true for infinite state space, geometric ergodicity remains a theoretically and practically important condition for Markov chains [e.g. Roberts and Rosenthal, 1998, Jones and Qin, 2022]. For example, geometric ergodicity provides one of the simplest sufficient conditions for the Markov chain central limit theorem (CLT) to hold. In fact, for a reversible geometrically ergodic Markov chain, a finite second moment of the function of interest g is sufficient to establish a CLT (e.g., Jones, 2004). The establishment of geometric ergodicity is usually done on a case-by-case analysis, and many works have studied geometric ergodicity of popular samplers (e.g., Mengersen and Tweedie, 1996, Roberts and Tweedie, 1996, Jarner and Hansen, 2000, Jarner and Tweedie, 2003, Johnson and Geyer, 2012, Chakraborty and Khare, 2017, Livingstone et al., 2019, Durmus et al., 2023).

The following proposition shows that both the autocovariance sequence γ and rolling sum Γ of the sequence γ with a window size of 2 from a reversible chain have the following *moment* representations, namely there exist measures F and G supported on $[-1, 1]$ and $[0, 1]$, respectively, such that $\gamma(k)$ and $\Gamma(k)$ are the kth moments of F and G, respectively. Let $\mathcal{M}_{\mathbb{R}}$ denote the set of finite regular measures on $\mathbb{R}$.

Proposition 1. *Assume (A.2) and (B.1).*

1. *The true autocovariance sequence $\gamma(k) = \langle g, Q_0^k g \rangle_\pi$, $k \in \mathbb{Z}$, has the following representation for some $F \in \mathcal{M}_{\mathbb{R}}$*

$$\gamma(k) = \int_{\sigma(Q_0)} x^{|k|} F(dx), \tag{13}$$

where $\sigma(Q_0)$ is the spectrum of the linear operator Q_0 defined as in (12). Moreover, $\sigma(Q_0)$ lies on the real axis, and $\sigma(Q_0) \subseteq [-1, 1]$.

2. *The sequence $\Gamma = \{\Gamma(k)\}_{k \in \mathbb{N}}$ defined by $\Gamma(k) = \gamma(2k) + \gamma(2k+1)$, $k \in \mathbb{N}$, has the following representation for some $G \in \mathcal{M}_{\mathbb{R}}$*

$$\Gamma(k) = \int_{\sigma(Q_0^2)} x^k G(dx), \tag{14}$$

and $\sigma(Q_0^2) \subseteq [0, 1]$.

3. *If we additionally assume (A.1) and (A.3) in addition to (A.2) and (B.1), there exists* $0 < \delta_0 \leq 1$ *such that* $\sigma(Q_0) \subseteq [-1+\delta_0, 1-\delta_0]$ *and* $\sigma(Q_0^2) \subseteq [0, (1-\delta_0)^2]$.

The proof of Proposition 1 is deferred to Supplementary Material S3.1 [Berg and Song, 2023]. In the example below, the moment representation of the autocovariance sequence from a reversible Markov chain is illustrated using an AR(1) chain.

Example 2.1. *(Autoregressive chain example) Consider an AR(1) autoregressive process with* $X_{t+1} = \rho X_t + \epsilon_{t+1}$, $t = 0, 1, 2, \ldots$, *where* $\epsilon_t \overset{iid}{\sim} N(0, \tau^2)$ *and* $\rho \in (-1, 1)$. *The stationary measure* π *for the* X_t *chain is the measure corresponding to a* $N(0, \tau^2/(1-\rho^2))$ *random variable, and the* X_t *chain can be shown to be reversible with respect to* π. *Consider the autocovariance sequence* $\gamma(k) = E_\pi[\bar{g}(X_0)\bar{g}(X_{|k|})]$ *with the identity function* $g(x) = x$. *Since* $E_\pi[g(X_0)] = 0$, *we have*

$$\gamma(k) = \mathrm{Cov}_\pi(g(X_0), g(X_k)) = \mathrm{Var}_\pi(g(X_0))\rho^{|k|} = \frac{\tau^2}{1-\rho^2}\rho^{|k|}.$$

Then $\gamma(k)$ *can be represented as* $\gamma(k) = \int x^{|k|} F(dx)$ *for all* $k \in \mathbb{Z}$ *by letting* $F = \frac{\tau^2}{1-\rho^2}\delta_\rho$, *where* δ_ρ *denotes a unit point mass measure at* ρ.

We note that the second statement of Proposition 1 implies that the $\Gamma(k)$ sequence is *completely monotone*, and therefore is a refinement of the result in Geyer [1992] which showed that $\Gamma(k)$ is 2-monotone. This is due to the theorem of Hausdorff [1921] below, in which an equivalence is shown between $[0,1]$-*moment* sequences (sequences with the representation $m(k) = \int x^k F(dx)$ for some F with $\mathrm{Supp}(F) \subseteq [0,1]$; see Definition 1 for the formal definition) and *completely monotone* sequences satisfying inequalities (7) for all $k, n \in \mathbb{N}$. The relationship between sequences admitting certain moment representations and their shape constraints will be further explored in the following Section 3.

Theorem 1 (Hausdorff moment theorem [Hausdorff, 1921])**.** *There exists a representing measure* μ *supported on* $[0,1]$ *for* $m \in \mathbb{R}^{\mathbb{N}}$ *if and only if* m *is a completely monotone sequence. Additionally, if* m *is a completely monotone sequence, the representing measure* μ *for* m *is unique.*

We have from Proposition 1 that γ is a $[-1,1]$-moment sequence. In general, γ is not a completely monotone sequence as its representing measure can have mass in $[-1, 0)$. A simple example is the autocovariance sequence from an AR(1) stationary chain with a

negative AR(1) coefficient. The autocovariances oscillate between positive and negative values as $k \to \infty$ and therefore cannot decrease monotonically.

Notations

We let $\mathbb{N}$ be the set of non-negative integers $\{0, 1, 2, ...\}$ and $\mathbb{Z}$ the set of integers $\{..., -1, 0, 1, ...\}$. For a sequence m on $\mathbb{N}$ or $\mathbb{Z}$, we define an ℓ_p norm for m by $\|m\|_p = (\sum_k |m(k)|^p)^{1/p}$ for $p = 1, 2, \ldots$, and $\|m\|_\infty = \max_k |m(k)|$. In addition, when $p = 2$, we omit the subscript and write $\|\cdot\| = \|\cdot\|_2$. We use $\ell_p(\mathbb{N})$ (or $\ell_p(\mathbb{Z})$) to denote the space of sequences on $\mathbb{N}$ (or $\mathbb{Z}$) with finite ℓ_p norms. In particular, $\ell_1(\mathbb{Z})$ is the space of absolutely summable sequences on $\mathbb{Z}$, i.e., $\ell_1(\mathbb{Z}) = \{m \in \mathbb{R}^{\mathbb{Z}}; \sum_{k=-\infty}^{\infty} |m(k)| < \infty\}$ and $\ell_2(\mathbb{Z})$ is the space of square summable sequences on $\mathbb{Z}$, i.e., $\ell_2(\mathbb{Z}) = \{m \in \mathbb{R}^{\mathbb{Z}}; \sum_{k=-\infty}^{\infty} m^2(k) < \infty\}$. We equip $\ell_2(\mathbb{Z})$ with a usual inner product $\langle x, y\rangle = \sum_{k=-\infty}^{\infty} x(k)y(k)$ for $x, y \in \ell_2(\mathbb{Z})$. Then $\|x\| = \sqrt{\langle x, x\rangle} = \|x\|_2$. Also, for $\alpha \in [-1, 1]$, we define $x_\alpha = \{x_\alpha(k)\}_{k\in\mathbb{Z}}$ such that $x_\alpha(k) = \alpha^{|k|}$ for $k \in \mathbb{Z}$. Note that for $\alpha \in (-1, 1)$, $x_\alpha \in \ell_2(\mathbb{Z})$. Finally, for a measure μ, we let $\mathrm{Supp}(\mu)$ denote the support of μ.

3 Moment least squares estimator (Moment LSE)

We now introduce the moment least squares estimator. We first formally define moment sequences and moment spaces.

Definition 1 (moment sequence and representing measure)**.** *We say that a sequence m is an $[a, b]$-moment sequence if there exists a positive measure μ supported on $[a, b]$ for some $-\infty < a \le b < \infty$ such that the equation*

$$m(k) = \int x^{|k|}\mu(dx) \tag{15}$$

holds for any $k \in \mathbb{N}$ (if $m = \{m(k)\}_{k=0}^{\infty}$ is a sequence defined on $\mathbb{N}$) or any $k \in \mathbb{Z}$ (if $m = \{m(k)\}_{k=-\infty}^{\infty}$ is a sequence defined on $\mathbb{Z}$). We say that μ is a representing measure for the sequence m.

For a closed set $C \subseteq \mathbb{R}$, we write $\mathscr{M}_\infty(C)$ to denote the the set of sequences on $\mathbb{R}^{\mathbb{Z}}$ with a moment representation with a measure supported on C. For example, $\mathscr{M}_\infty([a, b])$ is the set of $[a, b]$-moment sequences. By definition, we have $\mathscr{M}_\infty(I_1) \subseteq \mathscr{M}_\infty(I_2)$ if $I_1 \subseteq I_2$ for two closed intervals $I_1, I_2 \subseteq \mathbb{R}$. The support $[a, b]$ has a close relationship with the shape

constraints satisfied by sequences $m \in \mathscr{M}_\infty([a,b])$. When $[a,b] = [0,1]$, $\mathscr{M}_\infty([0,1])$ is the space of completely monotone sequences. In general, the true autocovariance γ does not belong to $\mathscr{M}_\infty([0,1])$, but does belong to $\mathscr{M}_\infty([-1,1])$. Additionally, for a geometrically ergodic chain, Proposition 1 shows $\gamma \in \mathscr{M}_\infty([-1+\delta, 1-\delta])$ for any $\delta \geq 0$ such that $\delta \leq \delta_0$, where δ_0 is the spectral gap of Q in Proposition 1. Throughout the remainder of the paper, we will consider projections onto the set $\mathscr{M}_\infty([-1+\delta, 1-\delta])$, and thus we let $\mathscr{M}_\infty(\delta) = \mathscr{M}_\infty([-1+\delta, 1-\delta])$ for notational simplicity.

Now we define the moment least squares estimator $\Pi_\delta(r_M)$ resulting from an initial autocovariance sequence estimator $r_M \in \ell_2(\mathbb{Z})$ by

$$\begin{aligned} \Pi_\delta(r_M) &= \underset{m \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})}{\arg\min} \|r_M - m\|^2 \\ &= \underset{m \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})}{\arg\min} \sum_{k \in \mathbb{Z}} \{r_M(k) - m(k)\}^2. \end{aligned} \tag{16}$$

Note that $\Pi_\delta(r_M)$ is the closest *moment* sequence with respect to some measure supported on $[-1+\delta, 1-\delta]$ to the input autocovariance sequence r_M, with respect to the ℓ_2 norm $\|\cdot\|$ on $\ell_2(\mathbb{Z})$. This optimization problem can be formulated as a convex quadratic problem, which we discuss further in Section 3.3.

The optimization problem (16) has one hyperparameter δ, which needs be chosen sufficiently small so that the true autocovariance sequence γ is a feasible solution, in the sense that $\gamma \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$, of the optimization problem (16). Note that any value of δ such that $0 \leq \delta \leq \delta_\gamma$ makes γ feasible for $\delta_\gamma = 1 - \sup\{|x|; x \in \mathrm{Supp}(F)\}$ where F is the representing measure for γ. Empirically, choosing δ as large as possible subject to $\delta \leq \delta_\gamma$ leads to better performance because, roughly speaking, larger δ corresponds to more shape regularization. However, the method appears to work for a wide range of δ as long as δ is chosen to be positive (see Section 5 for details). We also propose a practical choice of δ in Section 5. Theoretically, we showed the consistency of the proposed estimator $\Pi_\delta(r_M)$ for any $0 < \delta \leq \delta_\gamma$.

For the choice of the initial autocovariance sequence estimator, any estimator r_M from a Markov chain sample $X_0, X_1, ..., X_{M-1}$ of size M satisfying

(R.1) (a.s. elementwise convergence) $r_M(k) \underset{M\to\infty}{\to} \gamma(k)$ for each $k \in \mathbb{Z}$, P_x-almost surely, for any initial condition $x \in \mathsf{X}$,

(R.2) (finite support) $r_M(k) = 0$ for $k \geq n(M)$ for some $n(M) < \infty$, and

(R.3) (even function with a peak at 0) $r_M(k) = r_M(-k)$ and $r_M(0) \geq |r_M(k)|$ for each $k \in \mathbb{Z}$,

is allowed. As we demonstrate in Proposition 7, the empirical autocovariance sequence $\tilde{r}_M$ satisfies assumptions (R.1)–(R.3). In addition, (R.1)–(R.3) are satisfied by any windowed empirical autocovariance sequence $\check{r}_M$ of the form $\check{r}_M(k) = \tilde{r}_M(k)w_M(|k|)$, where $w_M(k)$ is any window function which meets the following conditions (W.1) - (W.3):

(W.1) $w_M(0) = 1$ for all $M \in \mathbb{N}$,

(W.2) $|w_M(k)| \leq 1$ for all $k \in \mathbb{N}$ and $M \in \mathbb{N}$,

(W.3) $w_M(k) \to 1$ for any fixed k as $M \to \infty$

In particular, conditions (W.1) - (W.3) are satisfied for some widely used window functions such as the simple truncation window $w_M(k) = I(k < b_M)$ and the Parzen window function $w_M(k) = [1 - k^q/b_M^q]I(k < b_M)$ for $q \in \{1, 2, 3, \dots\}$, which is the modified Bartlett window when $q = 1$.

In the following subsection, we provide some results relating to moment sequences, and provide an alternative characterization of moment sequences in relation to complete monotonicity.

3.1 Characterization of $[a, b]$-moment sequences

While γ is not completely monotone when the support of the representing measure for γ is not contained in $[0, 1]$, it still exhibits infinitely many constraints. Previous studies have provided characterizations of $[a, b]$-moment sequences [Krein and Nudelman, 1977, Chandler, 1988]. Specifically, an $[a, b]$-moment sequence m can be characterized equivalently by the non-negativity of a specific family of quadratic forms derived from m, a, and b (e.g., Theorem 3.13 in Schmüdgen, 2017).

In Proposition 2, we present an alternative characterization for an $[a, b]$-moment sequence m in terms of the complete monotonicity of a transformed sequence $T(m; [a, b])$. It is important to note that while Proposition 2 gives insights on which (infinite number of) constraints are imposed on an estimator at the sequence level by requiring the estimator to be in the $[a, b]$-moment space $\mathscr{M}_\infty([a, b])$, the actual enforcement of these constraints is achieved through a mixture representation as in (15). It is also technically convenient to have this alternative characterization for $[a, b]$-moment sequences because we can, e.g.,

verify that a sequence is an $[a,b]$-moment sequence by checking whether $T(m;[a,b])$ is completely monotone, and guarantee the uniqueness of the representing measures of $[a,b]$-moment sequences based on Theorem 1.

For a sequence $m = \{m(k)\}_{k=0}^{\infty}$ and constants $a < b$, we define $T : \mathbb{R}^{\mathbb{N}} \to \mathbb{R}^{\mathbb{N}}$ as follows:

$$T(m;a,b)(k) = (b-a)^{-k} \sum_{i=0}^{k} \binom{k}{i} m(i)(-a)^{k-i}, \quad k = 0,1,2,\ldots. \tag{17}$$

Note $T(m;a,b)(0) = m(0)$, and when $a = 0, b = 1$, we have $T(m;0,1) = m$.

Proposition 2 ($[a,b]$-moment sequences)**.** *For a sequence $m = \{m(k)\}_{k=0}^{\infty}$ and $a < b$, there exists a representing measure μ for m supported on $[a,b]$ if and only if the sequence $T(m;a,b)$ is completely monotone. Additionally, if $T(m;a,b)$ is completely monotone, then the representing measure for m is unique.*

The proof of Proposition 2 is deferred to Supplementary Material S4.1 [Berg and Song, 2023]. Since throughout this paper we will consider sequences $m = \{m(k)\}_{k=-\infty}^{\infty}$ satisfying the symmetry relation $m(k) = m(-k)$ for each $k \in \mathbb{Z}$, we state the following corollary.

Corollary 1. *Consider a sequence $m = \{m(k)\}_{k\in\mathbb{Z}}$ which is symmetric around 0, i.e., $m(k) = m(-k)$ for $k \in \mathbb{Z}$. Additionally, consider $a, b \in \mathbb{R}$ with $a < b$. Then there exists a measure μ supported on $[a,b]$ such that $m(k) = \int x^{|k|}\mu(dx)$ for all $k \in \mathbb{Z}$ if and only if the sequence $T(\{m(k)\}_{k\in\mathbb{N}};a,b)$ is completely monotone. Additionally, if $T(\{m(k)\}_{k\in\mathbb{N}};a,b)$ is completely monotone, then the measure corresponding to m is unique.*

3.2 Properties of the moment least squares estimator

The moment least squares estimator (moment LSE) $\Pi_\delta(r_M)$ from an initial autocovariance sequence estimator r_M involves a projection from $\ell_2(\mathbb{Z})$ to $\mathscr{M}_\infty([-1+\delta, 1-\delta]) \cap \ell_2(\mathbb{Z})$. In this section, we show the existence and uniqueness of projections from $\ell_2(\mathbb{Z})$ to a moment sequence space $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$ where $C \subseteq [-1,1]$ is a closed set. For an $r \in \ell_2(\mathbb{Z})$, define $\Pi(r;C)$ be the projection of r onto $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$. We present a variational characterization of the projection $\Pi(r;C)$. Finally, we obtain results on the properties of the representing measure of $\Pi(r;C)$. Namely, we show that for fixed sample size M, if $r(k) = 0$ for $k \geq n(M)$ for some $n(M) < \infty$, the representing measure $\hat{\mu}_C$ corresponding to $\Pi(r;C)$ is discrete, with finite support set $\mathrm{Supp}(\hat{\mu}_C)$ having cardinality $|\mathrm{Supp}(\hat{\mu}_C)| \leq n_0$, where n_0 is the smallest even number with $n_0 > (n(M)-1)$. Similar discreteness and finite support set

results appear in the setting of nonparametric maximum likelihood estimation for mixture models, as in Lindsay [1983], as well as in the least-squares estimation of a k-monotone or completely monotone pmf as in Giguelay [2017] and Balabdaoui and de Fournas-Labrosse [2020].

First of all, for any closed $C \subset [-1,1]$, we show that $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$ is a closed and convex subset of $\ell_2(\mathbb{Z})$. Then, since $\ell_2(\mathbb{Z})$ is a Hilbert space equipped with the inner product $\langle v, u\rangle = \sum_{k\in\mathbb{Z}} v(k)u(k)$, we obtain by the Hilbert space projection theorem the existence and uniqueness of projections from $\ell_2(\mathbb{Z})$ to $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$.

Proposition 3. *For any closed $C \subseteq [-1,1]$, the set $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$ is a closed, convex subset of $\ell_2(\mathbb{Z})$. In particular, for any given vector $r \in \ell_2(\mathbb{Z})$, $\Pi(r; C)$ exists and is unique in $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$.*

Note that for any M, an initial input autocovariance sequence r_M satisfying (R.1) -(R.3) is in $\ell_2(\mathbb{Z})$ since $r_M(k) = 0$ for $|k| \geq n(M)$, and therefore, the moment LSE $\Pi_\delta(r_M) = \Pi(r_M; [-1+\delta, 1-\delta])$ is well defined. In addition, the optimization problem (16) is convex. The proof of Proposition 3 uses the alternative characterization in Corollary 1 of an $[a,b]$-moment sequence and is deferred to Supplementary Material S4.2 [Berg and Song, 2023].

Next, we present a few results regarding the projection $\Pi(r; C)$ of $r \in \ell_2(\mathbb{Z})$ onto $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$. Proposition 4 provides a variational characterization of the projection $\Pi(r; C)$.

Proposition 4. *Let C be a closed subset of $[-1,1]$, and suppose $r \in \ell_2(\mathbb{Z})$. Then for $f \in \mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$, we have $f = \Pi(r; C)$ if and only if*

1. *for all $\alpha \in C \cap (-1,1)$, $\langle f, x_\alpha\rangle \geq \langle r, x_\alpha\rangle$, i.e.,*

$$\sum_{k=-\infty}^{\infty} f(k)\alpha^{|k|} \geq \sum_{k=-\infty}^{\infty} r(k)\alpha^{|k|}, \tag{18}$$

2. *$\langle f, f\rangle = \langle f, r\rangle$, i.e., $\sum_{k=-\infty}^{\infty} f(k)^2 = \sum_{k=-\infty}^{\infty} f(k)r(k)$.*

A similar characterization of $\Pi(r; C)$ was also presented in Balabdaoui and de Fournas-Labrosse [2020]. We omit the proof as the result can be obtained by a minor modification of Proposition 2.2 in Balabdaoui and de Fournas-Labrosse [2020].

Proposition 5 below shows that (18) holds with equality for α in the support of the representing measure for $\Pi(r; C)$ with $|\alpha| < 1$.

Proposition 5. *Let C be a closed subset of $[-1,1]$, and suppose $r \in \ell_2(\mathbb{Z})$. Let $\hat{\mu}_C$ denote the representing measure for $\Pi(r;C)$. Then for each $\alpha \in \mathrm{Supp}(\hat{\mu}_C) \cap (-1,1)$, we have*

$$\langle \Pi(r;C), x_\alpha \rangle = \langle r, x_\alpha \rangle .$$

The proof for Proposition 5 essentially follows from Proposition 4, as we have $\langle \Pi(r;C) - r, x_\alpha \rangle \geq 0$ for all $\alpha \in C \cap (-1,1)$ and from the second condition in Proposition 4

$$\int \langle \Pi(r;C) - r, x_\alpha \rangle \, \hat{\mu}_C(d\alpha) = \langle \Pi(r;C), \Pi(r;C) \rangle - \langle r, \Pi(r;C) \rangle = 0,$$

which implies $\langle \Pi(r;C) - r, x_\alpha \rangle = 0$, for $\hat{\mu}_C$-almost every α. We show that this implies that $\langle \Pi(r;C) - r, x_\alpha \rangle = 0$ for all $\alpha \in \mathrm{Supp}(\hat{\mu}_C) \cap (-1,1)$. The details are deferred to Supplementary Material S4.3 [Berg and Song, 2023].

Finally, we show that for an input sequence r with finite support, i.e., $r(k) = 0$ for $|k| \geq M$ for some M, then the representing measure for the projection $\Pi(r;C)$ is discrete, and the support of the representing measure contains at most a finite number of points. More concretely, we have the following result:

Proposition 6. *Let C be a closed subset of $[-1,1]$, and suppose $r \in \ell_2(\mathbb{Z})$ satisfies $r(k) = 0$ for all k with $|k| > M-1$ for $M < \infty$. Let $\Pi(r;C)$ denote the projection of r onto $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$. Let $\hat{\mu}_C$ denote the representing measure for $\Pi(r;C)$. Then $\mathrm{Supp}(\hat{\mu}_C)$ contains at most n points, where n is the smallest even number such that $n > (M-1)$. Additionally, the support of $\hat{\mu}_C$ is contained in $(-1,1)$, that is, $\mathrm{Supp}(\hat{\mu}_C) \cap \{-1,1\} = \emptyset$.*

The proof follows similar lines as in Balabdaoui and de Fournas-Labrosse [2020], but requires nontrivial modification to deal with the possible support of $\hat{\mu}_C$ in $[-1,0)$. We defer the proof to Supplementary Material S4.4 [Berg and Song, 2023]. In particular, a moment LSE $\Pi_\delta(r_M)$ for any initial estimator r_M satisfying condition (R.2) has a representing measure which is discrete and has support containing at most n_0 points, where n_0 is the smallest even number such that $n_0 > \{n(M)-1\}$. The representing measure for an arbitrary element of $f \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$ is in general neither finitely supported nor discrete. Thus Proposition 6 provides a considerable simplification of the form of the representing measure of $\Pi_\delta(r_M)$.

3.3 Computation of the moment least squares estimator

Recall that $\Pi_\delta(r_M)$ is the minimizer m of $\sum_{k\in\mathbb{Z}}\{r_M(k)-m(k)\}^2$ such that $m \in \mathscr{M}_\infty(\delta)\cap \ell_2(\mathbb{Z})$. By Proposition 6, since $\hat{\mu}_\delta$ is a discrete measure, we have

$$\Pi_\delta(r_M)(k) = \int \alpha^{|k|}\hat{\mu}_\delta(d\alpha) = \sum_{\alpha\in \mathrm{Supp}(\hat{\mu}_\delta)} \alpha^{|k|}\hat{\mu}_\delta(\{\alpha\}).$$

For a closed set $\Theta \subseteq [-1+\delta, 1-\delta]$, recall $\Pi(r;\Theta)$ is the projection of r to the set of $\ell_2(\mathbb{Z})$ moment sequences with representing measure supported on Θ. Note we have $\Pi_\delta(r) = \Pi(r;[-1+\delta,1-\delta]) = \Pi(r;\Theta_0)$ for any Θ_0 such that $\mathrm{Supp}(\hat{\mu}_\delta)\subseteq\Theta_0\subseteq[-1+\delta,1-\delta]$.

For a finite $\Theta=\{\alpha_1,\ldots,\alpha_s\}\subset(-1,1)$, $\Pi(r;\Theta)$ can be computed by solving a simple convex quadratic program. For $m\in\mathscr{M}_\infty(\Theta)\cap\ell_2(\mathbb{Z})$, the least squares objective in (16) becomes

$$\sum_{k\in\mathbb{Z}}(r_M(k)-m(k))^2 = \sum_{k\in\mathbb{Z}}(r_M(k)-\sum_{i=1}^{s}\alpha_i^{|k|}w_i)^2, \tag{19}$$

where we define $w_i=\mu_m(\{\alpha_i\})$ for $i=1,\ldots,s$ and $s=|\Theta|$ where μ_m denotes the representing measure for m. Define $\mathbf{w}=[w_1,\ldots,w_s]\in\mathbb{R}^s$. Define $\mathbf{a}=[a_1,\ldots,a_s]\in\mathbb{R}^s$ such that $a_i=\sum_{k\in\mathbb{Z}}\alpha_i^{|k|}r_M(k)=\sum_{k;r_M(k)\neq 0}\alpha_i^{|k|}r_M(k)$ and $\mathbf{B}\in\mathbb{R}^{s\times s}$ such that $\mathbf{B}_{ij}=\frac{1+\alpha_i\alpha_j}{1-\alpha_i\alpha_j}$. Note that $\mathbf{B}$ can be computed easily based on Θ and $\mathbf{a}$ can be computed easily based on Θ and r_M when r_M satisfies (R.2). Then with some algebra, we can show that $\sum_{k\in\mathbb{Z}}(r_M(k)-m(k))^2 = {r_M}^\top r_M - 2\mathbf{a}^\top\mathbf{w}+\mathbf{w}^\top\mathbf{B}\mathbf{w}$ (see Supplementary Material S1 of Berg and Song, 2023). Therefore the optimization problem becomes

$$\begin{aligned} \min_{\mathbf{w}} \quad & {r_M}^\top r_M - 2\mathbf{a}^\top\mathbf{w}+\mathbf{w}^\top\mathbf{B}\mathbf{w} \\ \text{subject to} \quad & \mathbf{w}\geq 0 \end{aligned} \tag{20}$$

which is a quadratic convex problem because $\mathbf{B}$ can be shown to be a positive definite matrix (Supplementary Material S1 in Berg and Song, 2023). Note that this objective is identical to the quadratic programming formulation of the non-negative least squares problem.

For computing $\Pi_\delta(r_M)$, in practice, we approximate the interval $[-1+\delta,1-\delta]$ with a finely spaced finite grid of s points $\Theta=\{\alpha_1,...,\alpha_s\}\subseteq[-1+\delta,1-\delta]$. We then approximate the solution $\Pi_\delta(r_M)=\Pi(r_M;[-1+\delta,1-\delta])$ by $\Pi(r_M;\Theta)$. Of course, if Θ contains the

support of $\hat{\mu}_\delta$, then $\Pi_\delta(r_M) = \Pi(r_M; \Theta)$. We used a grid of $s = 1001$ α values in $[-1 + \delta, 1 - \delta]$, where we first created an equally spaced grid $\mathcal{G}$ in a log-scale from $[0, 1 - \delta]$ and used $\mathcal{S} = -\mathcal{G} \cup \mathcal{G}$. We used the support reduction algorithm by Groeneboom et al. [2008] (ref. page 388) to solve (20) with this choice of Θ. In terms of run-time of our implementation, it took about .056 seconds on average to obtain $\Pi_\delta(r_M)$ for r_M from a length $M = 10000$ AR1 chain and the choice of grid above, on an author's typical personal laptop operating Mac OS with a 3.2 GHz processor. The implementation is available in `https://github.com/hsong1/momentLS`.

4 Statistical guarantee of the moment LS estimator

In this section, we analyze the statistical performance of the moment LS estimator. Specifically, we show that the moment least squares estimator $\Pi_\delta(r_M)$ obtained from any eligible initial autocovariance sequence estimator r_M satisfying (R.1)-(R.3) is ℓ_2-strongly consistent for the true autocovariance sequence, and the asymptotic variance estimate based on $\Pi_\delta(r_M)$ is strongly consistent for the true asymptotic variance σ^2 in (2).

First, the following Proposition shows that a wide range of estimators are allowed for the choice of the initial autocovariance sequence estimator r_M, including the empirical autocovariance estimator as well as windowed autocovariance estimators.

Proposition 7. *Assume that a Markov chain $X = \{X_0, X_1, \dots, \}$ with transition kernel Q satisfies conditions (A.1)-(A.3), and the function of interest g is in $L^2(\pi)$. The empirical autocovariance sequence $\tilde{r}_M$, defined as in (4), satisfies conditions (R.1)-(R.3) where $Y_M = M^{-1} \sum_{t=0}^{M-1} g(X_t)$. In addition, any windowed autocovariance sequence estimator $\check{r}_M$ such that $\check{r}_M(k) = \tilde{r}_M(k) w_M(|k|)$ for any window function w_M satisfying (W.1)-(W.3) satisfies (R.1)-(R.3).*

The proof is deferred to S5.1 in the Supplementary Material [Berg and Song, 2023].

4.1 L2 consistency of the moment LSE

We now show the strong consistency (with respect to the ℓ_2 metric) of the moment LSE $\Pi_\delta(r_M)$ for the true autocovariance sequence, that is, we show $\|\Pi_\delta(r_M) - \gamma\| \overset{a.s.}{\to} 0$, for any $\delta > 0$ satisfying $\mathrm{Supp}(F) \subseteq [-1 + \delta, 1 - \delta]$.

First of all, we present the following key lemma, which bounds the ℓ_2 distance between the projection $\Pi_\delta(r)$ of $r \in \ell_2(\mathbb{Z})$, and an element γ in $\mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$, with a mixture of geometrically weighted differences between the input r and γ. This lemma plays a crucial role in our convergence analysis. In our setting, the standard bound derived from the property of the projection

$$\|\Pi_\delta(r_M) - \gamma\|^2 \leq \|r_M - \gamma\|^2$$

for $\gamma \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$ is not helpful because we do not assume the consistency, with respect to the ℓ_2 metric, of r_M for the true autocovariance γ. In fact, the empirical autocovariance sequence seems not to converge to γ in the ℓ_2 sense. Even so, we can still show that a geometrically weighted difference between r_M and γ converges to 0, which leads to the convergence of $\Pi(r_M)$ to γ in the ℓ_2 sense.

Lemma 1. *Suppose $\delta \in [0, 1]$, and let $f \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$. Additionally, suppose $r \in \ell_2(\mathbb{Z})$. Then*

$$0 \leq \|\Pi_\delta(r) - f\|^2 \leq -\int \langle x_\alpha, r - f \rangle \, \mu_f(d\alpha) + \int \langle x_\alpha, r - f \rangle \, \hat{\mu}_\delta(d\alpha). \tag{21}$$

where $\hat{\mu}_\delta$ is the representing measure for $\Pi_\delta(r)$ and μ_f is the representing measure for f.

Proof. Clearly $0 \leq \|\Pi_\delta(r) - f\|^2$. We have,

$$\|f - \Pi_\delta(r)\|^2 = \langle f, f \rangle - 2 \langle \Pi_\delta(r), f \rangle + \langle \Pi_\delta(r), \Pi_\delta(r) \rangle . \tag{22}$$

First, for the third term in (22), by Proposition 5 and Lemma 4 in the Supplementary Material [Berg and Song, 2023], we have

$$\begin{aligned}
\langle \Pi_\delta(r), \Pi_\delta(r) \rangle &= \int \langle x_\alpha, \Pi_\delta(r) \rangle \, \hat{\mu}_\delta(d\alpha) \\
&= \int \langle x_\alpha, r \rangle \, \hat{\mu}_\delta(d\alpha) \\
&= \int \{ \langle x_\alpha, r - f \rangle + \langle x_\alpha, f \rangle \} \hat{\mu}_\delta(d\alpha) \\
&= \int \langle x_\alpha, r - f \rangle \, \hat{\mu}_\delta(d\alpha) + \langle \Pi_\delta(r), f \rangle ,
\end{aligned}$$

where the second equality follows from $\langle x_\alpha, \Pi_\delta(r) \rangle = \langle x_\alpha, r \rangle$ for all $\alpha \in \mathrm{Supp}(\hat{\mu})$. Thus,

(22) becomes,

$$\|f - \Pi_\delta(r)\|^2 = \langle f, f\rangle - \langle \Pi_\delta(r), f\rangle + \int \langle x_\alpha, r - f\rangle \, \hat{\mu}_\delta(d\alpha).$$

Now, for the second term in (22),

$$\begin{aligned}
\langle \Pi_\delta(r), f\rangle &= \int \langle x_\alpha, \Pi_\delta(r)\rangle \, \mu_f(d\alpha) \\
&\geq \int \langle x_\alpha, r\rangle \, \mu_f(d\alpha) \\
&= \int \{\langle x_\alpha, r - f\rangle + \langle x_\alpha, f\rangle\} \mu_f(d\alpha) \\
&= \int \langle x_\alpha, r - f\rangle \, \mu_f(d\alpha) + \langle f, f\rangle \, .
\end{aligned}$$

where for the second inequality we use Proposition 5 which states $\langle x_\alpha, \Pi_\delta(r)\rangle \geq \langle x_\alpha, r\rangle$ for all $\alpha \in [-1+\delta, 1-\delta] \cap (-1, 1)$, as well as Lemma 2 in the Supplementary Material [Berg and Song, 2023]. Therefore, we obtain,

$$\|f - \Pi_\delta(r)\|^2 \leq -\int \langle x_\alpha, r - f\rangle \, \mu_f(d\alpha) + \int \langle x_\alpha, r - f\rangle \, \hat{\mu}_\delta(d\alpha).$$

□

The next two propositions, Proposition 8 and 9, serve as the basis for proving the moment LS estimator's ℓ_2 consistency by proving the uniform convergence of the geometrically weighted difference between r_M and γ and the finiteness of the representing measure of $\Pi(r_M)$.

Proposition 8. *Let r_M denote an initial autocovariance sequence estimator satisfying (R.1)-(R.3). Let $\mathcal{K}$ denote a nonempty compact set with $\mathcal{K} \subseteq (-1, 1)$. Then we have*

$$\sup_{\alpha \in \mathcal{K}} |\, \langle r_M - \gamma, x_\alpha\rangle \,| \to 0 \quad P_x\text{-almost surely,} \tag{23}$$

as $M \to \infty$, for each initial condition $x \in \mathsf{X}$.

Proposition 9. *For a given $\delta > 0$ and an initial autocovariance sequence estimator r_M satisfying (R.1)-(R.3), let $\hat{\mu}_{\delta,M}$ denote the representing measure for $\Pi_\delta(r_M)$. Then there*

exists a constant $C_{\delta,\gamma} < \infty$ *with* $C_{\delta,\gamma}$ *depending only on* γ *and* δ *such that*

$$\limsup_{M\to\infty} \hat{\mu}_{\delta,M}([-1+\delta, 1-\delta]) \leq C_{\delta,\gamma}$$

P_x*-almost surely for any* $x \in \mathsf{X}$*. In particular,* $\hat{\mu}_{\delta,M}([-1+\delta, 1-\delta])$ *remains bounded almost surely.*

The proofs for Propositions 8 and 9 are in Supplementary Material S5.2 and S5.3 [Berg and Song, 2023]. Finally, we present the main result of this section in Theorem 2 below, which shows that the moment LSE is ℓ_2 consistent for the true autocovariance sequence γ. This result is the consequence of the key inequality in Lemma 1 as well as the uniform convergence of $\langle x_\alpha, r_M - \gamma\rangle$ and finiteness of the representing measure of $\Pi_\delta(r_M)$ in Proposition 8 and 9.

Theorem 2 (ℓ_2-consistency of Moment LSEs)**.** *Suppose* $X_0, X_1, ...,$ *is a Markov chain with transition kernel* Q *satisfying (A.1)-(A.3), and suppose* $g : \mathsf{X} \to \mathbb{R}$ *satisfies (B.1). Let* γ *denote the autocovariance sequence as defined in Proposition 1, and let* F *denote the representing measure for* γ*. Suppose* $\delta > 0$ *is chosen so that* F *is supported on* $[-1+\delta, 1-\delta]$*. Let* r_M *be an initial autocovariance sequence estimator satisfying conditions (R.1) - (R.3). Then*

$$\|\gamma - \Pi_\delta(r_M)\|^2 \underset{M\to\infty}{\to} 0, \quad P_x\text{-a.s.}$$

for each initial condition $x \in \mathsf{X}$*.*

Proof. From Proposition 1 and by the choice of δ, we have $\gamma \in \mathscr{M}_\infty(\delta)$ for $\delta > 0$. Then Lemma 3 in the Supplementary Material gives that $\gamma \in \ell_1(\mathbb{Z})$, and therefore $\gamma \in \ell_2(\mathbb{Z})$. Thus, we have $\gamma \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$. Additionally, $r_M \in \ell_2(\mathbb{Z})$ since r_M satisfies (R.2). Therefore, we can apply the result of Lemma 1, and we have the following inequality

$$\|\gamma - \Pi_\delta(r_M)\|^2 \leq -\int_{[-1,1]} \langle x_\alpha, r_M - \gamma\rangle \, F(d\alpha) + \int_{[-1,1]} \langle x_\alpha, r_M - \gamma\rangle \, \hat{\mu}_{\delta,M}(d\alpha).$$

where $\hat{\mu}_{\delta,M}$ is the representing measure for $\Pi_\delta(r_M)$. Note $\mathrm{Supp}(F) \subseteq [-1+\delta, 1-\delta]$ by the assumption on δ. Additionally, $\mathrm{Supp}(\hat{\mu}_{\delta,M}) \subseteq [-1+\delta, 1-\delta]$ since $\Pi_\delta(r_M) \in \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$.

Therefore, we have for any M

$$\|\gamma - \Pi_\delta(r_M)\|^2 \leq \left(\sup_{\alpha \in [-1+\delta, 1-\delta]} |\langle x_\alpha, r_M - \gamma \rangle| \right) \{F([-1,1]) + \hat{\mu}_{\delta,M}([-1,1])\}$$

and thus

$$\begin{aligned} \limsup_{M\to\infty} \|\gamma - \Pi_\delta(r_M)\|^2 \leq & \limsup_{M\to\infty} \left(\sup_{\alpha \in [-1+\delta, 1-\delta]} |\langle x_\alpha, r_M - \gamma \rangle| \right) F([-1,1]) \\ & + \limsup_{M\to\infty} \left(\sup_{\alpha \in [-1+\delta, 1-\delta]} |\langle x_\alpha, r_M - \gamma \rangle| \right) \limsup_{M\to\infty} \hat{\mu}_{\delta,M}([-1,1]). \end{aligned}$$

Let the initial condition for the chain $x \in \mathsf{X}$ be given. From Proposition 8, we know that $\left(\sup_{\alpha \in [-1+\delta, 1-\delta]} |\langle x_\alpha, r - \gamma \rangle|\right) \to 0$ P_x-a.s. Also we have $F([-1,1]) = \gamma(0) < \infty$ and $\limsup_{M\to\infty} \hat{\mu}_{\delta,M}([-1,1]) \leq C_{\delta,\gamma} < \infty$ P_x-a.s. from Proposition 9. Therefore, we have $\limsup_{M\to\infty} \|\gamma - \Pi_\delta(r_M)\|^2 = 0$ P_x-almost surely. Thus, $\|\gamma - \Pi_\delta(r_M)\|^2 \to 0$ P_x-almost surely as $M \to \infty$, as desired. □

An important consequence of Proposition 8, 9, and Theorem 2 is the measure convergence of $\hat{\mu}_{\delta,M}$ to the true representing measure F. Recall that for a sequence of measures $\{\nu_n\}_{n\in\mathbb{N}}$ and ν on $\mathbb{R}$, ν_n *converges vaguely* to ν if and only if $\int f d\nu_n \to \int f d\nu$ for all $f \in C_0(\mathbb{R})$ [e.g., Folland, 1999], where $C_0(\mathbb{R})$ is the space of continuous functions that vanish at infinity, i.e. $f \in C_0(\mathbb{R})$ iff f is continuous and the set $\{x; |f(x)| \geq \epsilon\}$ is compact for every $\epsilon > 0$.

Proposition 10 (vague convergence of $\hat{\mu}_{\delta,M}$)**.** *Assume the same conditions as in Theorem 2. For each initial condition $x \in \mathsf{X}$, we have $P_x(\hat{\mu}_{\delta,M} \to F$ vaguely, as $M \to \infty) = 1$, where $\hat{\mu}_{\delta,M}$ and F are the representing measures for $\Pi_\delta(r_M)$ and γ, respectively.*

This proposition is a direct consequence of the a.s. ℓ_2 convergence of $\Pi_\delta(r_M)$ to γ and Lemma 7 in the Supplementary Material S2 [Berg and Song, 2023].

4.2 Strong consistency of the asymptotic variance estimator based on the moment LSE

In this subsection, we present the strong consistency result for the asymptotic variance estimator based on the moment least squares estimators. It is well known that for a stationary, ψ-irreducible, geometrically ergodic, and reversible Markov chain and for a

square integrable g, the central limit theorem holds [e.g., see Corollary 6 in Haggstrom and Rosenthal, 2007], i.e.,

$$\sqrt{M}(Y_M - \mu) \xrightarrow{d} N(0, \sigma^2(\gamma)), \tag{24}$$

with

$$\sigma^2(\gamma) = \lim_{M\to\infty} M\mathrm{Var}(Y_M) = \sum_{k\in\mathbb{Z}} \gamma(k) = \int \frac{1+\alpha}{1-\alpha} F(d\alpha) < \infty, \tag{25}$$

where F denotes the representing measure associated with γ.

The main theorem for this subsection is Theorem 3, which shows that an asymptotic variance estimate based on the moment least squares estimator $\sigma^2(\Pi_\delta(r_M)) = \sum_{k\in\mathbb{Z}} \Pi_\delta(r_M)(k)$ is strongly consistent for $\sigma^2(\gamma)$ for any r_M which satisfies conditions (R.1) - (R.3).

Theorem 3 (strong consistency of asymptotic variance estimators based on Moment LSEs)**.** *Assume the same conditions as in Theorem 2. Let $\sigma^2(\gamma) = \sum_{k\in\mathbb{Z}} \gamma(k)$ be the asymptotic variance based on the true autocovariance sequence γ. We let $\sigma^2(\Pi_\delta(r_M)) = \sum_{k\in\mathbb{Z}} \Pi_\delta(r_M)(k)$ be an estimate of $\sigma^2(\gamma)$ based on the moment least squares estimator $\Pi_\delta(r_M)$. We have $\sigma^2(\Pi_\delta(r_M)) \to \sigma^2(\gamma)$ P_x-a.s., for each initial condition $x \in \mathsf{X}$, as $M \to \infty$.*

Proof. Let $\hat{\sigma}^2_M = \sigma^2(\Pi_\delta(r_M))$ and $\sigma^2 = \sigma^2(\gamma)$ for notational simplicity. Lemma 3 and Lemma 5 in the Supplementary Material give that $\hat{\sigma}^2_M = \int_{[-1+\delta,1-\delta]} \frac{1+\alpha}{1-\alpha} \hat{\mu}_{\delta,M}(d\alpha)$, and we have $\sigma^2 = \int_{[-1+\delta,1-\delta]} \frac{1+\alpha}{1-\alpha} F(d\alpha)$ from (25). Thus, we have

$$|\hat{\sigma}^2_M - \sigma^2| = \left| \int_{[-1+\delta,1-\delta]} \frac{1+\alpha}{1-\alpha} \hat{\mu}_{\delta,M}(d\alpha) - \int_{[-1+\delta,1-\delta]} \frac{1+\alpha}{1-\alpha} F(d\alpha) \right|.$$

We can obtain $f(\alpha) \in C_0(\mathbb{R})$ such that $f(\alpha) = \frac{1+\alpha}{1-\alpha}$ on $[-1+\delta, 1-\delta]$ by extending the two endpoints of $(1+\alpha)/(1-\alpha)$ at $\alpha \in \{-1+\delta, 1-\alpha\}$ to 0 linearly so that $f(\alpha) = 0$ for $|\alpha| \geq 1$.

More concretely, define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(\alpha) = \begin{cases} \frac{1+\alpha}{1-\alpha} & \alpha \in [-1+\delta, 1-\delta] \\ \frac{\alpha+1}{2-\delta} & -1 \leq \alpha \leq -1+\delta \\ \frac{-(2-\delta)\alpha+2-\delta}{\delta^2} & 1-\delta \leq \alpha \leq 1 \\ 0 & \alpha < -1 \text{ or } \alpha > 1. \end{cases}$$

Then $f \in C_0(\mathbb{R})$ and $f(\alpha) = \frac{1+\alpha}{1-\alpha}$ for $\alpha \in [-1+\delta, 1-\delta]$. Then, since $\text{Supp}(\hat{\mu}_{\delta,M}), \text{Supp}(F) \subseteq [-1+\delta, 1-\delta]$, we have

$$|\hat{\sigma}^2_M - \sigma^2| = \left| \int f(\alpha)\hat{\mu}_{\delta,M}(d\alpha) - \int f(\alpha)F(d\alpha) \right| \to 0, \quad P_x\text{-a.s.},$$

for any $x \in \mathsf{X}$ by the almost sure vague convergence of $\hat{\mu}_{\delta,M}$ to F in Proposition 10. □

5 Empirical studies

The goal of this section is two fold: first, we empirically illustrate some of the theoretical aspects discussed in the previous section, in particular, the ℓ_2 sequence consistency and asymptotic variance consistency of Moment LSEs. Second, we compare the performance of our method to the performance of other current state-of-the-art methods for autocovariance sequence estimation and asymptotic variance estimation. In Section 5.1.3, we propose a method for tuning the hyperparameter δ for moment LSEs. In Section 5.3, we use two simulation settings: one from a Metropolis-Hastings algorithm (Metropolis et al. [1953] and Hastings [1970]) with a discrete state space, and the other from a stationary AR(1) chain. In Section 5.4, we use a Bayesian probit regression.

5.1 Settings

5.1.1 Settings for simulated chains

Metropolis-Hastings chain We consider a Metropolis-Hastings chain on the discrete state space $(\mathsf{X}, 2^{\mathsf{X}})$ where $\mathsf{X} = \{1, 2, ..., d\}$, so that $d = 100$ states are possible. The stationary distribution for the simulation was constructed by normalizing a length d random vector $U = [U_1, U_2, ..., U_d]^T$ with $U_i \overset{iid}{\sim} \text{Uniform}(0, 1)$, so that $\pi(\{i\}) = U_i/(\sum_{i'=1}^{d} U_{i'})$. Each row of the proposal distribution P was constructed in the same manner, with $P(i, \{j\}) =$

$V_{ij}/\sum_{j=1}^{d} V_{ij}$ for random variables $V_{ij} \overset{iid}{\sim} \text{Uniform}(0,1)$. The Metropolis-Hastings algorithm was used to construct a transition kernel Q corresponding to the proposal distribution P. Finally, a function $g : \mathsf{X} \to \mathbb{R}$ was constructed via $g = [g_1, \ldots, g_d]$ with $g_j \overset{iid}{\sim} N(0,1)$. We generated a Markov chain $X_0, X_1, ...$ with stationary distribution π according to Q.

Since in this example, the transition kernel Q is on a small discrete state space, it is possible to compute the eigenvalues λ_i and eigenvectors ϕ_i corresponding to λ_i for $i = 1, \ldots, d$ numerically. Therefore, the true autocovariance sequence γ, the representing measure F for γ, and the asymptotic variance $\sigma^2(\gamma)$ can be all computed explicitly. More concretely, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$ denote the eigenvalues of Q. Suppose the eigenvectors ϕ_i are normalized so that $\langle \phi_i, \phi_j \rangle_\pi = 1[i = j]$. Note we have $\lambda_1 = 1$ and $\phi_1 = \mathbf{1}_d$ since $Q\mathbf{1}_d = \mathbf{1}_d$. We can write g and $\bar{g} = g - E_\pi[g(X_0)]\mathbf{1}_d$ as $g(k) = \sum_{i=1}^{d} \langle g, \phi_i \rangle_\pi \phi_i(k)$ and $\bar{g}(k) = \sum_{i=2}^{d} \langle g, \phi_i \rangle_\pi \phi_i(k)$ since $\langle g, \phi_1 \rangle_\pi = E_\pi[g(X_0)]$. Then, $\gamma(k) = \langle Q_0^{|k|} g, g \rangle_\pi = \langle Q^{|k|} \bar{g}, \bar{g} \rangle_\pi = \sum_{i=2}^{d} \langle g, \phi_i \rangle_\pi^2 \lambda_i^{|k|}$, and thus the representing measure for γ is

$$F = \sum_{i=2}^{d} \langle g, \phi_i \rangle_\pi^2 \delta_{\lambda_i} \tag{26}$$

where δ_a denotes a unit point mass at a. Finally, we have $\sigma^2(\gamma) = \sum_{i=2}^{d} \langle g, \phi_i \rangle_\pi^2 \frac{1+\lambda_i}{1-\lambda_i}$.

Autoregressive chain We also consider the autoregressive chain with the identity function $g(x) = x$ as in Example 2.1. We let $\tau^2 = 1$, and consider both positively and negatively correlated cases by setting $\rho = 0.9$ and $\rho = -0.9$ in each case, respectively.

5.1.2 Descriptions of estimators

We investigated the following autocovariance sequence estimators:

1. **(Empirical)** the empirical autocovariance sequence $\{\tilde{r}_M(k)\}_{k \in \mathbb{Z}}$,
2. **(Bartlett)** the windowed empirical autocovariance sequence $\check{r}_M(k) = w_M(|k|)\tilde{r}_M(k)$ with $w_M(k) = (1 - k/b_M^{(\text{Bart})})I(k < b_M^{(\text{Bart})})$ with threshold $b_M^{(\text{Bart})}$, and
3. **(MomentLS(Emp) and MomentLS(Bartlett))** our moment least squares estimators with the empirical autocovariance sequence $\Pi_\delta(\tilde{r}_M)$ and the windowed empirical autocovariance sequence $\Pi_\delta(\check{r}_M)$ as initial input sequences.

For all three sequence estimators (Empirical, Bartlett, and MomentLS), asymptotic variance estimates were obtained by summing up the sequence estimators over all $k \in \mathbb{Z}$. In the

case of Empirical and Bartlett estimators, this amounts to summing up the non-zero terms in the estimated autocovariance sequences $\tilde{r}_M$ or $\check{r}_M$. For MomentLS estimators, for each input sequence $r_M \in \{\tilde{r}_M, \check{r}_M\}$ and given $\delta > 0$, the sequence estimates were computed following steps outlined in Section 3.3. To elaborate further, we start by creating a grid $\Theta = \{\alpha_1, \ldots, \alpha_s\} \subseteq [-1+\delta, 1-\delta]$. We then solve the optimization problem (20) to obtain $\hat{\mathbf{w}} = [\hat{\mu}_\delta(\{\alpha_1\}), \ldots, \hat{\mu}_\delta(\{\alpha_s\})]^\top$. The momentLS sequence estimate for $\gamma(k)$ is $\Pi_\delta(r_M; \Theta)(k) = \sum_{\alpha; \hat{\mu}_\delta(\{\alpha\})>0} \alpha^{|k|} \hat{\mu}_\delta(\{\alpha\})$. The asymptotic variance estimate is

$$\sigma^2(\Pi_\delta(r_M; \Theta)) = \sum_{k \in \mathbb{Z}} \Pi_\delta(r_M; \Theta)(k) = \sum_{\alpha; \hat{\mu}_\delta(\{\alpha\})>0} \frac{1+\alpha}{1-\alpha} \hat{\mu}_\delta(\{\alpha\}).$$

The choice of δ is described in the next subsection 5.1.3.

For the comparison of asymptotic variance estimation performance, in addition to asymptotic variance estimates from the aforementioned estimators, we considered batch means, overlapping batch means, and initial sequence estimators. Let $Y_M = M^{-1} \sum_{t=0}^{M-1} g(X_t)$. For $i \leq M - b$, define the batch mean starting at i with batch length b by $Y_b(i) = b^{-1} \sum_{k=0}^{b-1} g(X_{i+k})$. Then the batch means, overlapping batch means, and initial sequence estimators are defined as

4. **(BM)** the batch mean estimator $\hat{\sigma}^2_{BM}$ with batch size $b_M^{(\mathrm{BM})}$,

$$\hat{\sigma}^2_{BM} = \frac{b_M}{a_M - 1} \sum_{k=0}^{a_M - 1} \{Y_{b_M^{(BM)}}(k b_M^{(BM)}) - Y_M\}^2,$$

   where $a_M = \lfloor M / b_M^{(\mathrm{BM})} \rfloor$ is the number of batches,

5. **(OLBM)** the overlapping batch mean estimator $\hat{\sigma}^2_{OLBM}$ with batch size $b_M^{(\mathrm{OLBM})}$,

$$\hat{\sigma}^2_{OLBM} = \frac{M b_M^{(\mathrm{OLBM})}}{(M - b_M^{(\mathrm{OLBM})})(M - b_M^{(\mathrm{OLBM})} + 1)} \sum_{j=0}^{M - b_M^{(\mathrm{OLBM})} + 1} \{Y_{b_M^{(\mathrm{OLBM})}}(j) - Y_M\}^2,$$

6. **(Init)** the initial (positive,monotone,convex) sequence estimator $\hat{\sigma}^2_{\mathrm{init,type}}$ computed as

$$\hat{\sigma}^2_{\mathrm{init,type}} = -\tilde{r}_M(0) + 2 \sum_{k=0}^{T-1} \hat{\Gamma}_M^{(\mathrm{type})}(k)$$

   for type $\in \{\mathrm{pos}, \mathrm{mono}, \mathrm{conv}\}$, where $\hat{\Gamma}_M(k) = \tilde{r}_M(2k) + \tilde{r}_M(2k+1)$, $T := \min\{k \in$

$\mathbb{N}; \hat{\Gamma}_M(j) < 0, \}$ is the first time point where $\hat{\Gamma}_M(k)$ becomes negative, and $\hat{\Gamma}_M^{(\text{pos})}(k)$, $\hat{\Gamma}_M^{(\text{mono})}(k)$, and $\hat{\Gamma}_M^{(\text{conv})}(k)$ are defined for $k < T$ as

- $\hat{\Gamma}_M^{(\text{pos})}(k) = \hat{\Gamma}_M(k)$,
- $\hat{\Gamma}_M^{(\text{mono})}(k) = \min_{j \leq k} \hat{\Gamma}_M^{(\text{pos})}(j)$, and
- $\hat{\Gamma}_M^{(\text{conv})}(k)$ is the kth element of the greatest convex minorant of $\hat{\Gamma}_M(0), \ldots, \hat{\Gamma}_M(T-1)$

for $k = 0, \ldots, T-1$.

The asymptotic variance estimator from the empirical autocovariance sequence is always 0, i.e., $\sigma^2(\tilde{r}_M) = \sum_{k \in \mathbb{Z}} \tilde{r}_M(k) = 0$ [e.g., Brockwell and Davis, 2009], and therefore is inconsistent for $\sigma^2(\gamma)$ whenever $\sigma^2(\gamma) > 0$. The asymptotic variance estimator from a windowed empirical autocovariance sequence is also sometimes called a spectral variance estimator since it corresponds to an estimated spectral density function at frequency 0.

5.1.3 Choice of hyperparameters

Hyperparameters are required for the Bartlett windowed estimators, BM, OLBM, and Moment LSEs. A batch size b_M needs to be specified a priori for the Bartlett windowed sequence estimate, BM, and OLBM, and δ determining the set $\mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$ onto which the initial autocovariance sequence r_M is projected must be specified for the MomentLS estimators.

For BM and OLBM, we used oracle hyperparameter settings when possible. From Flegal and Jones [2010], for the BM and OLBM methods, the mean-squared-error optimal batch sizes for estimating $\sigma^2(\gamma)$ are

$$b_M^{(\text{BM})} = \left(\frac{\Gamma^2 M}{\sigma^2(\gamma)} \right)^{1/3} \quad \text{and} \quad b_M^{(\text{OLBM})} = \left(\frac{8\Gamma^2 M}{3\sigma^2(\gamma)} \right)^{1/3} \tag{27}$$

respectively, where $\Gamma = -2 \sum_{s=1}^{\infty} s\gamma(s)$. Since the spectral variance estimator based on the Bartlett window is asymptotically equivalent to the overlapping batch mean estimator [Damerdji, 1991], we let $b_M^{(\text{Bart})} = b_M^{(\text{OLBM})}$. If oracle hyperparameters cannot be obtained because γ is unknown, we used the batch size tuning method implemented in the R package **mcmcse** [Liu et al., 2021].

For MomentLS estimators, we consider an oracle and data-driven choice of δ. An oracle choice of δ for MomentLS would be $\delta_\gamma = 1 - \sup\{|x|; x \in \text{Supp}(F)\}$ for the representing

measure F for the autocovariance sequence γ. For the data-driven choice of δ, we tune δ based on a modification of an adaptive bandwidth selection method proposed by Politis [2003].

Politis [2003] proposed an empirical rule of picking a lag $\hat{m}$ at which to truncate the autocovariance sequence. Under the assumption of uniform convergence of the empirical autocorrelations $\hat{\rho}_M(k) = \tilde{r}_M(k)/\tilde{r}_M(0)$ such that

$$\max_{k=0,\ldots,M-1} |\hat{\rho}_M(k) - \rho(k)| = O_P(\sqrt{\log M/M}) \tag{28}$$

(ref. eq (10) in Politis [2003]), Politis [2003] proposed the use of an estimator $\hat{m}$ satisfying $\hat{m}/\log(M) \to -1/(2\log|\alpha|)$ in probability, for stationary discrete-time process $X_1, \ldots, X_M$ with an exponentially-decaying autocovariance sequence satisfying $\gamma(k) = C\alpha^{|k|}, |k| > k_0$ for some $k_0 < \infty$ and $|\alpha| < 1$.

In our setting, $\gamma(k) = \int \alpha^{|k|} F(d\alpha)$ is a mixture of $\alpha^{|k|}$. Recall that any fixed choice of $\delta > 0$ such that $\text{Supp}(F) \subseteq [-1+\delta, 1-\delta]$ is valid to guarantee the a.s. sequence and asymptotic variance estimator convergences in Theorems 2 and 3. In particular, any fixed $\delta \leq \delta_\gamma$ is a valid choice for a moment LS estimator. Note that δ_γ can be larger than the spectral gap of the transition kernel Q. With a modification of the empirical rule in Politis [2003], we propose to use a data-driven $\tilde{\delta}_M$ such that $\tilde{\delta}_M < \delta_\gamma$ with high probability under the condition of (28).

Compared to the empirical rule by Politis [2003], our proposed rule focuses only on even lags of empirical autocorrelations. More concretely, we first choose $\hat{m}$ such that

$$\hat{m} = \min\{t \in 2\mathbb{N}; \hat{\rho}_M(t+2) \leq c_M\sqrt{\log M/M}\} \tag{29}$$

for some $c_M \geq 0$. This change is motivated by the fact that for reversible chains, $\gamma(k)$ is always nonnegative for even k, and the magnitude of $\gamma(k)$ can be arbitrarily small for odd k due to the potential cancellations of α^k terms from positive and negative α values. To illustrate this point, consider a simple example with $\gamma(k) = (-0.9)^k + 0.9^k$ for $k = 0, 1, 2, \ldots$; it is clear that $\gamma(k) = 0$ for any odd k.

Once we have determined $\hat{m}$, we let

$$\hat{\delta}_M = \max\{1 - \exp\{-\log(M)/(2\hat{m})\}, 1/M\}. \tag{30}$$

Under the condition (28), we show that $1 - \exp\{-\log(M)/(2\hat{m})\}$ is not asymptotically larger than δ_γ for any choice of $c_M \geq 0$, and converges to δ_γ in probability as $M \to \infty$ if we choose c_M so that $c_M \to \infty$ such that $c_M = O(\log(M))$ (see Supplementary Material S6 of Berg and Song, 2023). Therefore for any positive $c_\delta < 1$, $\tilde{\delta}_M = c_\delta \hat{\delta}_M$ should serve as an asymptotically conservative choice for δ_γ. We choose $c_\delta < 1$ in $\tilde{\delta}_M = c_\delta \hat{\delta}_M$, since for $c_\delta = 1$, the probability of $\tilde{\delta}_M > \delta_\gamma$ may not go to 0, even in the case that $\hat{\delta}_M$ converges to δ_γ in probability. We note that whereas the Politis [2003] procedure allows for nonincreasing $c_M = c$, we were unable to verify $\hat{m}/\log(M) \xrightarrow{p} -1/\{2\log(1-\delta_\gamma)\}$ without the condition $c_M \to \infty$.

Additionally, since $\hat{\delta}_M$ is random, the finite sample performance of momentLS estimators is influenced by the variability of $\hat{\delta}_M$. We use an averaging procedure in order to reduce the variability of $\hat{\delta}_M$, in which $\hat{\delta}$ estimates from separate segments of the observed chain $\{g(X_t)\}_{t=0}^{M-1}$ are averaged. Specifically, we partition the observed series $\{g(X_t)\}_{t=0}^{M-1}$ into L equal length splits, and compute the empirical autocovariances for each split in the following way. Let $B = \lfloor M/L \rfloor$. The kth autocovariance from the lth split, for $l = 1, \ldots, L$, is computed as

$$\tilde{r}_{M/L}^{(l)}(k) = \begin{cases} \frac{1}{B}\sum_{t=0}^{B-1-k} \tilde{g}(X_t)\tilde{g}(X_{t+k}) & l = 1 \\ \frac{1}{B}\sum_{t=(l-1)B-k}^{lB-1-k} \tilde{g}(X_t)\tilde{g}(X_{t+k}) & l > 1 \end{cases}$$

where we recall $\tilde{g}(X_t) = g(X_t) - M^{-1}\sum_{t=0}^{M-1} g(X_t)$. Then we computed $\hat{\delta}_{M/L}^{(l)}$ using $\{\tilde{r}_{M/L}^{(l)}(k)\}_{k=0}^{B-1}$ for $l = 1, \ldots, L$. Finally, we used

$$\tilde{\delta}_M = 0.8\frac{1}{L}\sum_{l=1}^{L} \hat{\delta}_{M/L}^{(l)}$$

with the choice $L = 5$ as the input for the Moment LS estimators in the experiments.

It is worth mentioning that in Theorems 2 and 3, the provided almost sure convergence guarantees are applicable to a Moment LS estimator $\Pi_\delta(r_M)$ with a valid, non-random δ. Also, while uniform convergence of empirical autocovariance sequences has been studied and the uniform bound (28) has been established for certain stationary time series whose examples include IID chains and the AR(1) chain of Example 2.1 with $g(x) = x$, see e.g., An et al. [1982], Kavalieris [2008], it is still an open question to establish similar results for a general geometrically ergodic Markov chain with arbitrary initial condition. We leave it

as a future work to provide a full justification for moment LS estimators with this tuned choice of δ.

5.2 Empirical illustration of the convergence properties of Moment LSEs

We recall that the convergence guarantees in Theorems 2 and 3 apply for Moment LS estimates with δ chosen such that $\delta > 0$ and $\text{Supp}(F) \subseteq [-1+\delta, 1-\delta]$, where F is the representing measure for the autocovariance sequence. Here, we empirically explore convergence of both the autocovariance sequence and the asymptotic variance estimators at varying δ levels, including cases in which the support of F is *not* contained in $[-1+\delta, 1-\delta]$. This latter setting is not covered by our Theorems 2 and 3, and in this case we expect the projection to $\mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$ to lead to bias in the corresponding Moment LSE.

For δ chosen such that $\text{Supp}(F) \subseteq [-1+\delta, 1-\delta]$, Figures 1 and 2 show that Moment LSEs lead to consistent estimates for both the autocovariance sequence (with respect to the ℓ_2 distance) and the asymptotic variance $\sigma^2(\gamma)$. Larger values of δ (subject to $\text{Supp}(F) \subseteq [-1+\delta, 1-\delta]$) lead to relatively better performance in the estimation of both the autocovariance sequence and the asymptotic variance, although the rates of convergence at different values of δ appear to be similar.

When $\delta = 0$, the moment LS estimator appears to be consistent for the true autocovariance sequence with respect to the ℓ_2 norm distance, but inconsistent with respect to the asymptotic variance (Figure 1). On $(-1, 1)$, the function $\alpha \to \frac{1+\alpha}{1-\alpha}$ is unbounded and can no longer be uniformly approximated by polynomials of finite degree. Thus the ℓ_2 sequence convergence property at $\delta = 0$ does not transfer, as in Theorem 3 with $\delta > 0$, to convergence of the estimated asymptotic variance.

In the setting where $\delta > 0$ is chosen so large that $\text{Supp}(F)$ is not contained in $[-1+\delta, 1-\delta]$, we observe an apparent bias variance trade-off. Our results in this setting suggest that an optimal choice of δ will strike a balance between the increase in variability expected in projecting to larger sets $\mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$ for small δ, and the increase in approximation error expected when $\gamma \notin \mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$ for large δ. In the discrete state space Metropolis-Hastings example, $\delta \leq 0.355$ is required for $\text{Supp}(F) \subseteq [-1+\delta, 1-\delta]$, yet for the smaller sample sizes in our study $\delta = 0.5$ leads to the best performance out of all values of δ considered for estimating both the autocovariance sequence and the asymptotic variance (Figure 1). We suspect that the improved performance for $\delta = 0.5$ results from decreased

variance, and that the bias introduced by restricting the support of $\hat{\mu}_{\delta,M}$ to $[-0.5, 0.5]$ is not too large since the representing measure F in this example has a substantial amount of mass between $[-0.5, 0.5]$. On the other hand, in the AR(1) example with $\rho = 0.9$, the representing measure F has no support within $[-0.8, 0.8]$, and the setting $\delta = 0.2$ leads to poor performance, suggesting that the bias introduced at this value of δ overcomes any gains in performance due to variance reduction.

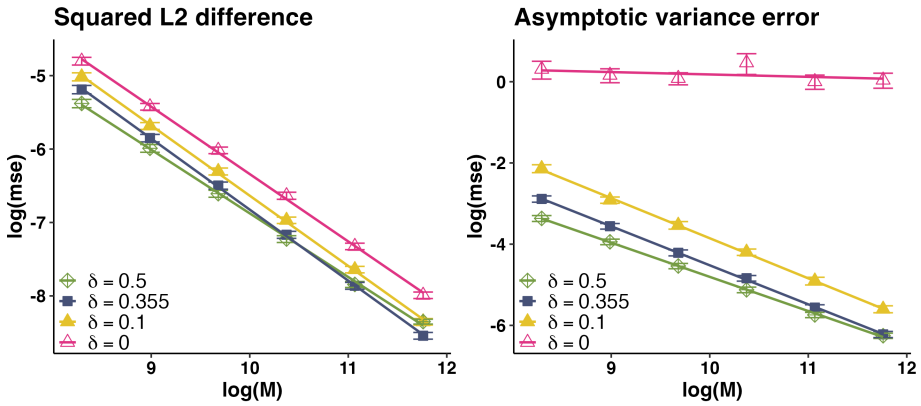


Figure 1: Metropolis-Hastings example. The support of the representing measure for γ is contained in $[-.645, .645]$, i.e., the valid δ range is $0 < \delta \leq .355$.

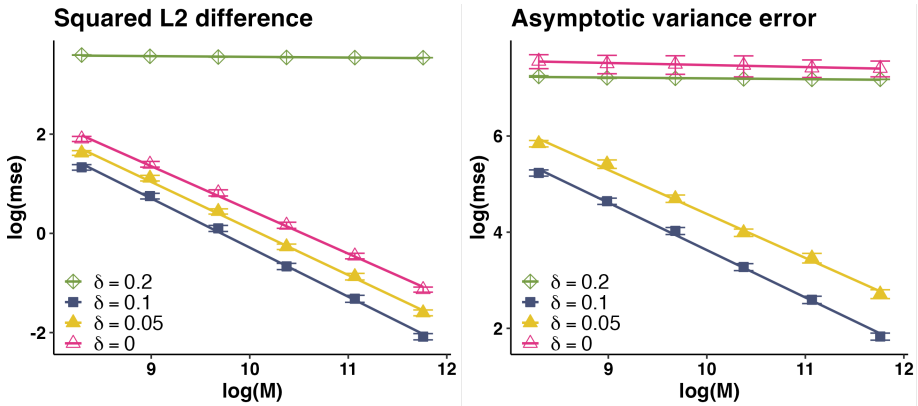


Figure 2: AR(1) example with a positive correlation ($\rho = 0.9$). The representing measure has a single support point at .9. The valid δ range is $0 < \delta \leq .1$.

5.3 Comparison with other state-of-the-art estimators for simulated chains

This subsection compares the performance of our method to the performance of other current state-of-the-art methods for autocovariance sequence estimation and asymptotic variance estimation using two simulated chains.

We computed the squared ℓ_2 autocovariance sequence error $\|\hat{r} - \gamma\|_2^2$ (when eligible)

and the squared asymptotic variance error $(\hat{\sigma}^2 - \sigma^2(\gamma))^2$ for $B = 400$ simulations from each method with varying chain lengths $M \in \{4000, 8000, 16000, 32000, 64000, 128000\}$. All simulations were performed using R software [R Core Team, 2020]. We used the **mcmcse** package [Flegal et al., 2021] for computing BM and OLBM estimators and the **mcmc** package [Geyer and Johnson, 2020] for computing initial positive, monotone, and convex sequence estimators.

The average squared ℓ_2 autocovariance sequence error and average squared asymptotic variance estimation error are reported in Figures 3-5 and in tables in Supplementary Material S7 [Berg and Song, 2023]. In these results,

- MomentLS(Tune,Emp) and MomentLS(Tune-Incr,Emp) refer to the moment LS estimators with the empirical autocovariance used for r_M and with δ chosen using the tuning procedure in Section 5.1.3, with the choices $c_M = 0$ and $c_M = 0.01\sqrt{\log M}$ in (29), and
- MomentLS(Orcl,Emp), MomentLS(Orcl,Brtl) refer to the moment LS estimates with oracle hyperparameter $\delta = \delta_\gamma$ and the empirical and Bartlett windowed autocovariances as inputs respectively.

We excluded the initial positive and monotone sequence estimators from the plots, since these generally performed similarly to or worse than the initial convex sequence estimator. To avoid overcrowding the plots, we also excluded the empirical estimator for the squared ℓ_2 error and the empirical, Bartlett, and MomentLS(Orcl,Brtl) estimators for the asymptotic variance error from Figures 3 - 5. We also reported only the MomentLS(Tune,Emp) results and excluded the MomentLS(Tune-Incr,Emp) results in Figures 3 - 5 because both sets of results were very similar. Tables that include these results can be found in Supplementary Material Section S7 [Berg and Song, 2023].

Metropolis-Hastings chain The first plot in Figure 3 displays the squared ℓ_2 error $\|\hat{r} - \gamma\|^2$ for several estimators $\hat{r}$. Notably, the moment LSEs using the empirical autocovariance sequence as r_M perform best out of the estimators considered for all sample sizes, with both the data driven and oracle tuning of δ. The moment LSE with the Bartlett windowed sequence as the input sequence (MomentLS(Orcl,Brtl)) has reduced ℓ_2 sequence error relative to the original Bartlett windowed autocovariance sequence (Bartlett). MomentLS(Orcl,Brtl) appear to converge slower than for the Moment LSEs with the empirical

autocovariance sequence as input. This decrease in convergence rate may be due to information loss from the thresholding of higher lag autocovariances in the Bartlett window sequences, which prevents information at higher lags from being used at all, in contrast to the empirical autocovariance sequence, where information from all lags can be used.

The second plot in Figure 3 compares mean squared errors for the asymptotic variance estimation. The MomentLSEs using the empirical autocovariance sequence as input again perform best out of the considered estimators. While the performance of the moment LSE with the data-driven selection of δ and that of the initial convex sequence estimator appear to be quite similar, the former shows a slightly superior performance, especially for larger values of M. The batch means estimator (BM) appears to perform slightly worse than the overlapping batch means estimator (OLBM).

Figure 4 shows a plot of the true, empirical, and moment LS estimated covariances for lags $k = 0, ..., 100$ based on a single simulation with sample size $M = 8000$. The empirical and moment LS estimated covariances are similar for very small k, but for larger k the empirical autocovariances clearly have large fluctuations about the true covariances relative to the moment LS covariances. These fluctuations apparently account for the large squared ℓ_2 error $\|\gamma - \tilde{r}_M\|^2$ of the empirical estimator.

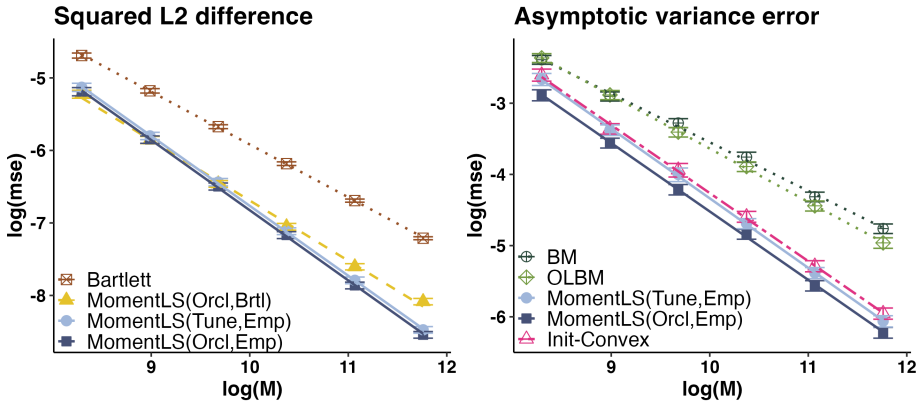


Figure 3: Plots for the discrete state space Metropolis-Hastings example. The first plot shows squared ℓ_2 error $\|\gamma - \hat{r}\|^2$ and the second plot shows mean squared error for the asymptotic variance estimation. The error bars represent 1 standard error from $B = 400$ simulations.

Autoregressive chain In Figure 5, we see generally comparable patterns in both of the AR(1) chain settings as in the discrete Metropolis-Hastings scenario. The estimated autocovariance sequences from the MomentLSEs with empirical autocovariances as the input sequences generally perform the best of the considered estimators in terms of squared

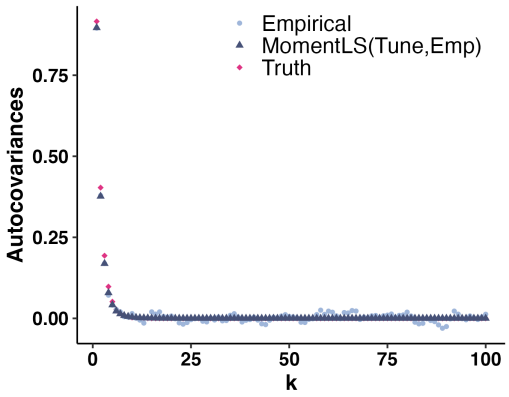


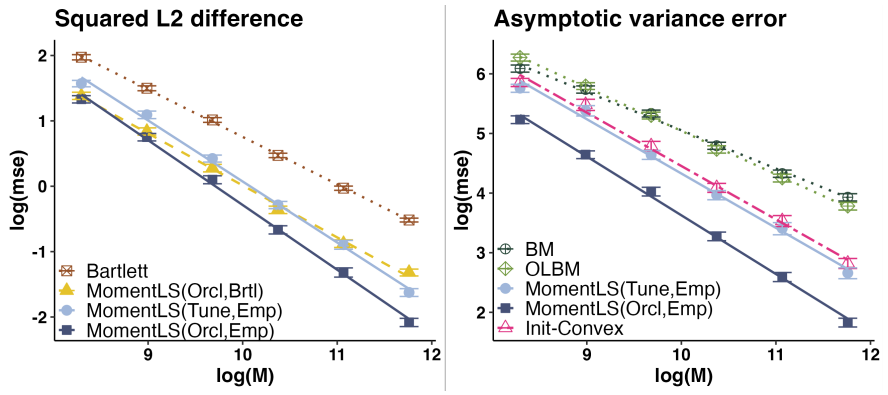
Figure 4: For the discrete state space Metropolis-Hastings example, a comparison of true, empirical, and moment LS estimated autocovariances from a single simulation with $M = 8000$.

ℓ_2 error and mean squared error for estimation of the asymptotic variance. In the $\rho = -0.9$ setting, the performance of the initial convex sequence estimator appears to be quite poor relative to the other estimators. Similarly to Figure 4 for the Metropolis-Hastings example, Figure 6 clearly shows the benefit of imposing shape constraints on the autocovariance sequence estimation, as the moment LS estimates $\Pi_\delta(\tilde{r}_M)(k)$ are much closer to the true autocovariance sequence than the empirical autocovariances $\tilde{r}_M(k)$, especially for large lags k.

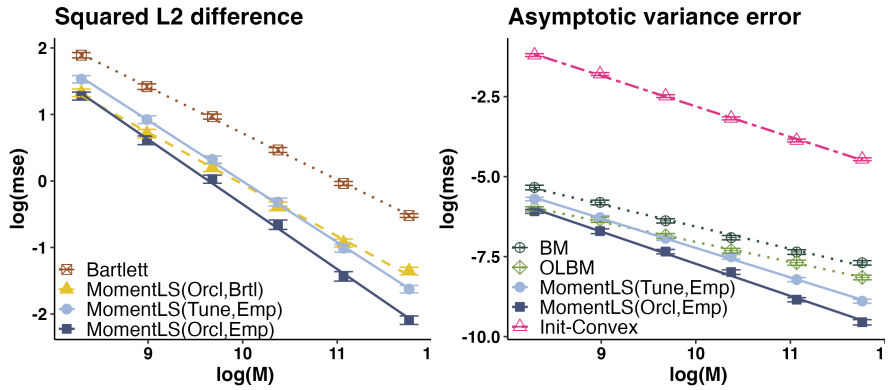
5.4 Bayesian probit regression

In this section, we illustrate the effectiveness of our method in a more realistic Bayesian probit regression model. We first compare the estimated asymptotic variances from the competing methods. In addition to this, as we mentioned in the Introduction, an asymptotic variance estimator is needed to quantify uncertainty in the MCMC estimates and to effectively terminate the chain based on the perceived precision of the MCMC estimates. We conduct two experiments in this regard: first, we construct confidence intervals based on the estimated asymptotic variances of competing methods for a fixed length chain and compare their coverage probabilities; and second, we compare the coverage probabilities of competing methods for a variable length chain, where for each method the chain length is determined by a fixed-width rule.

We consider the Glass identification data from the UCI machine learning repository. The dataset contains 214 examples of the chemical analysis of 7 different types of glass. We aim to predict the first glass type based on its 9 chemical properties $\mathbf{x} = (x_1, \ldots, x_9) \in \mathbb{R}^9$.



(a) $\rho = 0.9$



(b) $\rho = -0.9$

Figure 5: Plots for the autoregressive example. Plots in the left column show squared ℓ_2 error $\|\gamma - \hat{r}\|^2$ at $\rho = 0.9$ and $\rho = -0.9$. Plots in the right column show mean squared error for the asymptotic variance estimation at $\rho = 0.9$ and $\rho = -0.9$. The error bars represent 1 standard error from $B = 400$ simulations.

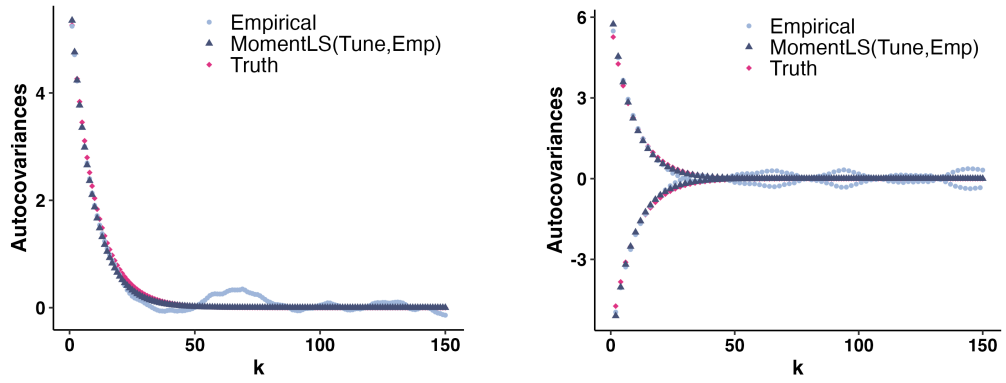
For the ith observation, we let $Y_i = 1$ if it is of the first glass type. We suppose

$$Pr(Y_i = 1) = \Phi(\beta_0 + \sum_{j=1}^{9} \beta_j x_{ij})$$

and assign independent $N(0, 1)$ priors on $\beta = (\beta_0, \ldots, \beta_9)$.

We sample $\{\beta(t)\}_{t=0}^{M-1}$ from the posterior distribution $\beta|\{Y_i\}_{i=1}^{214} \sim \pi(\cdot)$ using the data augmentation Gibbs sampler of Albert and Chib [1993]. This sampler is displayed in Algorithm 1. We let $\mathbf{X} \in \mathbb{R}^{n\times 10}$ be the design matrix where each row of $\mathbf{X}$ is $[1, \mathbf{x}_i]$. The marginal chain $\{\beta(t)\}_{t\geq 0}$, which we consider here, is reversible with respect to the posterior π [see, e.g., Liu et al., 1994, Robert and Casella, 2004]. Additionally, the $\{\beta(t)\}_{t\geq 0}$ chain has been shown to be geometrically ergodic [Chakraborty and Khare, 2017].

To compare estimated asymptotic variances and coverage probabilities from the com-



(a) $\rho = 0.9$ **(b)** $\rho = -0.9$

Figure 6: For the autoregressive example with (a) $\rho = 0.9$ and (b) $\rho = -0.9$, a comparison of true, empirical, and moment LS estimated autocovariances from a single simulation with $M = 8000$.

Algorithm 1 Albert and Chib [1993] sampler

1. Draw independent $z_1, \ldots, z_n$ with $z_i \sim TN(\mathbf{x}_i^\top \beta, 1, y_i)$, $i = 1, \ldots, n$.
   Let $\mathbf{z} = [z_1, \ldots, z_n]$.
2. Draw $\beta \sim N_p((\mathbf{X}^\top \mathbf{X} + I_n)^{-1} \mathbf{X}^\top \mathbf{z}, (\mathbf{X}^\top \mathbf{X} + I_n)^{-1})$.

peting methods, we need accurate reference estimates of posterior mean and asymptotic variance for each coefficient. Since both quantities are unknown, we independently generated a long chain $\{\beta_{\text{long}}(t)\}_{t=0}^{M_1-1}$ with $M_1 = 5 \times 10^6$ iterations to estimate posterior mean and also $B = 1000$ independent chains $\{\beta_{\text{par}}^{(b)}(t)\}_{t=0}^{M_2-1}$ with $M_2 = 5 \times 10^4$ to estimate asymptotic variance. Specifically, we use $\beta_{\text{orcl},j} = M_1^{-1} \sum_{t=0}^{M_1-1} \beta_{\text{long},j}(t)$ to estimate the posterior mean of the jth coefficient, and use $\sigma^2_{\text{orcl},j} = M \sum_{b=1}^{1000} (\bar{\beta}_{\text{par},j}^{(b)} - \bar{\bar{\beta}}_{\text{par},j})^2$ to estimate the asymptotic variance for the jth coefficient, where $\bar{\beta}_{\text{par},j}^{(b)}$ refers to the sample mean value of β_j from the bth chain and $\bar{\bar{\beta}}_{\text{par},j} = \frac{1}{1000} \sum_{b=1}^{1000} \bar{\beta}_{\text{par},j}^{(b)}$ refers to the sample mean of $\bar{\beta}_{\text{par},j}^{(b)}$.

Table 1 shows some estimated summary properties for the chains from Albert and Chib [1993] sampler, including the estimated posterior mean β_{orcl}, asymptotic variance σ^2_{orcl}, Monte Carlo standard error (MCSE) for β_{orcl}, as well as the estimated multiplier for the effective sample size $M_{\text{eff}}/M = 1/(1 + 2\sum_{t \in \mathbb{Z}} \rho(t))$, lag 1 autocorrelation $\rho(1)$, and δ_γ, the gap between 1 and the largest support point (in magnitude) for the representing measure of γ. Note that a smaller value of δ_γ implies slower mixing, as the spectral gap should be at least as small as δ_γ. In the table, $\text{MCSE}_j = \sigma_{\text{orcl},j}/\sqrt{M_1}$, M_{eff}/M was estimated based on σ^2_{orcl} and the lag 0 empirical autocovariances from the parallel chains, $\rho(1)$ was estimated based on the empirical autocovariances at lag 0 and 1 of the long chain $\{\beta_{\text{long}}(t)\}_{t=0}^{M_1-1}$, and δ_γ was estimated by $\hat{\delta}_{M_1}$ in (30), also using the long chain $\{\beta_{\text{long}}(t)\}_{t=0}^{M_1-1}$. For many of the

coefficients, the estimated gap δ_γ is relatively small.

Table 1: *Some estimated summary properties of the chains from the Albert and Chib [1993] sampler.*

Coef	β_{orcl}	σ^2_{orcl}	MCSE	M_{eff}/M	$\rho(1)$	δ_γ
β_0	-1.262	3.965	8.91×10^{-4}	0.013	0.912	0.025
β_1	0.301	0.337	2.56×10^{-4}	0.268	0.553	0.114
β_2	-0.198	1.187	4.87×10^{-4}	0.102	0.351	0.050
β_3	1.555	3.055	7.82×10^{-4}	0.111	0.257	0.039
β_4	-0.768	1.611	5.68×10^{-4}	0.062	0.599	0.040
β_5	0.451	0.772	3.93×10^{-4}	0.155	0.339	0.058
β_6	-0.016	7.863	1.25×10^{-3}	0.025	0.708	0.042
β_7	0.047	0.966	4.40×10^{-4}	0.347	0.217	0.114
β_8	0.080	9.235	1.36×10^{-3}	0.019	0.791	0.027
β_9	-0.103	0.056	1.06×10^{-4}	0.216	0.567	0.088

Comparison of asymptotic variance estimates We first compare the asymptotic variance estimates $\hat{\sigma}_{jM}$ obtained by BM, OLBM, Init-Convex, and MomentLS, for each coefficient β_j, $j = 0, \ldots, 9$.

Table 2: *Estimated mean squared relative errors (s.e.) for asymptotic variance estimates for the Glass data Bayesian probit regression with $B = 400$ simulations. For each simulation, we generated a length $M = 16000$ chain for β. The method with the smallest estimated mean squared errors is highlighted in bold for each coefficient.*

Coef	BM	OLBM	MomentLS.Tune.Emp.	Init.Convex
β_0	0.102 (0.003)	0.089 (0.003)	**0.048 (0.004)**	0.062 (0.006)
β_1	0.008 (0.000)	0.007 (0.000)	**0.003 (0.000)**	0.004 (0.000)
β_2	0.299 (0.002)	0.268 (0.002)	0.036 (0.002)	**0.033 (0.002)**
β_3	0.446 (0.002)	0.415 (0.002)	0.069 (0.003)	**0.046 (0.002)**
β_4	0.195 (0.002)	0.172 (0.002)	**0.029 (0.002)**	0.030 (0.002)
β_5	0.216 (0.002)	0.193 (0.002)	0.039 (0.002)	**0.031 (0.002)**
β_6	0.235 (0.003)	0.204 (0.003)	**0.024 (0.002)**	0.026 (0.002)
β_7	0.113 (0.001)	0.101 (0.001)	0.054 (0.001)	**0.035 (0.001)**
β_8	0.229 (0.004)	0.201 (0.004)	**0.052 (0.005)**	0.061 (0.005)
β_9	0.020 (0.001)	0.017 (0.001)	0.011 (0.001)	**0.011 (0.001)**

Table 2 shows the mean squared relative errors $\{(\hat{\sigma}^2_j - \sigma^2_{\text{orcl},j})/\sigma^2_{\text{orcl},j}\}^2$ from $B = 400$ simulated chains of length $M = 16000$. Generally, both moment LS and initial convex sequence estimators perform better than the batch means and overlapping batch means estimators. The Moment LS estimator and initial convex sequence estimator perform quite similarly.

Comparison of coverage probabilities We compare the coverage probabilities of the confidence intervals

$$\bar{\beta}_{jM} \pm t_{\alpha/2,M-1}\frac{\hat{\sigma}_{jM}}{\sqrt{M}} \tag{31}$$

for each coefficient β_j, $j = 0, \ldots, 9$, using $\hat{\sigma}_{jM}$ produced by BM, OLBM, Init-Convex, and MomentLS. For comparison, we also consider Oracle coverage probabilities based on the estimated "true" asymptotic variances $\sigma^2_{\text{orcl},j}$ as in the previous section.

Table 3 shows the estimated coverage probabilities for 95% confidence intervals (31) from length $M = 16000$ chains based on the asymptotic variances from the four methods (BM, OLBM, Init-Convex, and MomentLS) as well as using the Oracle asymptotic variance estimate. We used $B = 1000$ independent simulations. From Table 3, we observe that the coverage percentages for the BM and OLBM methods tend to be lower than the nominal 95% coverage probability. The moment LS and initial convex sequence estimates show more similar behavior, with the initial convex sequence estimates achieving coverage closest to the nominal 95% more often.

Table 3: *Estimated coverage probabilities for the Glass data Bayesian probit regression with $B = 1000$ simulations. For each simulation, we generated a length $M = 16000$ chain for β. The method whose coverage probability is closest to 95% (excluding the Oracle) is highlighted in bold for each coefficient.*

Estimator	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
BM	0.88	**0.93**	0.81	0.73	0.85	0.81	0.84	0.89	0.84	**0.92**
OLBM	0.89	**0.93**	0.83	0.74	0.86	0.82	0.85	0.89	0.85	**0.92**
MomentLS(Tune,Emp)	**0.94**	**0.93**	**0.93**	0.91	**0.93**	0.91	**0.94**	0.92	**0.93**	**0.92**
Init-Convex	0.93	**0.93**	**0.93**	**0.92**	**0.93**	**0.92**	**0.94**	**0.93**	**0.93**	**0.92**
Oracle	0.94	0.93	0.94	0.94	0.95	0.95	0.95	0.95	0.95	0.93

We also compared the coverage probabilities in the context of fixed-width methodology [Jones et al., 2006]. The idea of fixed-width rules is to terminate the simulation once a desirable confidence interval half-width ϵ for an MCMC estimate is achieved. For a specified accuracy ϵ, we terminate the chain the first time the following inequality holds:

$$\max\left\{t_{\alpha/2,M-1}\frac{\hat{\sigma}_{jM}}{\sqrt{M}},\ \text{for } j = 0, 1, 2, \ldots, 9\right\} + p(M) \leq \epsilon \tag{32}$$

where $p(M) = \epsilon I\left(M \leq M^*\right) + M^{-1}$ and M^* is a desirable minimum chain length. The role of $p(M)$ is to ensure that the simulation is not terminated too prematurely. Glynn and

Whitt [1992] established that if a functional central limit theorem holds and if a strongly consistent asymptotic variance estimator is used, the $1-\alpha$ confidence interval whose chain length M is chosen based on the fixed-width rule (32) is asymptotically valid as $\epsilon \to 0$.

We simulated $B = 1000$ chains using the fixed-width rules based on the BM, OLBM, Init-Convex, and Moment LS asymptotic variance estimates. As before, the Oracle row of the table refers to coverage probability and sample size selection based on the reference asymptotic variance values $\sigma^2_{\text{orcl},j}$ for each coefficient. We began each simulation with a minimum chain length of M^*, and if the criterion (32) is not satisfied, an additional 10% of the current number of iterations were performed before checking the criterion again. We computed the 95% confidence intervals based on the simulated chains (with random lengths) and checked whether the constructed confidence intervals included the true posterior mean or not. We used $\epsilon = 0.05$ and the minimum chain length $M^* = 1000$.

Table 4 reports the coverage probabilities. We observe a similar result as in the previous comparison. BM and OLBM tend to produce too liberal intervals. Moment LS and initial sequence estimates seem to achieve coverage probability closest to the nominal level on average, with the initial sequence estimates achieving coverage closer to nominal more often.

Table 4: *Average chain length at termination and coverage probabilities for the Glass data Bayesian probit regression with $B = 1000$ simulations using fixed-width methods. The first column displays the mean (s.e.) chain length at termination. The method whose coverage probability is closest to 95% (excluding the Oracle) is highlighted in bold for each coefficient.*

Estimator	M (s.e.)	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	β_9
BM	4,227 (40)	0.82	0.94	0.75	0.63	0.80	0.79	0.74	0.88	0.77	0.91
OLBM	4,563 (42)	0.83	0.94	0.76	0.66	0.80	0.81	0.75	0.90	0.80	0.91
MomentLS(Tune,Emp)	9,850 (70)	0.93	**0.95**	**0.93**	0.87	**0.93**	0.89	**0.94**	**0.93**	0.92	**0.94**
Init-Convex	10,022 (76)	**0.94**	**0.95**	**0.93**	**0.90**	0.92	**0.91**	**0.94**	**0.93**	**0.93**	**0.94**
Oracle	10,832 (0)	0.95	0.94	0.96	0.94	0.94	0.94	0.95	0.95	0.94	0.94

We note that in this section we have treated asymptotic variance estimation for the coefficient vector β in a component-wise fashion. It can be beneficial to also consider output analysis tools that take cross-covariance between components into consideration [e.g., Vats et al., 2019]. In this regard, extending the current framework to estimate the asymptotic variance matrix for multivariate functions of the Markov chain state, as in Dai and Jones [2017], Vats et al. [2018], is of interest.

6 Conclusion

In this work, we proposed a novel shape-constrained estimator for the autocovariance sequence from a reversible Markov chain. To the best of our knowledge, this is the first work in which the spectral representation of the autocovariance sequence is exploited to estimate the autocovariance sequence subject to infinitely many shape constraints. We have carried out a thorough analysis of the proposed Moment LS estimator, including its characterization and theoretical guarantees. Especially, we showed the strong consistency of the autocovariance sequence estimate from the Moment LS estimator in terms of an ℓ_2 error metric, convergence of the representing measure of the Moment LS estimator to the true representing measure, and the strong consistency of an estimate of the Markov chain CLT asymptotic variance based on our autocovariance sequence estimator. Our theoretical results hold for reversible and geometrically ergodic Markov chains. Finally, we empirically validated our theoretical findings and demonstrated the effectiveness of the proposed estimator compared to existing autocovariance estimators in both simulated and real data settings, including batch means, spectral variance estimators, and initial sequence estimators.

7 Acknowledgements

HS and SB gratefully acknowledge support from NSF DMS-2311141.

Supplement to "Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain"

Hyebin Song and Stephen Berg

Department of Statistics, Pennsylvania State University

S1 Computation of Moment LS estimators

In this section, we provide some details in obtaining the convex optimization problem in (20). Recall

$$\begin{aligned}\sum_{k\in\mathbb{Z}}(r_M(k)-m(k))^2 &= \sum_{k\in\mathbb{Z}}(r_M(k)-\sum_{i=1}^{s}\alpha_i^{|k|}w_i)^2 \\ &= \sum_{k;\in\mathbb{Z}} r_M(k)^2 - 2\sum_{k\in\mathbb{Z}} r_M(k)\left(\sum_{i=1}^{s}\alpha_i^{|k|}w_i\right)+\sum_{k\in\mathbb{Z}}\left(\sum_{i=1}^{s}\alpha_i^{|k|}w_i\right)^2.\end{aligned} \tag{S-1}$$

and the definitions of $\mathbf{w}$, $\mathbf{a}$, and $\mathbf{B}$.

The first term in (S-1) is simply ${r_M}^\top r_M$ for an input vector r_M. For the second term, we have

$$\begin{aligned}\sum_{k\in\mathbb{Z}} r_M(k)\left(\sum_{i=1}^{s}\alpha_i^{|k|}w_i\right) &= \sum_{|k|\le T_0-1} r_M(k)\left(\sum_{i=1}^{s}\alpha_i^{|k|}w_i\right) \\ &= \sum_{i=1}^{s}\left(\sum_{|k|\le T_0-1} r_M(k)\alpha_i^{|k|}\right)w_i = \mathbf{a}^\top\mathbf{w}.\end{aligned}$$

For the third term, we have

$$\begin{aligned}\sum_{k\in\mathbb{Z}}\left(\sum_{i=1}^{s}\alpha_i^{|k|}w_i\right)^2 &= \sum_{k\in\mathbb{Z}}\sum_{i=1}^{s}\sum_{j=1}^{s}\alpha_i^{|k|}\alpha_j^{|k|}w_iw_j \\ &= \sum_{i,j=1}^{s}\frac{1+\alpha_i\alpha_j}{1-\alpha_i\alpha_j}w_iw_j = \mathbf{w}^\top\mathbf{B}\mathbf{w}.\end{aligned}$$

Therefore we have

$$\sum_{k\in\mathbb{Z}}(r_M(k)-m(k))^2 = {r_M}^\top r_M - 2\mathbf{a}^\top\mathbf{w}+\mathbf{w}^\top\mathbf{B}\mathbf{w}$$

as desired. Since we minimize over $m \in \mathscr{M}_\infty(\Theta) \cap \ell_2(\mathbb{Z})$, we require

$$\mathbf{w} = [\mu_m(\{\alpha_1\}), \ldots, \mu_m(\{\alpha_s\})] \geq 0$$

elementwise. Finally, we note that $\mathbf{B}$ is a positive definite matrix because $\mathbf{w}^\top \mathbf{B}\mathbf{w} = 0$ implies that $\sum_{i=1}^s \alpha_i^{|k|} w_i = 0$ for all $k \in \mathbb{Z}$. By choosing at least s distinct $|k|$, we obtain $\mathbf{w} = 0$.

S2 A few technical Lemmas

Lemma 2. *Suppose $f \in \mathscr{M}_\infty(0) \cap \ell_2(\mathbb{Z})$, and let F be the representing measure for f, i.e., $f(k) = \int x^{|k|} F(dx)$. Then $F(\{-1, 1\}) = 0$. That is, the measure F does not have any point mass on -1 or 1.*

Proof. For any $k \in \mathbb{Z}$,

$$F(\{-1, 1\}) \leq \int_{[-1,1]} x^{2k} F(dx) = f(2k).$$

Since $f \in \ell_2(\mathbb{Z})$, we have $f(2k) \to 0$ as $k \to \infty$. Thus, $F(\{-1, 1\}) = 0$.

□

Lemma 3. *Suppose $f \in \mathscr{M}_\infty(\delta)$ for some $\delta > 0$. Then $f \in \ell_1(\mathbb{Z})$.*

Proof. Let F denote the representing measure for f. We have

$$\begin{aligned}
\sum_{k \in \mathbb{Z}} |f(k)| &= \sum_{k \in \mathbb{Z}} \left| \int x^{|k|} F(dx) \right| \\
&\leq \sum_{k \in \mathbb{Z}} \int |x|^{|k|} F(dx) \\
&= \int \sum_{k \in \mathbb{Z}} |x|^{|k|} F(dx)
\end{aligned}$$

where the last equality is due to Tonelli's theorem. Also since

$$\sum_{k \in \mathbb{Z}} |x|^{|k|} = 1 + 2 \sum_{k \geq 1} |x|^k = \frac{1 + |x|}{1 - |x|},$$

we have,

$$\sum_{k\in\mathbb{Z}} |f(k)| \leq \int \frac{1+|x|}{1-|x|} F(dx) \leq \sup_{x\in[-1+\delta,1-\delta]} \left(\frac{1+|x|}{1-|x|}\right) \int 1F(dx) = \frac{2-\delta}{\delta} f(0) < \infty.$$

where we used the fact $0 \leq \int 1F(dx) < \infty$ since $f \in \mathscr{M}_\infty(\delta)$ implies F is a finite, regular measure. Thus $f \in \ell_1(\mathbb{Z})$.

□

Lemma 4. *Suppose $f \in \mathscr{M}_\infty(0) \cap \ell_2(\mathbb{Z})$ with $f(k) = \int x^{|k|} F(dx)$ and $g \in \ell_2(\mathbb{Z})$. Then*

$$\langle f, g\rangle = \int \langle x_\alpha, g\rangle \, F(d\alpha) = \int_{[-1,1]} \langle x_\alpha, g\rangle \, F(d\alpha).$$

Proof. We have

$$\begin{aligned}
\langle f, g\rangle &= \sum_{k\in\mathbb{Z}} f(k)g(k) \\
&= \sum_{k\in\mathbb{Z}} g(k) \int_{[-1,1]} \alpha^{|k|} F(d\alpha) \\
&= \sum_{k\in\mathbb{Z}} \int_{[-1,1]} g(k)\alpha^{|k|} F(d\alpha).
\end{aligned}$$

We will show that $\sum_{k\in\mathbb{Z}} \int_{[-1,1]} |g(k)\alpha^{|k|}| F(d\alpha) < \infty$. Then, the desired result follows from Fubini's theorem, since

$$\sum_{k\in\mathbb{Z}} \int_{[-1,1]} g(k)\alpha^{|k|} F(d\alpha) = \int_{[-1,1]} \sum_{k\in\mathbb{Z}} g(k)\alpha^{|k|} F(d\alpha) = \int_{[-1,1]} \langle x_\alpha, g\rangle \, F(d\alpha).$$

We have

$$\sum_{k\in\mathbb{Z}} \int_{[-1,1]} |g(k)\alpha^{|k|}| F(d\alpha) = \sum_{k\in\mathbb{Z}} |g(k)| \int_{[-1,1]} |\alpha|^{|k|} F(d\alpha) = \sum_{k\in\mathbb{Z}} |g(k)| \tilde{f}(k) \leq \|g\| \|\tilde{f}\|,$$

where we define $\tilde{f}(k) = \int_{[-1,1]} |\alpha|^{|k|} F(d\alpha)$ and we use Cauchy-Schwarz for the last inequality. First, since $g \in \ell_2(\mathbb{Z})$, $\|g\| < \infty$. For $\|\tilde{f}\|$, we have $\tilde{f}(k) = \tilde{f}(-k)$, and for $0 \leq k_1 < k_2$, we have $\tilde{f}(k_1) \geq \tilde{f}(k_2)$. Additionally, for $n = 2k$ we have $\tilde{f}(n) = f(n)$. Thus

$$\|\tilde{f}\|_2^2 = \sum_{k\in\mathbb{Z}} \tilde{f}(k)^2 = \tilde{f}(0)^2 + 2\tilde{f}(1)^2 + 2\sum_{k=2}^{\infty} \tilde{f}(k)^2 \leq 3f(0)^2 + 4\sum_{k=1}^{\infty} f(2k)^2 < \infty$$

since $f \in \ell_2(\mathbb{Z})$. □

Corollary 2. *For $f, g \in \mathscr{M}_\infty(0) \cap \ell_2(\mathbb{Z})$ with $f(k) = \int_{[-1,1]} x^{|k|} F(dx)$ and $g(k) = \int_{[-1,1]} x^{|k|} dG(x)$,*

$$\begin{aligned}\langle f, g\rangle &= \int_{[-1,1]} \int_{[-1,1]} \langle x_{\alpha_1}, x_{\alpha_2}\rangle \, F(d\alpha_1) G(d\alpha_2) \\ &= \int_{[-1,1]} \int_{[-1,1]} \frac{1+\alpha_1\alpha_2}{1-\alpha_1\alpha_2} F(d\alpha_1) G(d\alpha_2).\end{aligned}$$

Additionally, the order of integration in both expressions can be interchanged.

Proof. Note than since both $f, g \in \mathscr{M}_\infty(0) \cap \ell_2(\mathbb{Z})$, by Lemma 2, $F(\{-1,1\}), G(\{-1,1\}) = 0$. We have

$$\begin{aligned}\langle f, g\rangle &= \int_{[-1,1]} \langle x_{\alpha_1}, g\rangle \, F(d\alpha_1) \\ &= \int_{(-1,1)} \langle x_{\alpha_1}, g\rangle \, F(d\alpha_1) \\ &= \int_{(-1,1)} \int_{[-1,1]} \langle x_{\alpha_1}, x_{\alpha_2}\rangle \, G(d\alpha_2) F(d\alpha_1) \\ &= \int_{[-1,1]} \int_{[-1,1]} \langle x_{\alpha_1}, x_{\alpha_2}\rangle \, G(d\alpha_2) F(d\alpha_1) \\ &= \int_{(-1,1)} \int_{(-1,1)} \langle x_{\alpha_1}, x_{\alpha_2}\rangle \, G(d\alpha_2) F(d\alpha_1) \\ &= \int_{(-1,1)} \int_{(-1,1)} \frac{1+\alpha_1\alpha_2}{1-\alpha_1\alpha_2} G(d\alpha_2) F(d\alpha_1) \\ &= \int_{[-1,1]} \int_{[-1,1]} \frac{1+\alpha_1\alpha_2}{1-\alpha_1\alpha_2} G(d\alpha_2) F(d\alpha_1) < \infty.\end{aligned}$$

Since $\langle x_{\alpha_1}, x_{\alpha_2}\rangle = \frac{1+\alpha_1\alpha_2}{1-\alpha_1\alpha_2} \geq 0$ for all $\alpha_1, \alpha_2 \in [-1,1]$, we can interchange the order of integration. □

A by-product of the Corollary is $\langle f, g\rangle \geq 0$ for all $f, g \in \mathscr{M}_\infty(0) \cap \ell_2(\mathbb{Z})$.

Lemma 5. *Suppose $f \in \mathscr{M}_\infty(0) \cap \ell_1(\mathbb{Z})$ with $f(k) = \int \alpha^{|k|} F(d\alpha)$. Define*

$$\sigma^2(f) = \sum_{k=-\infty}^{\infty} f(k).$$

Then

$$\sigma^2(f) = \int \frac{1+\alpha}{1-\alpha} F(d\alpha).$$

Proof. We have

$$\sum_{k=-\infty}^{\infty} \int |\alpha|^{|k|} F(d\alpha) = f(0) + 2\sum_{k=1}^{\infty} \int |\alpha|^k F(d\alpha) \leq 3f(0) + 4\sum_{j=1}^{\infty} f(2j) < \infty$$

where the first inequality follows from

$$\int |\alpha| F(d\alpha) \leq \int 1 F(d\alpha)$$

and

$$\int |\alpha|^{2k+1} F(d\alpha) \leq \int \alpha^{2k} F(d\alpha), \quad k \geq 1,$$

and the second inequality follows from $f \in \ell_1(\mathbb{Z})$. Thus, from Fubini's theorem, we have

$$\begin{aligned} \sigma^2(f) &= \sum_{k=-\infty}^{\infty} \int \alpha^{|k|} F(d\alpha) \\ &= \int \sum_{k=-\infty}^{\infty} \alpha^{|k|} F(d\alpha) \\ &= \int \frac{1+\alpha}{1-\alpha} F(d\alpha). \end{aligned}$$

□

Lemma 6. *Let I be a closed interval in $[-1, 1]$. The space $\mathscr{M}_\infty(I) \cap \ell_2(\mathbb{Z})$ is closed.*

Proof. Let $a = \inf I$ and $b = \sup I$. Consider a sequence of vectors $\{m_n\} \subseteq \mathscr{M}_\infty([a, b]) \cap \ell_2(\mathbb{Z})$ where $\|m_n - f\| \to 0$ for some $f = \{f(k)\}_{k=-\infty}^{\infty}$. We show that $f \in \mathscr{M}_\infty([a, b]) \cap \ell_2(\mathbb{Z})$.

First, $f \in \ell_2(\mathbb{Z})$ since

$$\|f\| \leq \|m_n\| + \|f - m_n\| < \infty$$

for large enough n.

Next, we show that $f \in \mathscr{M}_\infty([a, b])$. Note for any $j \in \mathbb{Z}$, $|m_n(j) - f(j)| \leq \|m_n - f\| \to 0$. We consider two cases where Case I: $b - a = 0$ and Case II: $b - a > 0$,

Case I: we have $f(0) = \lim_n m_n(0)$. Additionally, $m_n(k) = m_n(0)a^{|k|}$ for $k \neq 0$, and thus, $f(k) = \lim_n m_n(k) = f(0)a^{|k|}$ for $k \neq 0$. Then for the measure μ with point mass at a with mass $f(0)$, i.e., $\mu = f(0)\delta_a$, we have $f(k) = \int x^{|k|} \mu(dx)$ for $k \in \mathbb{Z}$, and μ is supported on $\{a\}$. Thus $f \in \mathscr{M}_\infty([a, b]) \cap \ell_2(\mathbb{Z})$.

Case II: we define $\tilde{f}$ on $\mathbb{N}$ as $\tilde{f}(k) = T(f; a, b)(k)$. We have for any $r, k \in \mathbb{N}$,

$$\begin{aligned}(-1)^r \Delta^r \tilde{f}(k) &= (-1)^r \Delta^r \lim_{n\to\infty} T(m_n; a, b)(k) \\ &= \lim_{n\to\infty} (-1)^r \Delta^r T(m_n; a, b)(k) \\ &\geq 0.\end{aligned}$$

where the last inequality holds since m_n are $[a, b]$-moment sequences. Thus $\tilde{f}$ is completely monotone, so by Corollary 1, f is an $[a, b]$-moment sequence. Since in addition $f \in \ell_2(\mathbb{Z})$, we have $f \in \mathscr{M}_\infty([a, b]) \cap \ell_2(\mathbb{Z})$. Thus $\mathscr{M}_\infty([a, b]) \cap \ell_2(\mathbb{Z})$ is closed. □

Lemma 7. *Let I be a closed interval in $[-1, 1]$. Consider a sequence of moment sequences $\{f_n\}_{n\in\mathbb{N}}$ and f such that $\{f_n\} \subseteq \mathscr{M}_\infty(I) \cap \ell_2(\mathbb{Z})$ and $\|f_n - f\| \to 0$. Then $f \in \mathscr{M}_\infty(I) \cap \ell_2(\mathbb{Z})$. Let μ_n and μ be the representing measures for f_n and f respectively. Then, we have $\mu_n \to \mu$ vaguely.*

Proof. First of all, $f \in \mathscr{M}_\infty(I) \cap \ell_2(\mathbb{Z})$ follows from Lemma 6. Let $\epsilon > 0$ and $h \in C_0(\mathbb{R})$ given. We want to show that $|\int h(\alpha)\mu_n(d\alpha) - \int h(\alpha)\mu(d\alpha)| \leq \epsilon$ for a sufficiently large n. Also, since $\|f_n - f\| \to 0$, we have $f_n(0) \to f(0)$. In other words, $\mu_n(I) \to \mu(I)$.

Now we approximate h on I. Since h is continuous, there exists a sequence of polynomials $p_N(\alpha) = \sum_{k=0}^{N} c_k \alpha^k$ which uniformly approximates $h(\alpha)$ on I, by the Weierstrass approximation theorem. Let $B = \mu(I) + \sup_n \mu_n(I)$. Suppose $B = 0$. That is, $f(0) = 0$ and $f_n(0) = 0$ for all n. Therefore, both μ_n and μ are null measures, and the conclusion trivially holds. Now suppose $B > 0$. We choose $N < \infty$ so that

$$\sup_{\alpha\in I} |h(\alpha) - p_N(\alpha)| \leq \frac{\epsilon}{2B}. \tag{S-2}$$

We have,

$$
\begin{aligned}
&|\int h(\alpha)\mu_n(d\alpha) - \int h(\alpha)\mu(d\alpha)| \\
&= |\int_I h(\alpha)\mu_n(d\alpha) - \int_I h(\alpha)\mu(d\alpha)| \\
&= \Big| \int_I \{h(\alpha) - p_N(\alpha)\}\mu_n(d\alpha) + \int_I p_N(\alpha)\mu_n(d\alpha) \\
&\qquad\qquad - \int_I \{h(\alpha) - p_N(\alpha)\}\mu(d\alpha) - \int_I p_N(\alpha)\mu(d\alpha) \Big| . \\
&\leq \underbrace{\int_I |h(\alpha) - p_N(\alpha)|\mu_n(d\alpha) + \int_I |h(\alpha) - p_N(\alpha)|\mu(d\alpha)}_{Term1} \\
&\qquad\qquad + \underbrace{\left| \int_I p_N(\alpha)\mu_n(d\alpha) - \int_I p_N(\alpha)\mu(d\alpha) \right|}_{Term2} .
\end{aligned}
$$

By the choice of N,

$$
\text{Term 1} \leq \epsilon \frac{\mu_n(I)}{2B} + \epsilon \frac{\mu(I)}{2B} \leq \frac{\epsilon}{2}, \tag{S-3}
$$

since $\mu(I) + \mu_n(I) \leq B$ for any n.

For term II, define $v_N = \{v_N(k)\}_{k\in\mathbb{Z}}$ such that

$$
v_N(k) = \begin{cases} c_k & 0 \leq k \leq N, \\ 0 & otherwise. \end{cases}
$$

for $\{c_k\}_{k=0}^N$ from the coefficients of the approximating polynomial $p_N(\alpha) = \sum_{k=0}^N c_k \alpha^k$. Note that $\|v_N\|^2 = \sum_{k=0}^N c_k^2 < \infty$ since $N < \infty$. In particular, $v_N \in \ell_2(\mathbb{Z})$, and thus by Lemma 4,

$$
\begin{aligned}
\int_I p_N(\alpha)\mu_n(d\alpha) &= \int_I \sum_{k=0}^N c_k \alpha^k \mu_n(d\alpha) \\
&= \int_I \sum_{k\in\mathbb{Z}} v_N(k)\alpha^{|k|}\mu_n(d\alpha) \\
&= \int_I \langle v_N, x_\alpha \rangle \, \mu_n(d\alpha) = \langle v_N, f_n \rangle .
\end{aligned}
$$

Similarly, we can show $\int_I p_N(\alpha)\mu(d\alpha) = \langle v_N, f\rangle$. Therefore,

$$\text{Term 2} = |\langle v_N, f_n\rangle - \langle v_N, f\rangle| \leq \|v_N\|\|f_n - f\|.$$

We can find $L < \infty$ such that for $n \geq L$, $\|v_N\|\|f_n - f\| \leq \epsilon/2$.

Combining these results for term I and term II, we have for $n \geq L$,

$$\text{Term 1+ Term 2} \leq \epsilon$$

But $\epsilon > 0$ was arbitrary. This proves the result. □

S3 Proofs for results in Section 2

S3.1 Proof of Proposition 1

Proof. The representation (13) is a consequence of the spectral theorem [e.g., Rudin, 1991] since Q_0 is a self-adjoint bounded linear operator on $L^2(\pi)$. The spectrum $\sigma(Q_0)$ lies on the real axis due to (A.2). Since the spectral radius $\rho(Q_0) = \sup\{|\rho|; \rho \in \sigma(Q_0)\}$ is equal to $\|Q_0\|_{L^2(\pi)}$ since Q_0 is self-adjoint, and $\|Q_0\|_{L^2(\pi)} \leq 1$, we have $\sigma(Q_0) \subseteq [-1, 1]$.

For Γ, we have

$$\Gamma(k) = \langle Q_0^{2k} g, g\rangle_\pi + \langle Q_0^{2k+1} g, g\rangle_\pi = \langle Q_0^{2k}(I + Q_0)g, g\rangle_\pi$$

for $k \in \mathbb{N}$. Let $(I + Q_0)^{1/2}$ denote the square root of $I + Q_0$, which is well defined because $I + Q_0$ is positive and self-adjoint, where the positivity of $I + Q_0$ is due to the fact that $\|Q_0\|_{L^2(\pi)} \leq 1$. Also, we have that $(I + Q_0)^{1/2}$ is positive, self-adjoint, and commutes with Q_0 [e.g., Theorem in Riesz and Sz.-Nagy, 2012, p265]. Therefore,

$$\begin{aligned}
\Gamma(k) &= \langle (I + Q_0)^{1/2} Q_0^{2k} (I + Q_0)^{1/2} g, g\rangle_\pi \\
&= \langle Q_0^{2k} (I + Q_0)^{1/2} g, (I + Q_0)^{1/2} g\rangle_\pi \\
&= \langle (Q_0^2)^k h, h\rangle_\pi
\end{aligned}$$

where $h = (I + Q_0)^{1/2} g$. Then by the spectral theorem, there exists a regular measure H supported on $\sigma(Q_0^2)$ such that $\Gamma(k) = \int x^k H(dx)$, $k \in \mathbb{N}$. Since Q_0^2 is positive and $\|Q_0^2\|_{L^2(\pi)} = \|Q_0\|^2_{L^2(\pi)}$, we have $\sigma(Q_0^2) \subseteq [0, 1]$.

Finally, with the additional assumption of (A.1) and (A.3), the spectral gap $1-\rho(Q_0) > 0$ [Roberts and Rosenthal, 1997, Kontoyiannis and Meyn, 2012]. We can find $\delta_0 > 0$ so that $\rho(Q_0) = \|Q_0\|_{L^2(\pi)} = 1-\delta_0$. Therefore $\sigma(Q_0) \subseteq [-1+\delta_0, 1-\delta_0]$. Since $\|Q_0^2\|_{L^2(\pi)} = \|Q_0\|_{L^2(\pi)}^2 = (1-\delta_0)^2$, we have $\sigma(Q_0^2) \subseteq [0, (1-\delta_0)^2]$. □

S4 Proofs for results in Section 3

S4.1 Proof of Proposition 2

Proof. First, suppose there is a measure μ on $[a,b]$ such that $m(k) = \int x^k \mu(dx)$ for all $k \in \mathbb{N}$. Define $g_k(x) = x^k$ and $f(x) = (x-a)/(b-a)$. Also, we define $\tilde{\mu}(A) = \mu(h(A))$ where $h(x) = f^{-1}(x) = (b-a)x + a$ and $h(A) := \{h(x); x \in A\}$. We show that $\tilde{\mu}$ is a representing measure for $T(m;a,b)$ and $\tilde{\mu}$ is supported on $[0,1]$. First of all, by the change of variable formula, for $k \in \mathbb{N}$,

$$\begin{aligned}
T(m;a,b)(k) &= (b-a)^{-k} \sum_{i=0}^{k} \binom{k}{i} m(i)(-a)^{k-i} \\
&= (b-a)^{-k} \int_{[a,b]} \sum_{i=0}^{k} \binom{k}{i} x^k (-a)^{k-i} \mu(dx) \\
&= (b-a)^{-k} \int_{[a,b]} (x-a)^k \mu(dx) \\
&= \int I\{f(x) \in [0,1]\} f(x)^k \mu(dx) \\
&= \int_{[0,1]} y^k \tilde{\mu}(dy).
\end{aligned}$$

Also, $\tilde{\mu}$ is supported on $[0,1]$ since

$$\tilde{\mu}(\mathbb{R} \setminus [0,1]) = \mu(\{h(x); x < 0 \text{ or } x > 1\}) = \mu(\mathbb{R} \setminus [a,b]) = 0.$$

Then by Theorem 1, $\tilde{\mu}$ is the unique representing measure for $T(m;a,b)$, and $T(m;a,b)$ is completely monotone.

Now suppose $T(m;a,b)$ is completely monotone. Then there exists a measure $\tilde{\mu}$ supported on $[0,1]$ such that $T(m;a,b)(k) = \int y^k \tilde{\mu}(dy)$. Define the measure μ by $\mu(A) = \tilde{\mu}(f(A))$ where $f(A) := \{f(x); x \in A\}$. First, note that μ is supported on $[a,b]$. From the

definition of $T(m;a,b)$, we have $m(0) = T(m;a,b)(0)$ and, recursively,

$$m(k) = (b-a)^k T(m;a,b)(k) - \sum_{i=0}^{k-1} \binom{k}{i} m(i)(-a)^{k-i}. \tag{S-4}$$

We now show that $m(k) = \int x^k \mu(dx)$ for $k \in \mathbb{N}$. Recall $h(x) = f^{-1}(x) = (b-a)x + a$. From the definition of $T(m;a,b)$ and change of variable formula, we have for any $k \in \mathbb{N}$,

$$\begin{aligned}
T(m;a,b)(k) &= \int_{[0,1]} y^k \tilde{\mu}(dy) \\
&= \int I\{f(h(y)) \in [0,1]\} f(h(y))^k \tilde{\mu}(dy) \\
&= \int I\{f(x) \in [0,1]\} f(x)^k (\tilde{\mu} \circ f)(dx) \\
&= \int_{[a,b]} \left(\frac{x-a}{b-a}\right)^k \mu(dx) \\
&= \int \left(\frac{x-a}{b-a}\right)^k \mu(dx).
\end{aligned}$$

When $k = 0$, $m(0) = T(m;a,b)(0) = \int 1 \mu(dx)$. Suppose $m(i) = \int x^i \mu(dx)$ for $i = 0, ..., k$. We show $m(k+1) = \int x^{k+1} \mu(dx)$. By (S-4),

$$\begin{aligned}
m(k+1) &= (b-a)^{k+1} T(m;a,b)(k+1) - \sum_{i=0}^{k} \binom{k+1}{i} m(i)(-a)^{k+1-i} \\
&= \int (x-a)^{k+1} \mu(dx) - \sum_{i=0}^{k} \binom{k+1}{i} m(i)(-a)^{k+1-i} \\
&= \int \sum_{i=0}^{k+1} \binom{k+1}{i} x^k (-a)^{k+1-i} \mu(dx) - \int \sum_{i=0}^{k} \binom{k+1}{i} x^k (-a)^{k+1-i} \mu(dx) \\
&= \int x^{k+1} \mu(dx).
\end{aligned}$$

Thus, by induction, $m(k) = \int x^k \mu(dx)$ for $k = 0, 1, ...$, for μ defined by $\mu(A) = \tilde{\mu}(f(A))$.

Finally, for uniqueness, let μ_1 and μ_2 be two representing measures for m. From the first part of the proof, we see that both $\tilde{\mu}_1(A) := \mu_1(h(A))$ and $\tilde{\mu}_2(A) := \mu_2(h(A))$ are representing measures supported on $[0,1]$ for $T(m;a,b)$. Then $\tilde{\mu}_1 = \tilde{\mu}_2 = \tilde{\mu}$ from Theorem 1. Then for any measurable set E,

$$\mu_1(E) = \tilde{\mu}(h^{-1}(E)) = \mu_2(E).$$

Thus, the measure μ corresponding to m is unique. □

S4.2 Proof of Proposition 3

Proof. We show $\mathscr{M}_\infty(C)\cap\ell_2(\mathbb{Z})$ is a convex and closed subset of $\ell_2(\mathbb{Z})$. Convexity holds since for $p, q \in \mathscr{M}_\infty(C)\cap\ell_2(\mathbb{Z})$ where $p(k) = \int_{[-1+\delta,1-\delta]} x^{|k|} F_1(dx)$ and $q(k) = \int_{[-1+\delta,1-\delta]} x^{|k|} F_2(dx)$, we have $u = \alpha p + (1-\alpha)q \in \ell_2(\mathbb{Z})$ and $u(k) = \int_{[-1+\delta,1-\delta]} x^{|k|}(\alpha F_1 + (1-\alpha)F_2)(dx)$, i.e., $u \in \mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$.

Now, we show $\mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$ is closed. In the case C is a closed interval, then from Lemma 6, $\mathscr{M}_\infty(C)\cap\ell_2(\mathbb{Z})$ is closed. Otherwise for a general closed set C, consider a sequence of vectors $\{m_n\} \subseteq \mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$ where $\|m_n - f\| \to 0$ for some $f = \{f(k)\}_{k=-\infty}^{\infty}$. Now, let $a = \inf C$ and $b = \sup C$. Then $m_n \in \mathscr{M}_\infty([a,b]) \cap \ell_2(\mathbb{Z})$ so from Lemma 6, $f \in \mathscr{M}_\infty([a,b]) \cap \ell_2(\mathbb{Z})$. In particular, f is an $[a,b]$-moment sequence. Let μ_n denote the representing measure for μ_n and let μ denote the representing measure for f. We now show $f \in \mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$.

Suppose $x \in [a,b]$ and $x \notin C$. We show $x \notin \mathrm{Supp}(\mu)$. We show that there exists $\epsilon > 0$ such that $\mu(N_\epsilon(x)) = 0$ where $N_\epsilon(x) = \{y; |y-x| < \epsilon\}$. Since C is closed we can find an $\epsilon' > 0$ such that $N_{\epsilon'}(x) \cap C = \emptyset$. Take $\psi : \mathbb{R} \to [0,1]$ to be the continuous function with

$$\psi(y) = \begin{cases} 0 & |y-x| > 3\epsilon'/4 \\ 1 & |y-x| < \epsilon'/2 \\ 1 - (4/\epsilon')\{y - (x+\epsilon'/2)\} & x+\epsilon'/2 \le y \le x+3\epsilon'/4 \\ (4/\epsilon')\{y-(x-3\epsilon'/4)\} & x-3\epsilon'/4 \le y \le x-\epsilon'/2. \end{cases}$$

From Lemma 7, $0 \le \mu(N_{\epsilon'/2}(x)) \le \int \psi(y)\,\mu(dy) = \lim_{n\to\infty} \int \psi(y)\,\mu_n(dy) = \lim_{n\to\infty} 0 = 0$. Taking $\epsilon = \epsilon'/2$, we obtain $x \notin \mathrm{Supp}(\mu)$. Since x was arbitrary, $\mathrm{Supp}(\mu) \subseteq C$ and $f \in \mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$.

Since $\mathscr{M}_\infty(\delta) \cap \ell_2(\mathbb{Z})$ is a closed, convex subset of the Hilbert space $\ell_2(\mathbb{Z})$, the existence and uniqueness of $\Pi(r; C)$ follows from the Hilbert space projection theorem.

□

S4.3 Proof of Proposition 5

Proof. In the case $\mathrm{Supp}(\hat{\mu}_C) \cap (-1,1) = \emptyset$, the statement in the Proposition is trivially true. Otherwise, suppose $\mathrm{Supp}(\hat{\mu}_C) \cap (-1,1)$ is nonempty. Let $\tilde{\alpha} \in \mathrm{Supp}(\hat{\mu}_C) \cap (-1,1)$ given. We show $\langle \Pi(r;C), x_{\tilde{\alpha}} \rangle = \langle r, x_{\tilde{\alpha}} \rangle$.

First, we show that $\langle x_\alpha, \Pi(r;C) - r \rangle = 0$ for $\hat{\mu}_C$-almost every α. Let $E = C \cap (-1,1)$. From Lemma 2, we have $\hat{\mu}_C(\{-1,1\}) = 0$, and from the definition of $\hat{\mu}_C$, we have $\mathrm{Supp}(\hat{\mu}_C) \subset C$, so $\hat{\mu}_C(E^c) = \hat{\mu}_C(C^c \cup \{-1,1\}) \leq \hat{\mu}_C(C^c) + \hat{\mu}_C(\{-1,1\}) = 0$. From Proposition 4 and Lemma 4, we have

$$\begin{aligned} 0 = \langle \Pi(r;C), \Pi(r;C) - r \rangle &= \int \langle x_\alpha, \Pi(r;C) - r \rangle \, \hat{\mu}_C(d\alpha) \\ &= \int_E \langle x_\alpha, \Pi(r;C) - r \rangle \, \hat{\mu}_C(d\alpha) \end{aligned} \tag{S-5}$$

From Proposition 4, $\langle x_\alpha, \Pi(r;C) - r \rangle \geq 0$ for all $\alpha \in E$. Thus, from (S-5) and the fact $\hat{\mu}_C(E^c) = 0$, we have

$$\langle x_\alpha, \Pi(r;C) - r \rangle = 0, \quad \text{for } \hat{\mu}_C\text{-a.e. } \alpha. \tag{S-6}$$

Now, we complete the proof that $\langle \Pi(r;C), x_{\tilde{\alpha}} \rangle = \langle r, x_{\tilde{\alpha}} \rangle$. Let $\epsilon > 0$ given. Choose $R > 0$ such that $| \langle x_{\tilde{\alpha}} - x_\alpha, \Pi(r;C) - r \rangle | \leq \epsilon$ for all $\alpha \in N_R(\tilde{\alpha})$, where $N_R(\tilde{\alpha}) = \{\alpha : |\alpha - \tilde{\alpha}| < R\}$. Since

$$| \langle x_{\tilde{\alpha}} - x_\alpha, \Pi(r;C) - r \rangle | \leq \|x_{\tilde{\alpha}} - x_\alpha\| \|\Pi(r;C) - r\|$$

and

$$\lim_{\alpha \to \tilde{\alpha}} \|x_{\tilde{\alpha}} - x_\alpha\| = 0,$$

such a choice of $R > 0$ exists. Now, since $\tilde{\alpha} \in \mathrm{Supp}(\hat{\mu}_C)$ and $N_R(\tilde{\alpha})$ is open, we have

$\hat{\mu}_C(N_R(\tilde{\alpha})) > 0$, so

$$\begin{aligned}
\langle x_{\tilde{\alpha}}, \Pi(r;C) - r\rangle &= \{\hat{\mu}_C(N_R(\tilde{\alpha}))\}^{-1} \int_{N_R(\tilde{\alpha})} \langle x_{\tilde{\alpha}}, \Pi(r;C) - r\rangle \, \hat{\mu}_C(d\alpha) \\
&= \{\hat{\mu}_C(N_R(\tilde{\alpha}))\}^{-1} \int_{N_R(\tilde{\alpha})} \langle x_{\alpha} + x_{\tilde{\alpha}} - x_{\alpha}, \Pi(r;C) - r\rangle \, \hat{\mu}_C(d\alpha) \\
&= \{\hat{\mu}_C(N_R(\tilde{\alpha}))\}^{-1} \int_{N_R(\tilde{\alpha})} \langle x_{\alpha}, \Pi(r;C) - r\rangle \, \hat{\mu}_C(d\alpha) \\
&+ \{\hat{\mu}_C(N_R(\tilde{\alpha}))\}^{-1} \int_{N_R(\tilde{\alpha})} \langle x_{\tilde{\alpha}} - x_{\alpha}, \Pi(r;C) - r\rangle \, \hat{\mu}_C(d\alpha) \\
&= \{\hat{\mu}_C(N_R(\tilde{\alpha}))\}^{-1} \int_{N_R(\tilde{\alpha})} \langle x_{\tilde{\alpha}} - x_{\alpha}, \Pi(r;C) - r\rangle \, \hat{\mu}_C(d\alpha)
\end{aligned}$$

where we used (S-6). From the choice of R, we have $|\langle x_{\tilde{\alpha}} - x_{\alpha}, \Pi(r;C) - r\rangle| \le \epsilon$ for all $\alpha \in N_R(\tilde{\alpha})$, and thus

$$-\epsilon \le \langle x_{\tilde{\alpha}}, \Pi(r;C) - r\rangle \le \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have $\langle x_{\tilde{\alpha}}, \Pi(r;C) - r\rangle = 0$ and thus $\langle \Pi(r;C), x_{\tilde{\alpha}}\rangle = \langle r, x_{\tilde{\alpha}}\rangle$. This proves the result. □

S4.4 Proof of Proposition 6

Proof. Define $g(\alpha) = \langle x_{\alpha}, r - \Pi(r;C)\rangle = \sum_{k\in\mathbb{Z}} \alpha^{|k|}\{r(k) - \Pi(r;C)(k)\}$. We consider derivatives of g, i.e., for $n \ge 1$,

$$g^{(n)}(\alpha) = \frac{d^n}{d\alpha^n} g(\alpha) = \frac{d^n}{d\alpha^n} \sum_{k\in\mathbb{Z}} \alpha^{|k|}\{r(k) - \Pi(r;C)(k)\}$$

We first show that the term-by-term differentiation of $g(\alpha)$ is justified, so that

$$g^{(1)}(\alpha) = \frac{d}{d\alpha} g(\alpha) = \sum_{k\in\mathbb{Z}:|k|\ge 1} |k|\alpha^{|k|-1}\{r(k) - \Pi(r;C)(k)\} \tag{S-7}$$

and, similarly,

$$g^{(n)}(\alpha) = \frac{d^n}{d\alpha^n} g(\alpha) = \sum_{k\in\mathbb{Z}:|k|\ge n} \frac{|k|!}{(|k|-n)!}\alpha^{|k|-n}\{r(k) - \Pi(r;C)(k)\}.$$

We first consider the case when $n = 1$. Let $\alpha_0 \in (-1, 1)$ be arbitrary. Let $\tilde{\alpha} = (|\alpha_0| + 1)/2$. Then $|\alpha_0| < \tilde{\alpha} < 1$. Take $\beta = \tilde{\alpha} - |\alpha_0| = 1 - |\alpha_0|/2$ and define $N_\beta(\alpha_0) = \{\alpha' :$

$|\alpha' - \alpha_0| \leq \beta\}$. We will show that the term by term differentiation of $g(\alpha)$ at $\alpha = \alpha_0$ is justifiable, by showing that each summand in (S-7) for $\alpha \in N_\beta(\alpha_0)$ is dominated by some absolutely summable $\tilde{g}_1(k)$.

For $k \in \mathbb{Z}$ and $\alpha \in (-1, 1)$, define

$$g_1(k, \alpha) = \frac{d[\alpha^{|k|}\{r(k) - \Pi(r; C)(k)\}]}{d\alpha} = \begin{cases} |k|\alpha^{|k|-1}\{r(k) - \Pi(r; C)(k)\} & k \geq 1 \\ 0 & k = 0, \end{cases}$$

and for $k \in \mathbb{Z}$, define $\tilde{g}_1(k) = |g_1(k, \tilde{\alpha})|$. Then $|g_1(k, \alpha)| \leq \tilde{g}_1(k)$ for all $\alpha \in N_\beta(\alpha_0)$.

Define the sequence $\Delta = \{\Delta(k)\}_{k=-\infty}^{\infty}$ by

$$\Delta(k) = |r(k) - \Pi(r; C)(k)|, \; k \in \mathbb{Z}.$$

Since $r(k) = 0$ for $|k| > M - 1$, we have $r \in \ell_2(\mathbb{Z})$, and so $\Pi(r; C) \in \ell_2(\mathbb{Z})$ also. Thus $\Delta \in \ell_2(\mathbb{Z})$, since

$$\|\Delta\|^2 = \sum_{k \in \mathbb{Z}} \Delta(k)^2 = \sum_{|k| \leq M-1} \{r(k) - \Pi(r; C)(k)\}^2 + \sum_{|k| > M-1} \Pi(r; C)(k)^2 < \infty. \tag{S-8}$$

For $n \geq 0$ and $\alpha \in (-1, 1)$, define the sequence $\tilde{x}_{\alpha,n} = \{\tilde{x}_{\alpha,n}(k)\}_{k=-\infty}^{\infty}$ by

$$\tilde{x}_{\alpha,n}(k) = \frac{d^n\{x_\alpha(k)\}}{d\alpha^n} = \begin{cases} \frac{|k|!}{(|k|-n)!}\alpha^{|k|-n} & |k| \geq n \\ 0 & |k| < n. \end{cases}$$

We have

$$\sum_{k \in \mathbb{Z}} \tilde{x}_{\alpha,n}^2 = 2\sum_{k \geq n} \frac{k!}{(k-n)!}|\alpha|^{k-n} = 2\sum_{l \geq 0} \frac{(l+n)!}{l!}|\alpha|^l < \infty, \tag{S-9}$$

for each $\alpha \in (-1, 1)$, $n \geq 0$, so $\tilde{x}_{\alpha,n} \in \ell_2(\mathbb{Z})$ for any $\alpha \in (-1, 1)$, $n \geq 0$. Therefore by (S-8) and (S-9), we can conclude that $\tilde{g}_1$ is absolutely summable,

$$\sum_{k \in \mathbb{Z}} \tilde{g}_1(k) = \langle \tilde{x}_{\tilde{\alpha},1}, \Delta \rangle \leq \|\tilde{x}_{\tilde{\alpha},1}\|\|\Delta\| < \infty.$$

Then, by the Lebesgue differentiation theorem, we have

$$g^{(1)}(\alpha_0) = \sum_{k\in\mathbb{Z}} g_1(k,\alpha_0) = \sum_{k\in\mathbb{Z}:|k|\geq 1} |k|\alpha_0^{|k|-1}\{r(k) - \Pi(r;C)(k)\}$$

[see, e.g., Theorem 2.27 in Folland, 1999]. Since $\alpha_0 \in (-1,1)$ was arbitrary, we have $g^{(1)}(\alpha) = \sum_{k\in\mathbb{Z}:|k|\geq 1} |k|\alpha^{|k|-1}\{r(k) - \Pi(r;C)(k)\}$ for each $\alpha \in (-1,1)$.

Proceeding similarly, we obtain

$$g^{(n)}(\alpha) = \frac{d^n}{d\alpha^n} g(\alpha) = \sum_{k\in\mathbb{Z}:|k|\geq n} \frac{|k|!}{(|k|-n)!}\alpha^{|k|-n}\{r(k) - \Pi(r;C)(k)\}.$$

for $\alpha \in (-1,1)$, $n \geq 1$.

We now show that $\hat{\mu}_C$ has finite support. Recall for $|k| > M-1$, $r(k) = 0$, so for $n > (M-1)$ we have

$$\begin{aligned} g^{(n)}(\alpha) &= -2\sum_{k=n}^{\infty} \frac{k!}{(k-n)!}\alpha^{k-n}\Pi(r;C)(k) \\ &= -2\sum_{k=n}^{\infty} \frac{k!}{(k-n)!}\alpha^{k-n}\int_{[-1,1]} x^k \hat{\mu}_C(dx). \end{aligned}$$

For $|\rho| < 1$, we have

$$\sum_{k=n}^{\infty} \frac{k!}{(k-n)!}\rho^{k-n} = \frac{n!}{(1-\rho)^{(n+1)}}. \tag{S-10}$$

Recall $n > (M-1)$. From Lemma 2, we have $\hat{\mu}_C(\{-1,1\}) = 0$ since $\Pi(r;C) \in \mathscr{M}_\infty(C) \cap \ell_2(\mathbb{Z})$. Thus, by Fubini's theorem,

$$\begin{aligned} &\sum_{k=n}^{\infty}\int_{[-1,1]} \frac{k!}{(k-n)!}|\alpha|^{k-n}|x|^k\hat{\mu}_C(dx) \\ &= \sum_{k=n}^{\infty}\int_{(-1,1)} \frac{k!}{(k-n)!}|\alpha|^{k-n}|x|^k\hat{\mu}_C(dx) \\ &= \int_{(-1,1)}\sum_{k=n}^{\infty} \frac{k!}{(k-n)!}|\alpha|^{k-n}|x|^k\hat{\mu}_C(dx) \\ &= \int_{(-1,1)}|x|^n\sum_{k=n}^{\infty} \frac{k!}{(k-n)!}|\alpha x|^{k-n}\hat{\mu}_C(dx) \\ &= \int_{(-1,1)}\frac{n!|x|^n}{(1-|\alpha x|)^{n+1}}\hat{\mu}_C(dx) < \infty \end{aligned}$$

for $|\alpha| < 1$ where the last equality is due to (S-10). Thus, the integral and summation in $g^{(n)}(\alpha)$ may be interchanged, so that

$$g^{(n)}(\alpha) = -2 \int_{(-1,1)} \frac{n! x^n}{(1-\alpha x)^{n+1}} \hat{\mu}_C(dx). \tag{S-11}$$

Now, we consider two subcases. In the first subcase, $\text{Supp}(\hat{\mu}_C) \subset \{0\}$ (that is, $\hat{\mu}_C$ is the null measure, or $\hat{\mu}_C$ puts point mass on 0 only.) Then $\text{Supp}(\hat{\mu}_C) \cap (-1,1)$ contains at most a single point. Otherwise, take n to be the smallest even number such that $n > (M-1)$. Then since the support of $\hat{\mu}_C$ contains points away from 0, $g^{(n)}(\alpha) < 0$ for all $\alpha \in (-1,1)$, since for even n the integrand in $g^{(n)}(\alpha)$ in (S-11) is positive for $x \neq 0$. Since $g^{(n)}(\alpha) \neq 0$ for all $\alpha \in (-1,1)$, there exist at most n points $-1 < \alpha_1 < \alpha_2 < ... < \alpha_n < 1$ such that $g(\alpha_i) = 0$. Thus $\text{Supp}(\hat{\mu}_C) \cap (-1,1)$ contains at most n points, where n is the smallest even number with $n > M-1$. Since $\hat{\mu}_C$ has a finite number of support points in $(-1,1)$, we have $\sup\{|x| : x \in (-1,1) \cap \text{Supp}(\hat{\mu}_C)\} < 1 - \epsilon$ for some $\epsilon > 0$ and thus $\{-1,1\} \cap \text{Supp}(\hat{\mu}_C) = \emptyset$. □

S5 Proofs for results in Section 4

S5.1 Proof of Proposition 7

First we show that (R.1)-(R.3) hold for the empirical autocovariance sequence $\tilde{r}_M$. The convergence in (R.1) is shown in Lemma 8 which is presented at the end of this proof. Assumption (R.2) holds from the definition of $\tilde{r}_M$ in (4), with the choice $n(M) = M$, and the symmetry in (R.3) also follows from the definition of $\tilde{r}_M(k)$ in (4) as (4) depends on k only through $|k|$. Finally, we show that

$$|\tilde{r}_M(k)| \leq \tilde{r}_M(0) \tag{S-12}$$

By symmetry, it is sufficient to prove the result for $k \in \mathbb{N}$. For notational simplicity, let $h(x) = g(x) - Y_M$. First, we consider $0 \leq k \leq M-1$. We define M length M vectors $v_j \in \mathbb{R}^M$, $j = 0, ..., M-1$ such that

$$v_0 = [h(X_0), h(X_1), \ldots, h(X_{M-1})],$$

and for $k = 1, \ldots, M-1$,

$$v_k = [h(X_k), h(X_{k+1}), \ldots, h(X_{M-1}), 0, 0, \ldots, 0].$$

Then, for $0 \leq k \leq M-1$, $\|v_k\| \leq \|v_0\|$ by the definition of the v_k's. Also, note that $\tilde{r}_M(k) = M^{-1} \langle v_0, v_k \rangle$ by the definition of empirical autocovariance. We have

$$|\tilde{r}_M(k)| = |\frac{1}{M} \langle v_0, v_k \rangle| \leq \frac{1}{M} \|v_0\| \|v_k\| \leq \frac{1}{M} \|v_0\|^2 = \tilde{r}_M(0).$$

Additionally, $\tilde{r}_M(k) = 0 \leq \|v_0\|^2 = \tilde{r}_M(0)$ for k with $|k| > (M-1)$.

Now, we argue that a windowed autocovariance $\check{r}_M(k) = w_M(|k|)\tilde{r}_M(k)$ satisfying conditions (W.1)-(W.3) satisfies (R.1)-(R.3). First of all, the symmetry in (R.3) holds since $\check{r}_M(-k) = w_M(|k|)\tilde{r}_M(-k) = w_M(|k|)\tilde{r}_M(k) = \check{r}_M(k), \forall k \in \mathbb{Z}$. Also, $\check{r}_M(0) = \tilde{r}_M(0) \geq |\tilde{r}_M(k)| \geq |\tilde{r}_M(k)||w_M(|k|)| = |\check{r}_M(k)|, \forall k \in \mathbb{Z}$ by (S-12), conditions (W.1) and (W.2). Assumption (R.2) holds with the choice $n(M) = \min\{b_M, M\}$ by conditions (W.3), and assumption (R.1) follows from Lemma 8, (W.3), and Slutsky's theorem.

Lemma 8. *Assume (A.1), (A.2) and (B.1). Let $k \in \mathbb{Z}$ given. Then*

$$\lim_{M \to \infty} \tilde{r}_M(k) = \gamma(k), \ P_x\text{-a.s.}$$

for each $x \in \mathsf{X}$.

Proof. First, we show that $\bar{r}_M(k) \overset{a.s.}{\to} \gamma(k)$ as $M \to \infty$, where $\bar{r}_M(k) = M^{-1} \sum_{t=0}^{M-1-|k|} \bar{g}(X_t)\bar{g}(X_{t+|k|})$ where we define $\bar{g}(x) = g(x) - E_\pi[g(X_0)]$. Without loss of generality, assume $k > 0$. We define $h_k : \Omega \to \mathbb{R}$ such that $h_k(\omega) = \bar{g}(X_0(\omega))\bar{g}(X_k(\omega))$ for $k \geq 0$. Note $h_k \in L_1(\Omega, \mathcal{F}, P_\pi)$ by (B.1). Let $\theta : \Omega \to \Omega$ be the shift operator. We first want to show that

$$\bar{r}_M(k) = \frac{1}{M} \sum_{t=0}^{M-1-k} \bar{g}(X_t)\bar{g}(X_{t+k}) = \frac{1}{M} \sum_{t=0}^{M-1-k} \theta^t h_k \to \gamma(k)$$

P_x-almost surely, for any initial condition $x \in \mathsf{X}$.

Using the fact that the set of P_π-invariant events is trivial due to X being Harris recurrent and from Theorem 17.1.2 in Meyn and Tweedie [2009], we have a set F_h of full

π-measure such that for any initial condition in $x \in F_h$,

$$\lim_{M\to\infty} \frac{1}{M} \sum_{t=0}^{M-1-k} \theta^t h_k = \lim_{M\to\infty} \frac{M-k}{M} \frac{1}{M-k} \sum_{t=0}^{M-1-k} \theta^t h_k = E_\pi[h_k] = \gamma(k), \text{ a.s. } P_x. \quad \text{(S-13)}$$

Now via a modification of Proposition 17.1.6 in Meyn and Tweedie [2009], we show (S-13) holds for all $x \in \mathsf{X}$. Define $h_\infty(x) = P_x(\lim_{M\to\infty} M^{-1} \sum_{t=0}^{M-1-k} \bar{g}(X_t)\bar{g}(X_{t+k}) = \gamma(k))$. We know that $h_\infty(x) = 1$ for $x \in F_h$. If we show $h_\infty(x) = 1$ for all $x \in \mathsf{X}$, we have the desirable result. We show that $h_\infty(x)$ is harmonic.

$$\begin{aligned}
&Qh_\infty(x) \\
&= E_x[P_{X_1}\{\lim_{M\to\infty} M^{-1} \sum_{t=0}^{M-1-k} \bar{g}(X_t)\bar{g}(X_{t+k}) = \gamma(k)\}] \\
&= E_x[P_x\{\lim_{M\to\infty} M^{-1} \sum_{t=0}^{M-1-k} \bar{g}(X_{t+1})\bar{g}(X_{t+1+k}) = \gamma(k)|\mathcal{F}_1\}] \\
&= E_x[P_x\{\lim_{M\to\infty} \left[\frac{M+1}{M}\frac{1}{M+1} \sum_{t=0}^{M-k} \bar{g}(X_t)\bar{g}(X_{t+k}) - \frac{1}{M}\bar{g}(X_0)\bar{g}(X_k)\right] = \gamma(k)\}] \\
&= h_\infty(x).
\end{aligned}$$

Therefore $h_\infty(x) = 1$ for any $x \in \mathsf{X}$, and (S-13) holds for any initial condition $x \in \mathsf{X}$.

Finally, we show that $\tilde{r}_M(k) - \bar{r}_M(k) \to 0$ as $M \to \infty$, P_x-a.s., for all $x \in \mathsf{X}$, where $\tilde{r}_M(k) = M^{-1} \sum_{t=0}^{M-1-k} (g(X_t) - Y_M)(g(X_{t+k}) - Y_M)$ for $Y_M = \sum_{t=0}^{M-1} g(X_t)$. First we have $Y_M \to \mu$, P_x-almost surely for all $x \in \mathsf{X}$ by SLLN in Theorem 17.1.7 in Meyn and Tweedie [2009]. For any $k \in \mathbb{N}$, we have,

$$\begin{aligned}
\tilde{r}_M(k) - \bar{r}_M(k) &= \frac{1}{M} \sum_{t=0}^{M-1-k} \{(g(X_t) - Y_M)(g(X_{t+k}) - Y_M) - (g(X_t) - \mu)(g(X_{t+k}) - \mu)\} \\
&= \frac{1}{M} \sum_{t=0}^{M-1-k} \{(Y_M - \mu)(g(X_{t+k}) + g(X_t)) + \mu^2 - Y_M^2\} \\
&= (Y_M - \mu)\frac{1}{M} \sum_{t=0}^{M-1-k} \{g(X_{t+k}) + g(X_t)\} + \frac{M-k}{M}(\mu^2 - Y_M^2). \qquad \text{(S-14)}
\end{aligned}$$

Since both $\frac{1}{M-k} \sum_{t=0}^{M-1-k} g(X_{t+k})$ and $\frac{1}{M-k} \sum_{t=0}^{M-1-k} g(X_t)$ converge to μ by SLLN in Theorem 17.1.7 in Meyn and Tweedie [2009], $\frac{1}{M} \sum_{t=0}^{M-1-k} \{g(X_{t+k}) + g(X_t)\} \to 2\mu$, P_x-almost surely for any $x \in \mathsf{X}$. Therefore, by continuous mapping theorem, (S-14) $\to 0$ as

$M \to \infty$, P_x-almost surely for any $x \in \mathsf{X}$, which proves the result. □

S5.2 Proof of Proposition 8

Proof. First, we show that

$$|\langle x_\alpha, r_M - \gamma \rangle| \to 0 \tag{S-15}$$

P_x-a.s. for any $x \in \mathsf{X}$, for any $\alpha \in (-1, 1)$. For the ease of notation, if a P_x-almost sure convergence holds for any $x \in \mathsf{X}$, we will just say the convergence holds almost surely. Let $\epsilon > 0$ given. Note for any $B > 0$,

$$\langle x_\alpha, r_M - \gamma \rangle = \sum_{k=-(B-1)}^{(B-1)} \alpha^{|k|}\{r_M(k) - \gamma(k)\} + 2\sum_{k=B}^{\infty} \alpha^{|k|}\{r_M(k) - \gamma(k)\}.$$

since $r_M(k) = r_M(-k)$ by (R.3). Choose B such that

$$\sum_{k=B}^{\infty} |\alpha|^k \gamma(0) = \frac{|\alpha|^B}{1 - |\alpha|}\gamma(0) \le \epsilon/4.$$

Then

$$\begin{aligned}
&\limsup_{M\to\infty} |\sum_{k=B}^{\infty} \alpha^k\{r_M(k) - \gamma(k)\}| \\
&\le \limsup_{M\to\infty} \sum_{k=B}^{\infty} |\alpha|^k\{|r_M(k)| + |\gamma(k)|\} \\
&\le \limsup_{M\to\infty} \sum_{k=B}^{\infty} |\alpha|^k\{r_M(0) + \gamma(0)\} \\
&= \epsilon/2
\end{aligned}$$

where the second inequality uses $|r_M(k)| \le r_M(0)$ in (R.3) and the equality uses $r_M(0) \overset{a.s.}{\to} \gamma(0)$ by (R.1).

Furthermore, we have

$$\sum_{k=-(B-1)}^{(B-1)} \alpha^{|k|}\{r_M(k) - \gamma(k)\} \overset{a.s.}{\to} 0$$

since $r_M(k) \overset{a.s.}{\to} \gamma(k)$ for each $k \in \mathbb{Z}$. Thus

$$\limsup_{M\to\infty} |\langle x_\alpha, r_M - \gamma\rangle| \le \epsilon.$$

Since $\epsilon > 0$ was arbitrary, we have $\langle x_\alpha, r_M - \gamma\rangle \overset{a.s.}{\to} 0$ as $M \to \infty$. This proves the a.s. convergence result for each $\alpha \in (-1, 1)$.

Now, we show that the convergence is uniform over $\mathcal{K}$. First, let δ_0 denote the minimum distance between $\mathcal{K}$ and $\{-1, 1\}$, i.e., $\delta_0 = \inf\{\min(|1 - x|, |-1 - x|) : x \in \mathcal{K}\}$. Since $\mathcal{K} \subset (-1, 1)$, the gap $\delta_0 > 0$. Since for $x \in \mathcal{K}$, $x \in (-1, 1)$, we have $\delta_0 \le 1$. If $\delta_0 = 1$, then $\mathcal{K} = \{0\}$ since $\mathcal{K}$ is nonempty by assumption, and $\sup_{\alpha\in\mathcal{K}} |\langle x_\alpha, \Pi_\delta(r_M) - \gamma\rangle| \overset{a.s.}{\to} 0$ from the previously shown convergence for each $\alpha \in (-1, 1)$.

Otherwise, suppose $\delta_0 < 1$. Let $\epsilon_1 > 0$ be given, and choose $\beta > 0$ such that

$$\beta = \epsilon_1/(4\delta_0^2\gamma(0)). \tag{S-16}$$

For $\alpha \in (-1, 1)$, define $B_\beta(\alpha) = \{|x - \alpha| \le \beta\} \cap [-1 + \delta_0, 1 - \delta_0]$. Take $N(\beta) = \lceil 2(1 - \delta)/\beta \rceil$ and define $\alpha_j = (-1 + \delta_0) + j\beta$, $j = 0, ..., N(\beta) - 1$. Then $\mathcal{K} \subset \cup_{j=0}^{N(\beta)-1} B_\beta(\alpha_j)$ and

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{K}} |\langle x_\alpha, r_M - \gamma\rangle| \\
&\le \max_{j=0,...,N(\beta)-1} \sup_{\alpha\in B_\beta(\alpha_j)} |\langle x_\alpha, r_M - \gamma\rangle| \\
&\le \max_{j=0,...,N(\beta)-1} \sup_{\alpha\in B_\beta(\alpha_j)} |\langle x_\alpha - x_{\alpha_j}, r_M - \gamma\rangle| + |\langle x_{\alpha_j}, r_M - \gamma\rangle| \\
&\le \max_{j=0,...,N(\beta)-1} \sup_{\alpha\in B_\beta(\alpha_j)} |\langle x_\alpha - x_{\alpha_j}, r_M - \gamma\rangle| + \max_{j=0,...,N(\beta)-1} |\langle x_{\alpha_j}, r_M - \gamma\rangle|
\end{aligned} \tag{S-17}$$

From convergence (S-15), we have

$$\limsup_{M\to\infty} \max_{j=0,...,N(\beta)-1} |\langle x_{\alpha_j}, r_M - \gamma\rangle| = 0 \tag{S-18}$$

almost surely since $N(\beta)$ is finite.

To control the size of the first term by the distance between α and α_i, we have the following Lemma, the proof of which is deferred to the end of this proof.

Lemma 9. *For any r such that $r(0) \ge 0$, $r(k) = r(-k)$, and $|r(k)| \le r(0)$ for $k \in \mathbb{Z}$, and*

$\alpha, \beta \in (-1, 1)$, *we have*

$$|\langle r, x_\alpha - x_\beta \rangle| \leq \frac{2r(0)}{(1-|\alpha|)(1-|\beta|)}|\alpha - \beta|.$$

From Lemma 9, we have, for $\alpha \in B_\beta(\alpha_j)$,

$$\begin{aligned}
|\langle x_\alpha - x_{\alpha_j}, r_M - \gamma \rangle| &= |\langle x_\alpha - x_{\alpha_j}, r_M \rangle - \langle x_\alpha - x_{\alpha_j}, \gamma \rangle| \\
&\leq |\langle x_\alpha - x_{\alpha_j}, r_M \rangle| + |\langle x_\alpha - x_{\alpha_j}, \gamma \rangle| \\
&\leq \frac{2|\alpha - \alpha_j|}{(1-|\alpha|)(1-|\alpha_j|)}(r_M(0) + \gamma(0)) \\
&\leq 2(\beta/\delta_0^2)\{r_M(0) + \gamma(0)\}.
\end{aligned}$$

Since this bound does not depend on j, we have,

$$\begin{aligned}
&\limsup_{M\to\infty} \max_{j=0,\ldots,N(\beta)-1} \sup_{\alpha \in B_\beta(\alpha_j)} |\langle x_\alpha - x_{\alpha_j}, r_M - \gamma \rangle| \\
&\leq \limsup_{M\to\infty} 2(\beta/\delta_0^2)\{r_M(0) + \gamma(0)\} \\
&= 4\beta\delta_0^{-2}\gamma(0) \qquad \text{(S-19)}
\end{aligned}$$

since $r_M(0) \overset{a.s.}{\to} \gamma(0)$. Thus, from (S-17), (S-18), and (S-19), we have

$$\begin{aligned}
&\limsup_{M\to\infty} \sup_{\alpha \in \mathcal{K}} |\langle x_\alpha, r_M - \gamma \rangle| \\
&\leq \limsup_{M\to\infty} \max_{j=0,\ldots,N(\beta)-1} \sup_{\alpha \in B_\beta(\alpha_j)} |\langle x_\alpha - x_{\alpha_j}, r_M - \gamma \rangle| \\
&+ \limsup_{M\to\infty} \max_{j=0,\ldots,N(\beta)-1} |\langle x_{\alpha_j}, r_M - \gamma \rangle| \\
&\leq 4\beta\delta_0^{-2}\gamma(0) = \epsilon_1.
\end{aligned}$$

where the last equality is due to the choice of β in (S-16). But ϵ_1 was arbitrary, so $\lim_{M\to\infty} \sup_{\alpha \in \mathcal{K}} |\langle x_\alpha, r_M - \gamma \rangle| = 0$ almost surely. This proves the result.

□

Now we present the proof for Lemma 9.

Proof of Lemma 9. By definition,

$$\begin{aligned}|\langle r_M, x_\alpha - x_\beta\rangle| &= |2\sum_{k=1}^{\infty} r(k)\{\alpha^k - \beta^k\}| \\ &\leq 2r(0)\sum_{k=1}^{\infty}|\alpha^k - \beta^k|\end{aligned}$$

where the second inequality uses the fact that $\max_{k\geq 1}|r(k)| \leq r(0)$. Using the following equality:

$$\alpha^k - \beta^k = (\alpha - \beta)\sum_{j=1}^{k}\alpha^{k-j}\beta^{j-1}$$

we have,

$$\begin{aligned}|\langle r, x_\alpha - x_\beta\rangle| &\leq 2r(0)\sum_{k=1}^{\infty}|(\alpha-\beta)\sum_{j=1}^{k}\alpha^{k-j}\beta^{j-1}| \\ &\leq 2r(0)|\alpha-\beta|\sum_{k=1}^{\infty}\sum_{j=1}^{k}|\alpha^{k-j}\beta^{j-1}| \\ &= 2r(0)|\alpha-\beta|\sum_{j=1}^{\infty}\sum_{k=j}^{\infty}|\alpha|^{k-j}|\beta|^{j-1} \\ &= 2r(0)|\alpha-\beta|\frac{1}{(1-|\alpha|)(1-|\beta|)}\end{aligned}$$

where the last equality follows from

$$\sum_{j=1}^{\infty}\sum_{k=j}^{\infty}|\alpha|^{k-j}|\beta|^{j-1} = \sum_{j=1}^{\infty}|\beta|^{j-1}\sum_{k=j}^{\infty}|\alpha|^{k-j} = \frac{1}{(1-|\alpha|)(1-|\beta|)}.$$

□

S5.3 Proof of Proposition 9

Proof. By Lemma 4 and Proposition 6, we have,

$$\begin{aligned}\langle \Pi_\delta(r_M), \Pi_\delta(r_M)\rangle &= \int_{[-1,1]} \langle \Pi_\delta(r_M), x_\alpha\rangle \, \hat{\mu}_{\delta,M}(d\alpha) \\ &= \sum_{\alpha\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \langle \Pi_\delta(r_M), x_\alpha\rangle \, \hat{\mu}_{\delta,M}(\{\alpha\}) \\ &= \sum_{\alpha\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \langle r_M, x_\alpha\rangle \, \hat{\mu}_{\delta,M}(\{\alpha\}) \end{aligned} \tag{S-20}$$

where the last equality is due to Proposition 5. On the one hand,

$$\begin{aligned}\langle \Pi_\delta(r_M), \Pi_\delta(r_M)\rangle &= \sum_{\alpha\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \langle \Pi_\delta(r_M), x_\alpha\rangle \, \hat{\mu}_{\delta,M}(\{\alpha\}) \\ &= \sum_{\alpha,\alpha'\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \langle x_\alpha, x_{\alpha'}\rangle \, \hat{\mu}_{\delta,M}(\{\alpha\})\hat{\mu}_{\delta,M}(\{\alpha'\}) \\ &\geq \inf_{\alpha,\alpha'\in[-1+\delta,1-\delta]} \langle x_\alpha, x_{\alpha'}\rangle \sum_{\alpha,\alpha'\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \hat{\mu}_{\delta,M}(\{\alpha\})\hat{\mu}_{\delta,M}(\{\alpha'\}) \end{aligned} \tag{S-21}$$

since $\langle x_\alpha, x_{\alpha'}\rangle = \frac{1+\alpha\alpha'}{1-\alpha\alpha'} > 0$ for any $\alpha, \alpha' \in (-1,1)$. On the other hand, we have from (S-20) that

$$\langle \Pi_\delta(r_M), \Pi_\delta(r_M)\rangle \leq \sup_{\alpha\in[-1+\delta,1-\delta]} |\langle r_M, x_\alpha\rangle| \sum_{\alpha\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \hat{\mu}_{\delta,M}(\{\alpha\}). \tag{S-22}$$

Thus, from (S-21) and (S-22), we have

$$\begin{aligned}&\inf_{\alpha,\alpha'\in[-1+\delta,1-\delta]} \langle x_\alpha, x_{\alpha'}\rangle \sum_{\alpha,\alpha'\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \hat{\mu}_{\delta,M}(\{\alpha\})\hat{\mu}_{\delta,M}(\{\alpha'\}) \\ &\leq \sup_{\alpha\in[-1+\delta,1-\delta]} |\langle r_M, x_\alpha\rangle| \sum_{\alpha\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \hat{\mu}_{\delta,M}(\{\alpha\}).\end{aligned}$$

That is,

$$\sum_{\alpha\in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \hat{\mu}_{\delta,M}(\{\alpha\}) \leq \frac{\sup_{\alpha\in[-1+\delta,1-\delta]} |\langle r_M, x_\alpha\rangle|}{\inf_{\alpha,\alpha'\in[-1+\delta,1-\delta]} \langle x_\alpha, x_{\alpha'}\rangle}. \tag{S-23}$$

The denominator is deterministic, and we let $C_0 := \inf_{\alpha,\alpha' \in [-1+\delta,1-\delta]} \langle x_\alpha, x_{\alpha'} \rangle$. Now we show that the numerator is bounded almost surely. We have,

$$\sup_{\alpha \in [-1+\delta,1-\delta]} |\langle r_M, x_\alpha \rangle| \leq \sup_{\alpha \in [-1+\delta,1-\delta]} |\langle r_M - \gamma, x_\alpha \rangle| + \sup_{\alpha \in [-1+\delta,1-\delta]} |\langle \gamma, x_\alpha \rangle|$$

The second term $\sup_{\alpha \in [-1+\delta,1-\delta]} |\langle \gamma, x_\alpha \rangle|$ is deterministic and bounded by $\gamma(0)(2-\delta)/\delta$ from Holder's inequality. For the first term, from Proposition 8, we have

$$\limsup_{M \to \infty} \sup_{\alpha \in [-1+\delta,1-\delta]} |\langle r_M - \gamma, x_\alpha \rangle| = 0$$

almost surely. Define $C_{\delta,\gamma} = \frac{\gamma(0)(2-\delta)}{(\delta)(C_0)}$. Then

$$\begin{aligned} &\limsup_{M \to \infty} \hat{\mu}_{\delta,M}([-1+\delta, 1-\delta]) \\ &= \limsup_{M \to \infty} \sum_{\alpha \in \mathrm{Supp}(\hat{\mu}_{\delta,M})} \hat{\mu}_{\delta,M}(\{\alpha\}) \\ &\leq \frac{\gamma(0)(2-\delta)}{(\delta)(C_0)} = C_{\delta,\gamma} \end{aligned}$$

almost surely by (S-23). □

S6 Proofs for results in Section 5

Proposition 11. *Suppose $X_0, X_1, ...,$ is a Markov chain with transition kernel Q satisfying (A.1)-(A.3), and suppose $g : \mathsf{X} \to \mathbb{R}$ satisfies (B.1). Let γ denote the autocovariance sequence as defined in Proposition 1, and let F denote the representing measure for γ. Assume that $\gamma(0) = \mathrm{Var}_\pi(g(X_0)) > 0$. Let $\rho(k) = \gamma(k)/\gamma(0)$, $k \in \mathbb{Z}$ denote the autocorrelation sequence and $\hat{\rho}_M(k) = \tilde{r}_M(k)/\tilde{r}_M(0)$ denote the empirical autocorrelation sequence. Define δ_γ such that $\delta_\gamma = 1 - \sup\{|x|; x \in \mathrm{Supp}(F)\}$ where F is the representing measure for γ. Let*

$$\hat{\delta}_M = 1 - \exp\{-\log M/(2\hat{m})\}, \tag{S-24}$$

with $\hat{\delta}_M := 1$ in the case $\hat{m} = 0$. Suppose in addition to (A.1)-(A.3) and (B.1) that

$$\sup_{k=0,...,M-1} |\hat{\rho}_M(k) - \rho(k)| = O_{P_x}(\sqrt{\frac{\log M}{M}}) \tag{S-25}$$

with respect to the Markov chain law P_x for each $x \in \mathsf{X}$. Choose $\hat{m}$ such that

$$\hat{m} = \min\{t \in 2\mathbb{N}; \hat{\rho}_M(t+2) \leq c_M \sqrt{\frac{\log M}{M}}\} \tag{S-26}$$

for some $c_M \geq 0$. Then we have $\hat{\delta}_M$ is asymptotically not larger than δ_γ, i.e., for any $\epsilon > 0$, we have $\lim_{M\to\infty} P_x(\hat{\delta}_M \geq \delta_\gamma + \epsilon) = 0$.

Furthermore, under the assumption of $c_M \to \infty$ such that $c_M = O(\log M)$, $\hat{\delta}_M$ converges in P_x-probability to δ_γ, i.e., for any $\epsilon > 0$, we have $\lim_{M\to\infty} P_x(|\hat{\delta}_M - \delta_\gamma| \geq \epsilon) = 0$.

Proof. Define $\tilde{F}$ such that

$$\tilde{F}((-\infty, t]) = \begin{cases} 0 & t < 0 \\ \gamma(0)^{-1} F([-t, t]) & t \geq 0 \end{cases}$$

Let $H_{\tilde{F}}$ be the distribution function of $\tilde{F}$. Note δ_γ exists and is finite since $\gamma(0) > 0$ implies $\mathrm{Supp}(F)$ is nonempty, and $\mathrm{Supp}(F) \subset [-1, 1]$. Also we note $\tilde{F}((-\infty, \infty)) = \gamma(0)^{-1} F((-\infty, \infty)) = 1$ since $\gamma(0) = \int \alpha^0 F(d\alpha) = F((-\infty, \infty))$.

We first show that $1 - \delta_\gamma$ is the smallest value of b such that $[-b, b]$ has full F-measure, i.e., $1 - \delta_\gamma = \inf\{t; F([-t, t]) \geq \gamma(0)\} = \inf\{t; H_{\tilde{F}}(t) \geq 1\}$, and for $a > \delta_\gamma$, $F([-(1-a), (1-a)]) < \gamma(0)$.

In the case $\delta_\gamma = 1$, then $\mathrm{Supp}(F) = \{0\}$, and for $a > \delta_\gamma$, $[-(1-a), (1-a)] = \emptyset$, which is not full measure since $\gamma(0) > 0$. We next consider the case $\delta_\gamma < 1$. From the definition of δ_γ, $\mathrm{Supp}(F) \subset [-(1-\delta_\gamma), 1-\delta_\gamma]$. Now, consider a such that $\delta_\gamma < a \leq 1$. We show $[-(1-a), 1-a]$ is not full F measure. Since $\mathrm{Supp}(F)$ is closed, we have $\{-(1-\delta_\gamma), 1-\delta_\gamma\} \cap \mathrm{Supp}(F) \neq \emptyset$. Let $N_\theta(x) = \{y : |y - x| < \theta\}$ denote the open θ-neighborhood of x. Define $\theta_0 = (a - \delta_\gamma)/2$. Then $\{-(1-\delta_\gamma), 1-\delta_\gamma\} \cap \mathrm{Supp}(F) \neq \emptyset$ implies the open set $A = N_{\theta_0}(1-\delta_\gamma) \cup N_{\theta_0}(-(1-\delta_\gamma))$ has $F(A) > 0$, but $A \cap [-(1-a), 1-a] = \emptyset$, and so $F([-(1-a), 1-a]) < F((-\infty, \infty)) = \gamma(0)$. Thus

$$1 - \delta_\gamma = \inf\{t; F([-t, t]) \geq \gamma(0)\} = \inf\{t; H_{\tilde{F}}(t) \geq 1\} \tag{S-27}$$

From the definition of $\hat{m}$, we have

$$\hat{\rho}_M(\hat{m}) \geq c_M \sqrt{\frac{\log M}{M}} \qquad \text{and} \qquad \hat{\rho}_M(\hat{m}+2) \leq c_M \sqrt{\frac{\log M}{M}}. \tag{S-28}$$

First, we consider the case $\delta_\gamma < 1$.

Let Δ_M be $\Delta_M = \sup_{k=0,\dots,M-1} |\rho(k) - \hat{\rho}_M(k)|$. Since $\Delta_M = O_{P_x}(\sqrt{\log M/M})$, we have $C_\beta > 0$ and a finite M_1 such that $\Delta_M \le C_\beta \sqrt{\log M/M}$ with probability at least $1-\beta$ for all $M \ge M_1$. Let $\mathcal{E}_M$ be the event such that this inequality holds.

On the event $\mathcal{E}_M$, the second condition in (S-28) implies

$$\begin{aligned}
&\hat{\rho}_M(\hat{m}+2) \le c_M \sqrt{\frac{\log M}{M}} \\
&\Rightarrow \rho(\hat{m}+2) - |\hat{\rho}_M(\hat{m}+2) - \rho(\hat{m}+2)| \le c_M \sqrt{\frac{\log M}{M}} \\
&\Rightarrow \rho(\hat{m}+2) \le (C_\beta + c_M) \sqrt{\frac{\log M}{M}}.
\end{aligned}$$

Note by definition of ρ,

$$\rho(\hat{m}+2) = \gamma(0)^{-1} \int \alpha^{\hat{m}+2} F(d\alpha) = \gamma(0)^{-1} \int |\alpha|^{\hat{m}+2} F(d\alpha) = \int \alpha^{\hat{m}+2} \tilde{F}(d\alpha).$$

We will lower-bound $\int \alpha^{\hat{m}+2} \tilde{F}(d\alpha)$. First, define $\{a_k\}_{k=1}^\infty$ such that

$$a_k = \sup\{t \ge 0; H_{\tilde{F}}(t) < 1 - 1/\sqrt{\log(k)}\}, \tag{S-29}$$

where we take the convention of $\sup\{\emptyset\} = -\infty$. By definition of $H_{\tilde{F}}$, we have $H_{\tilde{F}}(1-\delta_\gamma) = 1$ and

$$H_{\tilde{F}}(1-\delta_\gamma) - H_{\tilde{F}}(a_k) \ge 1 - (1 - 1/\sqrt{\log(k)}) = 1/\sqrt{\log(k)} \tag{S-30}$$

Also, $a_k \ge a_{k+1}$ since $H_{\tilde{F}}$ is an increasing function. Now, we show $\lim_{k\to\infty} a_k = 1-\delta_\gamma$: first, we have $a_k \le 1-\delta_\gamma$ for all k. Therefore the limit of a_k exists. Now suppose to the contrary that $\lim_k a_k = c_a < 1-\delta_\gamma$. Then $H_{\tilde{F}}(c_a) < 1$ by (S-27). Choose $\{\epsilon_k\}$ such that $\epsilon_k \ge 0$, $a_k + \epsilon_k \le 1-\delta_\gamma$ and $\epsilon_k \to 0$. For each k, we have $H_{\tilde{F}}(a_k+\epsilon_k) \ge 1 - 1/\sqrt{\log(k)}$ by the definition of a_k. Then taking k limit to both sides, we have $\lim_k H_{\tilde{F}}(a_k+\epsilon_k) = H_{\tilde{F}}(c_a) \ge 1$ since $H_{\tilde{F}}$ is a right continuous function, and we have a contradiction. Therefore we have $\lim_k a_k = 1-\delta_\gamma$.

We have,

$$\int \alpha^{\hat{m}+2} \tilde{F}(d\alpha) \ge \int_{(a_M, 1-\delta_\gamma]} \alpha^{\hat{m}+2} \tilde{F}(d\alpha) \ge a_M^{\hat{m}+2} \int_{(a_M, 1-\delta_\gamma]} \tilde{F}(d\alpha) = a_M^{\hat{m}+2} (H_{\tilde{F}}(1-\delta_\gamma) - H_{\tilde{F}}(a_M)).$$

Since

$$a_M^{\hat{m}+2}(H_{\tilde{F}}(1-\delta_\gamma) - H_{\tilde{F}}(a_M)) \geq a_M^{\hat{m}+2}/\sqrt{\log M},$$

we have,

$$\begin{aligned}
&\int \alpha^{\hat{m}+2}\tilde{F}(d\alpha) \leq (c_M + C_\beta)\sqrt{\frac{\log M}{M}} \\
&\Rightarrow a_M^{\hat{m}+2} \leq \sqrt{\log M}(c_M + C_\beta)\sqrt{\frac{\log M}{M}} \\
&\Rightarrow (\hat{m}+2)\log a_M \leq \log(\sqrt{\log M}) + \log(c_M + C_\beta) + \frac{1}{2}\{\log(\log M) - \log M\} \\
&\Rightarrow \frac{\hat{m}}{\log M} \geq \frac{-2\log a_M + \log(\sqrt{\log M}) + \log(c_M + C_\beta) + \frac{1}{2}\{\log(\log M) - \log M\}}{\log a_M \log M} \\
&\Rightarrow 1 - \exp(-\frac{\log M}{2\hat{m}}) \leq \\
&\qquad 1 - \exp\left\{-\frac{1}{2}\left[\frac{\log a_M \log M}{-2\log a_M + \log(\sqrt{\log M}) + \log(c_M + C_\beta) + \frac{1}{2}\{\log(\log M) - \log M\}}\right]\right\}
\end{aligned}$$

Since $\log a_M \to \log(1-\delta_\gamma)$ as $M \to \infty$, the RHS converges to δ_γ. In other words, there exists a finite M_2 such that the RHS is $\delta_\gamma + \epsilon_0$. Therefore,

$$\hat{\delta}_M \leq \delta_\gamma + \epsilon_0 \tag{S-31}$$

on $\mathcal{E}_M$ for $M \geq M_2$. Therefore, for $M \geq \max\{M_1, M_2\}$, we have $P_x(\hat{\delta}_M \leq \delta_\gamma + \epsilon_0) \geq P(\mathcal{E}_M) \geq 1 - \beta$ for arbitrarily chosen β and ϵ_0, i.e., $\hat{\delta}_M$ is asymptotically not greater than δ_γ.

Now under the condition that $c_M \to \infty$ and $c_M = O(\log M)$, we show that $\hat{\delta}_M$ is asymptotically not smaller than δ_γ as well. The first condition in (S-28) implies on the event $\mathcal{E}_M$,

$$\begin{aligned}
&\hat{\rho}_M(\hat{m}) \geq c_M\sqrt{\frac{\log M}{M}} \\
&\Rightarrow \rho(\hat{m}) + |\rho(\hat{m}) - \hat{\rho}_M(\hat{m})| \geq c_M\sqrt{\frac{\log M}{M}} \\
&\Rightarrow \rho(\hat{m}) + C_\beta\sqrt{\frac{\log M}{M}} \geq c_M\sqrt{\frac{\log M}{M}} \\
&\Rightarrow \gamma(0)^{-1}\int \alpha^{\hat{m}}F(d\alpha) \geq (c_M - C_\beta)\sqrt{\frac{\log M}{M}}
\end{aligned}$$

Note that if we only require $c_M \geq 0$, the RHS can be negative depending on c_M and C_β. However, with the choice of $c_M \to \infty$, there exists a finite M_3 such that $c_M - C_\beta > 0$ for $M \geq M_3$.

We continue to upper bound the LHS. Since $\hat{m}$ is even, $\gamma(0)^{-1}\int \alpha^{\hat{m}} F(d\alpha) = \gamma(0)^{-1}\int |\alpha|^{\hat{m}} F(d\alpha) = \int \alpha^{\hat{m}} \tilde{F}(d\alpha)$. In particular, $\mathrm{Supp}(\tilde{F}) \subseteq [0, 1-\delta_\gamma]$, since $\tilde{F}([0,1-\delta_\gamma]) = \tilde{F}((-\infty, 1-\delta_\gamma]) = \gamma(0)^{-1}F([-1+\delta_\gamma, 1-\delta_\gamma]) = 1$ by the definition of δ_γ. Therefore,

$$
\begin{aligned}
&\int \alpha^{\hat{m}} \tilde{F}(d\alpha) \geq (c_M - C_\beta)\sqrt{\frac{\log M}{M}} \\
&\Rightarrow (1-\delta_\gamma)^{\hat{m}} \geq (c_M - C_\beta)\sqrt{\frac{\log M}{M}} \\
&\Rightarrow \hat{m}\log(1-\delta_\gamma) \geq \log(c_M - C_\beta) + \frac{1}{2}\{\log(\log M) - \log M\} \\
&\Rightarrow \hat{m} \leq \frac{\log(c_M - C_\beta)}{\log(1-\delta_\gamma)} + \frac{1}{2\log(1-\delta_\gamma)}\{\log(\log M) - \log M\}
\end{aligned}
$$

since $\log(1-\delta_\gamma) < 0$. By dividing both sides by $\log(M)/2$,

$$
\begin{aligned}
&\frac{2\hat{m}}{\log M} \leq \frac{2\log(c_M - C_\beta) + \log(\log M) - \log M}{\log(1-\delta_\gamma)\log(M)} \\
&\Rightarrow 1 - \exp(-\frac{2\hat{m}}{\log M}) \geq 1 - \exp\left(-\frac{\log(1-\delta_\gamma)\log(M)}{2\log(c_M - C_\beta) + \log(\log M) - \log M}\right)
\end{aligned}
$$

By the condition of $c_M = O(\log(M))$, $c_M/\log(M) \leq C$ for some constant C for a sufficiently large M.

$$
\frac{\log(c_M - C_\beta)}{\log(M)} \leq \frac{\log(c_M)}{\log(M)} \leq \frac{\log(C\log(M))}{\log(M)} = o(1).
$$

Therefore, the RHS converges to δ_γ, and we can find a finite M_4 such that

$$
\hat{\delta}_M \geq \delta_\gamma - \epsilon_0 \tag{S-32}
$$

on $\mathcal{E}_M$, for $M \geq \max\{M_3, M_4\}$. Thus, under the additional condition that $c_M \to \infty$ and $c_M = O(\log M)$, combining (S-31) and (S-32) yields $P_x(\{|\hat{\delta}_M - \delta_\gamma| > \epsilon_0\}) \leq P(\mathcal{E}_M^c) \leq \beta$, for $M \geq \max_{i=1,\ldots,4} M_i$, i.e., $\hat{\delta}_M \to \delta_\gamma$ in probability since β and ϵ_0 were arbitrary. This shows the result in the case $\delta_\gamma < 1$.

Now we consider the case when $\delta_\gamma = 1$. In this case, the inequality $\hat{\delta}_M \leq \delta_\gamma = 1$ is trivially true with probability 1 from the definition of $\hat{m}$, and therefore, we have $P_x(\hat{\delta}_M \leq$

$\delta_\gamma + \epsilon_0) = 1$ for all M, for each $\epsilon_0 > 0$. Thus $\hat{\delta}_M$ is not asymptotically larger than δ_γ.

Now, under the additional assumption $c_M \to \infty$ and $c_M = O(\log(M))$, we show $\hat{\delta}_M \xrightarrow{p} \delta_\gamma = 1$. Let $\epsilon_0, \beta > 0$ given. As before, let Δ_M be $\Delta_M = \sup_{k=0,\ldots,M-1} |\rho(k) - \hat{\rho}_M(k)|$. Since $\Delta_M = O_{P_x}(\sqrt{\log M/M})$, we have $C_\beta > 0$ and a finite M_1 such that $\Delta_M \leq C_\beta \sqrt{\log M/M}$ with probability at least $1 - \beta$ for all $M \geq M_1$. Let $\mathcal{E}_M$ denote the event $\Delta_M \leq C_\beta \sqrt{\frac{\log M}{M}}$. We also have a finite M_2 such that $c_M \geq C_\beta$ for $M \geq M_2$. Therefore, on $\mathcal{E}_M$,

$$\hat{\rho}_M(2) \leq C_\beta \sqrt{\frac{\log M}{M}} \Rightarrow \hat{\rho}_M(2) \leq c_M \sqrt{\frac{\log M}{M}} \Rightarrow \hat{m} = 0,$$

holds for all $M \geq M_2$. Note $\hat{\delta}_M = 1$ whenever $\hat{m} = 0$. Thus for $M \geq \max\{M_1, M_2\}$,

$$P_x(|\hat{\delta}_M - \delta_\gamma| \geq \epsilon_0) \leq P_x(\hat{\delta}_M \neq \delta_\gamma) = P_x(\hat{m} \neq 0) \leq \beta.$$

Since β was arbitrary, this proves the result $\hat{\delta}_M \xrightarrow{p} \delta_\gamma = 1$. □

S7 Supplementary Tables for Section 5

Here we present some supplementary tables for Section 5.

Table S1: *Estimated average ℓ_2 error (s.e.) for the autocovariance sequence estimators and mean squared error (s.e.) for the asymptotic variance estimators for discrete state space Metropolis-Hastings example*

(a) ℓ_2 error

Estimator	4000	8000	16000	32000	64000	128000
Empirical	1.2732	1.2639	1.2616	1.2668	1.2683	1.2666
	(0.0074)	(0.0051)	(0.0035)	(0.0025)	(0.0018)	(0.0013)
Bartlett	0.0092	0.0056	0.0034	0.0021	0.0012	0.0007
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
MomentLS(Orcl,Brtl)	**0.0054**	**0.0029**	0.0016	0.0009	0.0005	0.0003
	(0.0003)	**(0.0001)**	(0.0001)	(0.0000)	(0.0000)	(0.0000)
MomentLS(Tune,Emp)	0.0059	0.0030	0.0016	**0.0008**	**0.0004**	**0.0002**
	(0.0003)	(0.0002)	(0.0001)	**(0.0000)**	**(0.0000)**	**(0.0000)**
MomentLS(Tune-Incr,Emp)	0.0059	0.0030	0.0016	**0.0008**	**0.0004**	**0.0002**
	(0.0003)	(0.0002)	(0.0001)	**(0.0000)**	**(0.0000)**	**(0.0000)**
MomentLS(Orcl,Emp)	0.0056	**0.0029**	**0.0015**	**0.0008**	**0.0004**	**0.0002**
	(0.0003)	**(0.0001)**	**(0.0001)**	**(0.0000)**	**(0.0000)**	**(0.0000)**

(b) Asymptotic variance mean squared error

Estimator	4000	8000	16000	32000	64000	128000
BM	0.0917	0.0548	0.0377	0.0234	0.0134	0.0086
	(0.0056)	(0.0035)	(0.0024)	(0.0017)	(0.0009)	(0.0006)
OLBM	0.0940	0.0559	0.0332	0.0204	0.0118	0.0070
	(0.0058)	(0.0036)	(0.0022)	(0.0014)	(0.0008)	(0.0005)
Empirical	6.4180	6.4180	6.4180	6.4180	6.4180	6.4180
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Bartlett	0.0921	0.0541	0.0325	0.0201	0.0115	0.0069
	(0.0058)	(0.0034)	(0.0022)	(0.0014)	(0.0008)	(0.0005)
Init-Positive	0.1236	0.0571	0.0332	0.0154	0.0084	0.0038
	(0.0146)	(0.0051)	(0.0041)	(0.0015)	(0.0011)	(0.0003)
Init-Decr	0.0878	0.0400	0.0230	0.0114	0.0058	0.0030
	(0.0079)	(0.0030)	(0.0022)	(0.0009)	(0.0005)	(0.0002)
Init-Convex	0.0741	0.0348	0.0195	0.0101	0.0051	0.0026
	(0.0063)	(0.0026)	(0.0019)	(0.0008)	(0.0004)	(0.0002)
MomentLS(Orcl,Brtl)	**0.0521**	**0.0274**	**0.0148**	0.0085	0.0048	0.0030
	(0.0038)	**(0.0018)**	**(0.0010)**	(0.0005)	(0.0003)	(0.0002)
MomentLS(Tune,Emp)	0.0697	0.0342	0.0183	0.0092	0.0046	0.0023
	(0.0059)	(0.0025)	(0.0017)	(0.0006)	(0.0004)	(0.0002)
MomentLS(Tune-Incr,Emp)	0.0675	0.0336	0.0179	0.0090	0.0046	0.0023
	(0.0056)	(0.0025)	(0.0017)	(0.0006)	(0.0004)	(0.0002)
MomentLS(Orcl,Emp)	0.0558	0.0285	**0.0148**	**0.0079**	**0.0038**	**0.0020**
	(0.0043)	(0.0020)	**(0.0011)**	**(0.0005)**	**(0.0003)**	**(0.0002)**

Table S2: *Estimated average ℓ_2 error (s.e.) for the autocovariance sequence estimators and mean squared error (s.e.) for the asymptotic variance estimators for AR1 example with* $\rho = 0.9$

(a) ℓ_2 error

Estimator	4000	8000	16000	32000	64000	128000
Empirical	260.4808	262.8923	261.6753	261.7598	263.6259	263.6811
	(3.3948)	(2.6730)	(1.8479)	(1.3180)	(0.9379)	(0.6410)
Bartlett	7.1962	4.4781	2.7572	1.6033	0.9710	0.5964
	(0.2886)	(0.1685)	(0.0995)	(0.0540)	(0.0312)	(0.0175)
MomentLS(Orcl,Brtl)	3.9916	2.3139	1.3216	0.6979	0.4162	0.2675
	(0.2211)	(0.1283)	(0.0762)	(0.0396)	(0.0241)	(0.0140)
MomentLS(Tune,Emp)	4.8135	2.9854	1.5289	0.7512	0.4092	0.1970
	(0.2362)	(0.1719)	(0.0819)	(0.0416)	(0.0276)	(0.0117)
MomentLS(Tune-Incr,Emp)	4.7958	2.9674	1.5241	0.7509	0.4082	0.1973
	(0.2356)	(0.1700)	(0.0816)	(0.0416)	(0.0275)	(0.0117)
MomentLS(Orcl,Emp)	**3.7863**	**2.1209**	**1.1063**	**0.5141**	**0.2680**	**0.1247**
	(0.2114)	**(0.1209)**	**(0.0675)**	**(0.0322)**	**(0.0189)**	**(0.0078)**

(b) Asymptotic variance mean squared error

Estimator	4000	8000	16000	32000	64000	128000
BM	441.3225	310.6322	207.5458	120.8707	75.7108	50.9232
	(26.3386)	(18.7844)	(12.0461)	(7.1677)	(4.5795)	(3.1021)
OLBM	530.6842	329.0258	200.7452	113.6880	70.3730	43.9050
	(30.8907)	(18.6833)	(11.5641)	(6.9584)	(4.6056)	(2.8280)
Empirical	10,000.0000	10,000.0000	10,000.0000	10,000.0000	10,000.0000	10,000.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Bartlett	480.1497	304.9819	188.2221	108.3490	67.6490	42.5953
	(28.6448)	(18.0249)	(11.0456)	(6.7378)	(4.5006)	(2.7620)
Init-Positive	727.0617	384.9189	200.7905	94.0732	53.6517	25.4210
	(102.9961)	(52.4958)	(24.7859)	(9.7885)	(5.0255)	(2.3289)
Init-Decr	442.0562	289.4032	141.7235	70.1543	39.3842	19.6494
	(37.6246)	(30.6812)	(12.9728)	(5.7635)	(3.5904)	(1.7146)
Init-Convex	349.7933	240.7814	120.4678	59.7319	34.3007	16.8201
	(23.8527)	(22.5304)	(9.4864)	(4.4887)	(3.1422)	(1.4084)
MomentLS(Orcl,Brtl)	206.4103	121.6584	73.4146	40.7843	24.8478	16.6819
	(13.2641)	(7.7392)	(4.9726)	(2.6673)	(1.5947)	(0.9671)
MomentLS(Tune,Emp)	317.3013	217.8352	103.8053	52.8779	30.1932	14.2880
	(21.2726)	(18.9499)	(7.7694)	(4.0613)	(3.0648)	(1.3012)
MomentLS(Tune-Incr,Emp)	313.1335	214.4295	103.0196	52.5359	30.0120	14.2162
	(20.8190)	(18.4725)	(7.7013)	(4.0221)	(3.0511)	(1.2912)
MomentLS(Orcl,Emp)	**187.2898**	**104.0514**	**56.0977**	**26.4687**	**13.3904**	**6.2344**
	(12.1154)	**(6.7444)**	**(4.0863)**	**(1.9108)**	**(1.0194)**	**(0.4571)**

Table S3: *Estimated average ℓ_2 error (s.e.) for the autocovariance sequence estimators and mean squared error (s.e.) for the asymptotic variance estimators for AR1 example with $\rho = -0.9$*

(a) ℓ_2 error

Estimator	4000	8000	16000	32000	64000	128000
Empirical	260.2496	260.7740	260.8336	264.7212	263.1831	263.6014
	(3.5967)	(2.6222)	(1.8960)	(1.2805)	(0.8522)	(0.6256)
Bartlett	6.6264	4.1319	2.6333	1.5941	0.9635	0.5946
	(0.2658)	(0.1757)	(0.0993)	(0.0591)	(0.0292)	(0.0165)
MomentLS(Orcl,Brtl)	3.7670	2.0407	1.2248	0.6844	0.3941	0.2585
	(0.2187)	(0.1308)	(0.0719)	(0.0449)	(0.0237)	(0.0146)
MomentLS(Tune,Emp)	4.6199	2.5181	1.3821	0.7286	0.3639	0.1968
	(0.2543)	(0.1457)	(0.0754)	(0.0432)	(0.0217)	(0.0110)
MomentLS(Tune-Incr,Emp)	4.6113	2.5137	1.3793	0.7305	0.3638	0.1974
	(0.2541)	(0.1459)	(0.0753)	(0.0436)	(0.0219)	(0.0111)
MomentLS(Orcl,Emp)	**3.5910**	**1.8474**	**1.0294**	**0.5192**	**0.2388**	**0.1236**
	(0.2156)	**(0.1215)**	**(0.0627)**	**(0.0368)**	**(0.0169)**	**(0.0079)**

(b) Asymptotic variance mean squared error

Estimator	4000	8000	16000	32000	64000	128000
BM	0.0048	0.0030	0.0017	0.0010	0.0006	0.0005
	(0.0003)	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
OLBM	0.0025	0.0018	0.0011	0.0007	0.0005	0.0003
	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
Empirical	0.0767	0.0767	0.0767	0.0767	0.0767	0.0767
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)
Bartlett	0.0027	0.0019	0.0011	0.0007	0.0005	0.0003
	(0.0002)	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)
Init-Positive	0.0973	0.0526	0.0232	0.0114	0.0059	0.0032
	(0.0076)	(0.0036)	(0.0016)	(0.0008)	(0.0004)	(0.0002)
Init-Decr	0.1186	0.0663	0.0309	0.0149	0.0078	0.0044
	(0.0079)	(0.0041)	(0.0020)	(0.0010)	(0.0005)	(0.0003)
Init-Convex	0.3009	0.1664	0.0835	0.0417	0.0209	0.0116
	(0.0134)	(0.0076)	(0.0037)	(0.0019)	(0.0009)	(0.0005)
MomentLS(Orcl,Brtl)	0.0028	0.0017	0.0010	0.0007	0.0004	0.0003
	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	(0.0000)
MomentLS(Tune,Emp)	0.0034	0.0019	0.0010	0.0006	0.0003	**0.0001**
	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	**(0.0000)**
MomentLS(Tune-Incr,Emp)	0.0034	0.0019	0.0010	0.0006	0.0003	**0.0001**
	(0.0002)	(0.0001)	(0.0001)	(0.0000)	(0.0000)	**(0.0000)**
MomentLS(Orcl,Emp)	**0.0023**	**0.0012**	**0.0006**	**0.0003**	**0.0001**	**0.0001**
	(0.0001)	**(0.0001)**	**(0.0000)**	**(0.0000)**	**(0.0000)**	**(0.0000)**

References

James H. Albert and Siddhartha Chib. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679, 1993.

Hong-Zhi An, Zhao-Guo Chen, and Edward J Hannan. Autocorrelation, autoregression and autoregressive approximation. *The Annals of Statistics*, pages 926–936, 1982.

T. W. Anderson. *The statistical analysis of time series.* John Wiley & Sons, Inc., New York-London-Sydney, 1971.

Fadoua Balabdaoui and Gabriella de Fournas-Labrosse. Least squares estimation of a completely monotone pmf: From Analysis to Statistics. *Journal of Statistical Planning and Inference*, 204:55–71, 2020.

Fadoua Balabdaoui and Cécile Durot. Marshall lemma in discrete convex estimation. *Statistics & Probability Letters*, 99:143–148, 2015.

Fadoua Balabdaoui and Jon A Wellner. Estimation of a k-monotone density: limit distribution theory and the spline connection. *The Annals of Statistics*, 35(6):2536–2564, 2007.

Richard E Barlow, D J Bartholomew, J M Bremner, and H D Brunk. *Statistical inference under order restrictions: The theory and application of isotonic regression (Wiley series in probability and mathematical statistics, no. 8).* Wiley, January 1972.

Witold Bednorz and Krzysztof Latuszyński. A few remarks on "Fixed-width output analysis for Markov chain Monte Carlo" by Jones et al. *Journal of the American Statistical Association*, 102(480):1485–1486, 2007.

Stephen Berg and Hyebin Song. Supplement to "efficient shape-constrained inference for the autocovariance sequence from a reversible markov chain". 2023.

Peter J Brockwell and Richard A Davis. *Time series: theory and methods.* Springer Science & Business Media, 2009.

Steve Brooks, Andrew Gelman, Galin Jones, and Xiao-Li Meng. *Handbook of Markov Chain Monte Carlo.* CRC press, 2011.

Saptarshi Chakraborty and Kshitij Khare. Convergence properties of Gibbs samplers for Bayesian probit regression with proper priors. *Electronic Journal of Statistics*, 11(1):177 – 210, 2017.

Saptarshi Chakraborty, Suman K Bhattacharya, and Kshitij Khare. Estimating accuracy of the MCMC variance estimator: Asymptotic normality for batch means estimators. *Statistics & Probability Letters*, 183:109337, 2022.

JD Chandler. Moment problems for compact sets. *Proceedings of the American Mathematical Society*, 104(4):1134–1140, 1988.

Chew-Seng Chee and Yong Wang. Nonparametric estimation of species richness using discrete k-monotone distributions. *Computational Statistics & Data Analysis*, 93:107–118, 2016.

Ning Dai and Galin L. Jones. Multivariate initial sequence estimators in Markov chain Monte Carlo. *Journal of Multivariate Analysis*, 159:184–199, 2017.

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

Lutz Dümbgen and Kaspar Rufibach. logcondens: Computations related to univariate log-concave density estimation. *Journal of Statistical Software*, 39:1–28, 2011.

Alain Durmus, Samuel Gruffaz, Miika Kailas, Eero Saksman, and Matti Vihola. On the convergence of dynamic implementations of hamiltonian monte carlo and no u-turn samplers. *arXiv preprint arXiv:2307.03460*, 2023.

Cécile Durot, Sylvie Huet, François Koladjo, and Stéphane Robin. Nonparametric species richness estimation under convexity constraint. *Environmetrics*, 26(7):502–513, November 2015.

Willy Feller. Completely monotone functions and sequences. *Duke Mathematical Journal*, 5(3):661 – 674, 1939.

James M Flegal and Lei Gong. Relative fixed-width stopping rules for Markov chain Monte Carlo simulations. *Statistica Sinica*, pages 655–675, 2015.

James M. Flegal and Galin L. Jones. Batch means and spectral variance estimators in Markov chain Monte Carlo. *The Annals of Statistics*, 38(2):1034 – 1070, 2010.

James M Flegal, Murali Haran, and Galin L Jones. Markov chain Monte Carlo: Can we trust the third significant figure? *Statistical Science*, pages 250–260, 2008.

James M. Flegal, John Hughes, Dootika Vats, Ning Dai, Kushagra Gupta, and Uttiya Maji. *mcmcse: Monte Carlo Standard Errors for MCMC*. Riverside, CA, and Kanpur, India, 2021. R package version 1.5-0.

Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

Charles J. Geyer. Practical Markov chain Monte Carlo. *Statistical Science*, 7(4):473–483, 1992.

Charles J. Geyer and Leif T. Johnson. *mcmc: Markov Chain Monte Carlo*, 2020. R package version 0.9-7.

Jade Giguelay. Estimation of a discrete probability under constraint of k-monotonicity. *Electronic journal of statistics*, 11(1):1–49, 2017.

Peter W Glynn and Ward Whitt. The asymptotic validity of sequential stopping rules for stochastic simulations. *The Annals of Applied Probability*, 2(1):180–198, 1992.

Ulf Grenander. On the theory of mortality measurement. *Scandinavian Actuarial Journal*, 1956(1):70–96, 1956.

Piet Groeneboom, Geurt Jongbloed, and Jon A Wellner. The support reduction algorithm for computing non-parametric function estimates in mixture models. *Scandinavian Journal of Statistics*, 35(3):385–399, 2008.

Olle Haggstrom and Jeffrey Rosenthal. On variance conditions for Markov chain CLTs. *Electronic Communications in Probability*, 12:454–464, 2007.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970.

Felix Hausdorff. Summationsmethoden und momentfolgen. I. *Mathematische Zeitschrift*, 9 (1):74–109, 1921.

Hanna K Jankowski and Jon A Wellner. Estimation of a discrete monotone distribution. *Electronic journal of statistics*, 3:1567, 2009.

Søren F Jarner and Richard L Tweedie. Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains. *Bernoulli*, 9(4):559–578, 2003.

Søren Fiig Jarner and Ernst Hansen. Geometric ergodicity of metropolis algorithms. *Stochastic Processes and their Applications*, 85(2):341–361, 2000.

Nicholas P Jewell. Mixtures of exponential distributions. *The Annals of Statistics*, pages 479–484, 1982.

Leif T Johnson and Charles J Geyer. Variable transformation to obtain geometric ergodicity in the random-walk metropolis algorithm. *The Annals of Statistics*, pages 3050–3076, 2012.

Galin L. Jones. On the Markov chain central limit theorem. *Probability Surveys*, 1:299 – 320, 2004.

Galin L Jones and Qian Qin. Markov chain monte carlo in practice. *Annual Review of Statistics and Its Application*, 9:557–578, 2022.

Galin L Jones, Murali Haran, Brian S Caffo, and Ronald Neath. Fixed-width output analysis for Markov chain Monte Carlo. *Journal of the American Statistical Association*, 101(476):1537–1547, 2006.

Laimonis Kavalieris. Uniform convergence of autocovariances. *Statistics & probability letters*, 78(6):830–838, 2008.

Ioannis Kontoyiannis and Sean P Meyn. Geometric ergodicity and the spectral gap of non-reversible Markov chains. *Probability Theory and Related Fields*, 154(1):327–339, 2012.

Michael R. Kosorok. Monte Carlo error estimation for multivariate Markov chains. *Statistics & Probability Letters*, 46(1):85–93, 2000.

M. G. Krein and Adolf Abramovich Nudelman. *The Markov moment problem and extremal problems : ideas and problems of P. L. Cebysev and A. A. Markov and their further development*. American Mathematical Society Providence, R.I, 1977.

Arun K. Kuchibhotla, Rohit K. Patra, and Bodhisattva Sen. Semiparametric efficiency in convexity constrained single-index model. *Journal of the American Statistical Association*, 0:1–15, 2021.

Krzysztof Latuszynski. Regeneration and fixed-width analysis of Markov chain Monte Carlo algorithms. 2009.

Claude Lefèvre and Stéphane Loisel. On multiply monotone distributions, continuous or discrete, with applications. *Journal of Applied Probability*, 50(3):827–847, 2013.

Bruce G Lindsay. The geometry of mixture likelihoods: a general theory. *The Annals of Statistics*, pages 86–94, 1983.

Jun S. Liu, Wing Hung Wong, and Augustine Kong. Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81:27–40, 1994.

Ying Liu, Dootika Vats, and James M Flegal. Batch size selection for variance estimators in mcmc. *Methodology and Computing in Applied Probability*, pages 1–29, 2021.

Samuel Livingstone, Michael Betancourt, Simon Byrne, and Mark Girolami. On the geometric ergodicity of Hamiltonian Monte Carlo. *Bernoulli*, 25(4A):3109 – 3138, 2019.

Kerrie L Mengersen and Richard L Tweedie. Rates of convergence of the hastings and metropolis algorithms. *The Annals of Statistics*, 24(1):101–121, 1996.

Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge University Press, Cambridge, second edition, 2009.

Dimitris N Politis. Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 15 (4-5):517–533, 2003.

M.B. Priestley. *Spectral Analysis and Time Series*. Number Bd. 1-2 in Probability and mathematical statistics : a series of monographs and textbooks. Academic Press, 1981.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2020.

Frigyes Riesz and Béla Sz.-Nagy. *Functional Analysis*. Courier Corporation, December 2012.

Christian P. Robert and George Casella. *Monte Carlo Statistical Methods*. Springer-Verlag, New York, second edition, 2004.

Gareth Roberts and Jeffrey Rosenthal. Geometric ergodicity and hybrid Markov chains. *Electronic Communications in Probability*, 2:13–25, 1997.

Gareth O Roberts and Jeffrey S Rosenthal. Markov-chain monte carlo: some practical implications of theoretical results. *The Canadian Journal of Statistics/La Revue Canadienne de Statistique*, pages 5–20, 1998.

Gareth O Roberts and Richard L Tweedie. Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. *Biometrika*, 83(1):95–110, 1996.

Tim Robertson. Order restricted statistical inference. Technical report, 1988.

W. Rudin. *Functional Analysis*. International series in pure and applied mathematics. McGraw-Hill, 1991.

Konrad Schmüdgen. *The moment problem*, volume 9. Springer, 2017.

Stan Development Team. Stan reference manual, version 2.28, 2019.

F. W. Steutel. Note on completely monotone densities. *The Annals of Mathematical Statistics*, 40:1130 – 1131, 1969.

Dootika Vats, James M. Flegal, and Galin L. Jones. Strong consistency of multivariate spectral variance estimators in Markov chain Monte Carlo. *Bernoulli*, 24(3):1860 – 1909, 2018.

Dootika Vats, James M Flegal, and Galin L Jones. Multivariate output analysis for Markov chain Monte Carlo. *Biometrika*, 106(2):321–337, 2019.