Statistical and computation aspects of shape-constrained inference for autocovariance sequence estimation[1]

Stephen Berg

Department of Statistics, Penn State University

Joint work with Hyebin Song (Stat, Penn State)

JSM 2024

[1] This material is based upon work supported by the National Science Foundation under Grant DMS-2311141.

Introduction: MCMC for Bayesian statistics

- probability measure π on $(\mathsf{X}, \mathscr{X})$, e.g., Bayesian posterior distribution
- want $\mu = \int g(x)\,\pi(dx) = E_\pi[g]$
- construct Markov chain $X_0, X_1, \ldots$ with stationary distribution π
- estimate μ by

$$\hat{\mu}_M = M^{-1} \sum_{t=0}^{M-1} g(X_t)$$

- quantify uncertainty in $\hat{\mu}_M$
- turns out, this problem is closely related to estimating the autocovariance function $\gamma : \mathbb{Z} \to \mathbb{R}$, defined as

$$\gamma(k) = \mathrm{Cov}_\pi(g(X_0), g(X_k)),\ k \in \mathbb{Z},$$

Asymptotic variance

Suppose $X_0, X_1, \ldots$ are a Markov chain sequence with a stationary probability measure π and transition kernel Q

- The asymptotic variance $\sigma^2 = \sum_{k=-\infty}^{\infty} \gamma(k)$ aggregates the covariance from all lags. Given an estimate $\hat{\gamma}$ of γ, we may consider $\hat{\sigma}^2 = \sum_{k=-\infty}^{\infty} \hat{\gamma}(k)$.
- $\text{Var}(\hat{\mu}_M) \approx \sigma^2/M$ for large M.
  - σ^2 is the CLT variance, or the limit of $M\text{Var}(\hat{\mu}_M)$ as $M \to \infty$ [e.g. Haggstrom and Rosenthal, 2007]

Sample autocovariances

- For a given sample $X_0, X_1, \dots, X_{M-1}$ of size M, the empirical autocovariance sequence $r_M : \mathbb{Z} \to \mathbb{R}$ defined as

$$r_M(k) = \begin{cases} \frac{1}{M} \sum_{t=0}^{M-k-1} \tilde{g}(X_t)\tilde{g}(X_{t+k}) & , |k| \leq M-1 \\ 0 & , |k| \geq M \end{cases}$$

  is a natural estimator for γ, where $\tilde{g}(X_t) = g(X_t) - \frac{1}{M} \sum_{t=0}^{M-1} g(X_t)$.

- while $r_M(k) \to \gamma(k)$ for each k, r_M is rather terrible as an estimator for γ.
  - For example, summing the empirical autocovariances $\hat{\sigma}^2_{Emp} = \sum_{k=-\infty}^{\infty} r_M(k)$ leads to an inconsistent estimator of σ^2.
  - variance of $r_M(k)$ is "too large" compared to the signal $\gamma(k)$ for large k.

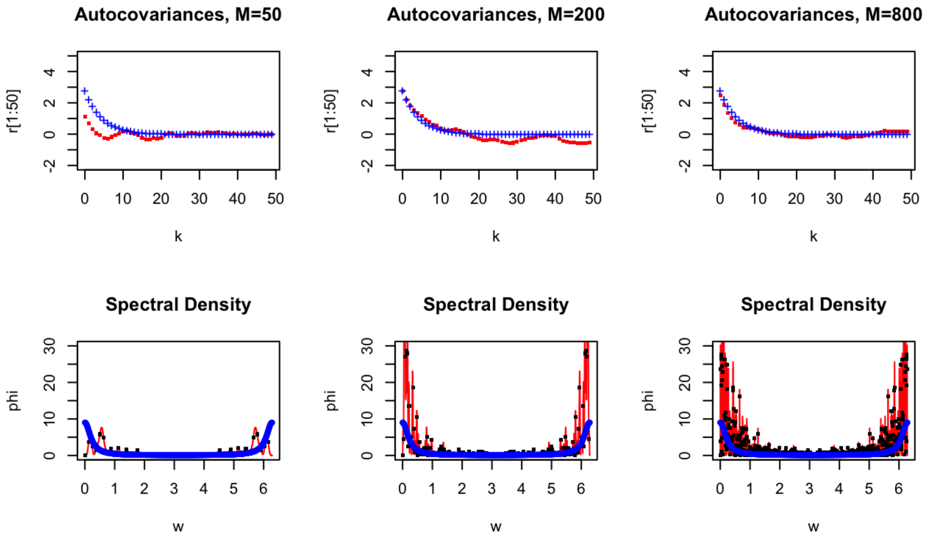
Spectral density

For a sequence $f : \mathbb{Z} \to \mathbb{R}$, define the discrete-time Fourier transform of f: $\hat{f} : [-\pi, \pi] \to \mathbb{R}$ such that

$$\hat{f}(\omega) = \sum_{k \in \mathbb{Z}} f(k) e^{-i\omega k}$$

- (spectral density) $\phi_\gamma(\omega) = \sum_{k \in \mathbb{Z}} \gamma(k) e^{-i\omega k}$
- (periodogram) $\hat{r}_M(\omega) = \sum_{k \in \mathbb{Z}} r_M(k) e^{-i\omega k}$, which is an estimate of ϕ_γ
- By Parseval's identity, $\sum_{k \in \mathbb{Z}} (\gamma(k) - r_M(k))^2 = \frac{1}{2\pi} \int_{[-\pi,\pi]} (\phi_\gamma(\omega) - \hat{r}_M(\omega))^2 d\omega$
  - ℓ_2 square distance between r_M and γ = integrated square error between $\tilde{\phi}_M$ and ϕ_γ
- $\sigma^2 = \phi_\gamma(0)$

Red = Empirical autocovariance (periodogram)
Blue = True autocovariance (spectral density)

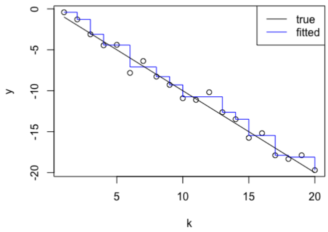


Estimation with shape constraints

Estimation with various shape constraints can be of interest:

- Monotonicity
  - Isotonic regression [e.g.,Barlow et al. [1972]]: for finite $y \in \mathbb{R}^n$, $y_k = f_k + \epsilon_k$, $f_k \geq f_{k+1}$ for $k = 1, \ldots, n$.

$$\hat{f}_{iso} = \underset{f; f_k \geq f_{k+1}, k=1,\ldots,d-1}{\arg\min} \|y - f\|^2$$



Estimation with shape constraints

- Monotonicity (cont'd)
  - Estimation of a discrete monotone pmf [Jankowski and Wellner, 2009]

$$\hat{p}_M(k+1) \geq \hat{p}_M(k) \geq 0, \text{ for } n, k \in \mathbb{N}$$

  - Estimation of a discrete completely monotone pmf [Balabdaoui and de Fournas-Labrosse, 2020]

$$(-1)^n \Delta^n \hat{p}(k) \geq 0, \text{ for } n \in \mathbb{N}$$

    where $\Delta^0 p(k) = p(k)$, $\Delta^n p(k) = \Delta^{n-1} p(k+1) - \Delta^{n-1} p(k)$, for $n = 1, 2, 3, \ldots,$ $k \in \mathbb{N}$
- Convexity, Log-concavity, etc. [e.g., Dümbgen and Rufibach [2011], Balabdaoui and Durot [2015], Kuchibhotla et al. [2017]]
- initial sequence estimators of Geyer [1992]

Mixture representations

Commonly, shape constraints on a function are related to a mixture representation for the function

- Complete monotonicity of sequence m (positive, decreasing, convex, ...):

  $(-1)^n \Delta^n m(k) \geq 0$, $\forall n, k \in \mathbb{N} \qquad \Longleftrightarrow$

  $\exists \mu$ s.t. $m(k) = \int_{[0,1]} x^k \, \mu(dx)$, $\forall k \in \mathbb{N}$
  [Hausdorff moment theorem Hausdorff, 1921]
- convex sequence: mixture of “trifunctions” [Durot et al., 2013]
- k-monotone: mixture of Beta(1,k) densities [Balabdaoui and Wellner, 2005]
- (Spatial statistics) Mátern covariance: mixture of Gaussians

Shape constraints in autocovariance function γ

- It is a well known result that the true autocovariance sequence γ for a reversible Markov chain admits the following representation [Rudin, 1973]:

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

  for a positive measure F supported on $[-1, 1]$

- Moreover, if a chain has a positive spectral gap, then F is supported on $[-1+\delta_\gamma, 1-\delta_\gamma]$ for some $\delta_\gamma > 0$ (true for e.g., an IID sample or a reversible chain with geometric ergodicity [Roberts and Rosenthal, 1997]).

- γ is a $[-1+\delta_\gamma, 1-\delta_\gamma]$ moment sequence for a reversible and geometrically ergodic chain.

Our approach 1: L2 projection of empirical autocovariace sequence

Let $\mathscr{M}_\infty(\delta)$ denote the set of $[-1+\delta, 1-\delta]$ moment sequences

(Un-weighted) Moment LSE [Berg and Song, 2023]:

$$\Pi(r_M;\delta) = \underset{m\in\mathscr{M}_\infty(\delta)\cap\ell_2(\mathbb{Z})}{\arg\min} \sum_{k\in\mathbb{Z}} \{r_M(k) - m(k)\}^2 \tag{2}$$

- projection onto the ℓ_2 moment sequence set
- here δ is a hyperparameter
- ignores unequal variances and non-zero covariances in $\{r_M(k)\}_{k\in\mathbb{Z}}$

Our approach 2: Weighted inner-product norm projection

- previously: covariance fitting objective $\|r_M - m\|_2^2 = \sum_{k=-\infty}^{\infty}(r_M(k) - m(k))^2$
- what about a weighted squared error loss function?
  - In regression, $Y = X\beta + \epsilon$, $\quad Var(Y) = \Sigma$
  - $\hat{\beta}_{OLS} = \text{argmin}_\beta \sum_{k=1}^{n}(Y_k - x_k^\top \beta)^2 = (Y - X\beta)^\top(Y - X\beta)$
    $\hat{\beta}_{GLS} = \text{argmin}_\beta (Y - X\beta)^\top \Sigma^{-1}(Y - X\beta)$

- covariances of r_M not easy to work with
  - $\text{Cov}(r_M(j), r_M(k)) \neq 0$ for all j, k such that $|j|, |k| < M \rightarrow$ covariance matrix not sparse

Our approach 2: Weighted inner-product norm projection

- recall the Parseval's identity: $\sum_{k\in\mathbb{Z}}(r_M(k) - m(k))^2 = \frac{1}{2\pi}\int_{[-\pi,\pi]}(\hat{r}_M(\omega) - \hat{m}(\omega))^2 d\omega$

- Periodogram at Fourier frequencies:

$$\hat{r}_M(\omega_k) \overset{ind}{\approx} \phi(\omega_k)\mathsf{Exp}(1) \qquad k = 0, ..., \lfloor M/2 \rfloor.$$

$\omega_k = 2\pi k/M$, $k = 0, ..., M-1$[e.g., Brockwell and Davis, 1991, Kokoszka and Mikosch, 2000]

Our approach 2: Weighted inner-product norm projection

Given a good estimate $\phi_M(\omega)$ of true spectral density, define
Weighted Moment LSE [Song and Berg, 2024+]

$$\Pi^{\phi_M}(r_M;\delta) = \underset{f \in \mathscr{M}_\infty(\delta)\cap \ell_2(\mathbb{Z},\mathbb{R})}{\arg\min} \|r_M - f\|^2_{\phi_M}$$

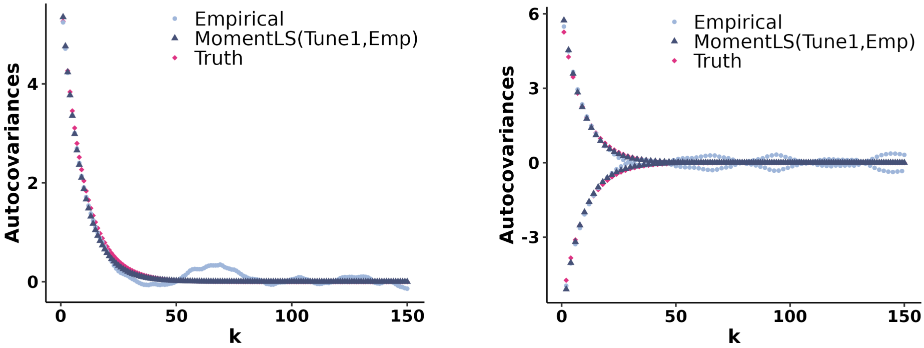
where the weighted inner-product norm $\|\cdot\|_{\phi_M}$ is defined as

$$\|r_M - f\|^2_{\phi_M} = (2\pi)^{-1} \int_{[-\pi,\pi]} \frac{\{\hat{r}_M(\omega) - \hat{f}(\omega)\}^2}{\phi_M(\omega)^2} d\omega$$

and $\hat{r}_M$ and $\hat{f}$ are Fourier transform of r_M and f.

- another interpretation of the objective: estimate spectral density using **shape-constraints on Fourier coefficients**

Moment LSE in practice



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

Figure: *For an AR(1) chain 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* $M = 8000$.

Estimator

$$\Pi^{\phi_M}(r_M;\delta) = \underset{f \in \mathscr{M}_\infty(\delta)\cap \ell_2(\mathbb{Z},\mathbb{R})}{\arg\min} \|r_M - f\|^2_{\phi_M}$$

Objective: minimize $\|r_M - f\|^2_{\phi_M}$ over valid autocovariances f

- ie sequences satisfying

$$f(k) = \int_{[-1+\delta,1-\delta]} \alpha^{|k|}\,\mu(d\alpha), \quad \forall k \in \mathbb{Z}$$

  for a positive measure μ with $\mathrm{Supp}(\mu) \subseteq [-1+\delta, 1-\delta]$.
- $0 < \delta \leq 1$ is a tuning parameter (estimated from data)

Finite, discrete support of estimated mixing measure

Proposition (ref Song and Berg [2024+])

- Let $C = [-1 + \delta, 1 - \delta]$
- Suppose $r(k) = r(-k)$, $\forall k$, and $r(k) = 0$ for $|k| > M_0 \geq 0$
- μ_C^ϕ denote the representing measure for $\Pi^\phi(r; C)$

Then $|\operatorname{Supp}(\mu_C^\phi)| \leq \frac{n}{2} + 1$, where n is the smallest even number s.t. $n > M_0$.

- The mixing measure μ_C^ϕ for $\Pi^\phi(r; C)$ is finite and discrete, with at most $\approx M/2$ support points, where M is the MCMC sample size
- proof uses techniques from total positivity [e.g. Karlin, 1968] to bound number of support points
- in practice much smaller than $M/2$

Computation

- The objective is:

$$\|r_M - f\|_{\phi_M}^2 = \text{const} + \frac{1}{2\pi}\left\{\int_{[-\pi,\pi]} -2\frac{\hat{r}_M(\omega)\hat{f}(\omega)}{\phi_M(\omega)^2}d\omega + \int_{[-\pi,\pi]} \frac{\hat{f}(\omega)^2}{\phi_M(\omega)^2}d\omega\right\}$$

- approximate $\Pi^{\phi_M}(r_M;\delta)$ with $\Pi^{\phi_M}(r_M;C)$ where $C = \{\beta_1, \ldots, \beta_s\} \subset [-1+\delta, 1-\delta]$
- optimize over f with representing measures μ of form

$$\mu = \sum_{i=1}^{s} c_i \delta_{\beta_i}$$

Computation

For f with representing measure $\mu = \sum_{i=1}^{s} c_i \delta_{\beta_i}$

$$\|r_m - f\|_{\phi_M}^2 = \text{const} - 2\sum_{i=1}^{s} c_i a_i + \sum_{i=1}^{s}\sum_{j=1}^{s} c_i c_j B_{ij}$$

where

$$B_{ij} = \langle x_{\beta_i}, x_{\beta_j} \rangle_{\phi_M} = \frac{1}{2\pi}\int_{[-\pi,\pi]} \frac{K(\beta_i,\omega)K(\beta_j,\omega)}{\phi_M(\omega)^2} d\omega$$

$$a_i = \langle x_{\beta_i}, r_M \rangle_{\phi_M} = \frac{1}{2\pi}\int_{[-\pi,\pi]} \frac{\hat{r}_M(\omega)K(\beta_i,\omega)}{\phi_M(\omega)^2} d\omega$$

$$K(\beta,\omega) = \frac{1-\beta^2}{1-2\beta\cos(\omega)+\beta^2} \quad \text{(Poisson kernel)}$$

- B_{ij}, a_i are computable from data. This is a quadratic programming problem with non-negativity constraints on c_i.

Plug-in spectral density and asymptotic variance estimators

For a weighted momentLS estimator $\Pi^{\phi_M}(r_M;\delta)$, we consider

$$\phi^W_{\delta M}(\omega) = \sum_{k\in\mathbb{Z}} \Pi^{\phi_M}(r_M;\delta)(k)e^{-i\omega k} \qquad \text{(spectral density)}$$

$$\sigma^2_{\delta M,W} = \sum_{k\in\mathbb{Z}} \Pi^{\phi_M}(r_M;\delta)(k) \qquad \text{(asymptotic variance)}$$

Statistical guarantee

Theorem ([Song and Berg, 2024+])

Consider a Harris ergodic, π-reversible, and geometrically ergodic Markov chain $X_0, X_1, \ldots$. Suppose a weight function ϕ_M is a twice continuously differentiable, positive function with bounded derivatives and does not vanish asymptotically. Let μ_γ denote the representing measure for γ. Suppose $\delta > 0$ is chosen so that $\mathrm{Supp}(\mu_\gamma) \subseteq [-1+\delta, 1-\delta]$. Then

1. *(ℓ_1-consistency)* $\|\gamma - \Pi^{\phi_M}(r_M; \delta)\|_1 \to 0$, P_x-a.s.
2. *(spectral density)* $\sup_{\omega \in [-\pi, \pi]} |\phi^W_{\delta M}(\omega) - \phi_\gamma(\omega)| \to 0$ P_x-a.s.
3. *(asymptotic variance)* $\sigma^2_{\delta M, W} \to \sigma^2(\gamma)$ P_x-a.s.

for each initial condition $x \in \mathsf{X}$, where we define $\sigma^2(\gamma) = \sum_{k \in \mathbb{Z}} \gamma(k)$

- spectral density estimate $\phi_{\delta M}$ from the unweighted momentLS satisfies the regularity conditions

Empirical performance (asymptotic variance)

Bayesian LASSO example [Rajaratnam et al., 2019]

Table: *Estimated average mean squared error (s.e.) for the asymptotic variance estimators*

(a) Asymptotic variance mean squared error[1]

	Bartlett	OBM	Init-con	IO-kernel	mLS_uw	mLS_w
β_0	333.45 (14.90)	343.99 (15.19)	162.57 (10.41)	164.04 (9.82)	163.10 (10.78)	**153.18 (10.37)**
β_1	291.61 (13.49)	300.97 (13.75)	166.04 (11.06)	160.04 (9.99)	164.35 (11.04)	**148.67 (10.02)**
β_2	368.42 (18.31)	383.17 (18.83)	227.44 (15.42)	230.95 (14.68)	229.72 (16.54)	**214.63 (15.28)**
β_3	98.26 (4.89)	101.43 (5.01)	55.91 (3.46)	58.10 (3.32)	52.96 (3.27)	**49.24 (3.14)**
β_4	63.69 (3.57)	65.62 (3.64)	41.21 (2.66)	37.00 (2.32)	38.75 (2.59)	**35.53 (2.39)**
β_5	19.65 (1.00)	20.10 (1.01)	11.89 (0.84)	12.27 (0.81)	12.06 (0.89)	**11.62 (0.84)**
β_6	141.89 (6.94)	147.49 (7.17)	87.85 (5.50)	82.63 (4.68)	86.93 (5.54)	**81.21 (5.27)**
β_7	0.54 (0.03)	0.55 (0.03)	0.71 (0.05)	0.53 (0.03)	0.49 (0.03)	**0.45 (0.03)**
σ^2	351.56 (18.28)	357.09 (18.47)	474.50 (32.61)	490.00 (17.44)	337.27 (23.94)	**311.73 (21.89)**

[1] values for β_0-β_7 are scaled by 10^4, and σ^2 is scaled by 10^2

Empirical performance (spectral density)

$$\text{Integrated squared error} = (2\pi)^{-1} \int_{[-\pi,\pi]} \{\hat{\phi}(\omega) - \phi(\omega)\}^2 \, d\omega$$

(a) spectral density mean integrated squared error[2]

	Bartlett	IO-kernel	mLS_uw	mLS_w
β_0	177.58 (6.11)	101.31 (3.95)	80.85 (3.64)	**74.76 (3.45)**
β_1	153.17 (5.42)	92.70 (3.70)	75.28 (3.49)	**68.34 (3.29)**
β_2	198.89 (6.98)	128.41 (5.34)	103.72 (5.01)	**94.98 (4.75)**
β_3	76.46 (2.53)	50.57 (1.85)	37.32 (1.66)	**34.89 (1.63)**
β_4	61.09 (1.97)	37.88 (1.46)	29.41 (1.34)	**27.04 (1.27)**
β_5	12.52 (0.50)	9.07 (0.40)	7.37 (0.38)	**7.10 (0.37)**
β_6	73.69 (2.69)	48.30 (1.85)	40.49 (1.81)	**37.58 (1.77)**
β_7	2.17 (0.06)	2.17 (0.04)	1.31 (0.05)	**1.27 (0.05)**
σ^2	113.55 (3.49)	168.96 (2.39)	84.01 (2.99)	**82.44 (2.91)**

[2] values for β_0-β_7 are scaled by 10^5, and σ^2 is scaled by 10^2

Summary

- We propose novel shape-constrained estimators for the autocovariance sequence and weighing framework resulting from a reversible Markov chain.
  - exploits the representability of the autocovariances of reversible Markov chains as the moments of a unique positive measure supported on $[-1, 1]$.
  - leverages asymptotic independence in the Fourier transform of the periodogram
- We provide a theoretical analysis of the proposed estimators, in particular, we proved
  - a.s. ℓ_2 consistency of the momentLSE sequence,
  - a.s. consistency of the asymptotic variance estimator based on the momentLSE sequence for the true asymptotic variance σ^2.
- We show our asymptotic variance estimators empirically outperform other state-of-the art estimators.

This talk is based on the following two papers:

- Berg S and Song H, Efficient shape-constrained inference for the autocovariance sequence from a reversible Markov chain, Annals of Statistics, 2023
- Song H and Berg S, Weighted shape-constrained estimation for autocovariance sequence from a reversible Markov chain, 2024+, on ArXiv soon

Thank you!

References I

Fadoua Balabdaoui and Gabriella de Fournas-Labrosse. Least squares estimation of a completely monotone pmf: From analysis to statistics. *J. Stat. Plan. Inference*, 204: 55–71, January 2020.

Fadoua Balabdaoui and Cécile Durot. Marshall lemma in discrete convex estimation. *Stat. Probab. Lett.*, 99:143–148, April 2015.

Fadoua Balabdaoui and Jon A Wellner. Estimation of a k-monotone density, part 1: characterizations consistency, and minimax lower bounds. *arXiv preprint math/0509080*, 2005.

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.

Stephen Berg and Hyebin Song. Efficient shape-constrained inference for the autocovariance sequence from a reversible markov chain. *The Annals of Statistics*, 51(6):2440–2470, 2023.

References II

Peter J Brockwell and Richard A Davis. *Time series: theory and methods*. Springer science & business media, 1991.

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

Lutz Dümbgen and Kaspar Rufibach. logcondens: Computations related to univariate Log-Concave density estimation. *J. Stat. Softw.*, 39:1–28, March 2011.

Cécile Durot, Sylvie Huet, François Koladjo, and Stéphane Robin. Least-squares estimation of a convex discrete distribution. *Computational Statistics & Data Analysis*, 67:282–298, 2013.

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.

Charles J. Geyer. Practical markov chain monte carlo. *Statistical Science*, 7(4): 473–483, 1992. ISSN 08834237.

References III

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

Felix Hausdorff. Summationsmethoden und momentfolgen. I. *Math. Z.*, 9(1):74–109, March 1921.

Hanna K Jankowski and Jon A Wellner. Estimation of a discrete monotone distribution. *Electron. J. Stat.*, 3:1567–1605, 2009.

Samuel Karlin. *Total positivity*. Stanford University Press, 1968.

Piotr Kokoszka and Thomas Mikosch. The periodogram at the fourier frequencies. *Stochastic processes and their applications*, 86(1):49–79, 2000.

Arun K Kuchibhotla, Rohit K Patra, and Bodhisattva Sen. Semiparametric efficiency in convexity constrained single index model. August 2017.

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

References IV

Bala Rajaratnam, Doug Sparks, Kshitij Khare, and Liyuan Zhang. Uncertainty quantification for modern high-dimensional regression via scalable bayesian methods. *Journal of Computational and Graphical Statistics*, 28(1):174–184, 2019.

Gareth Roberts and Jeffrey Rosenthal. Geometric ergodicity and hybrid markov chains. *Eff. Clin. Pract.*, 2(none):13–25, January 1997.

W Rudin. Functional analysis, 1973.

Hyebin Song and Stephen Berg. Weighted shape-constrained estimation for the autocovariance sequence from a reversible markov chain. 2024+.

Choice of δ

Our theoretical results cover the case of fixed tuning parameter δ satisfying $0 < \delta \leq \Delta(F)$

Empirically,

- ℓ_2 norm convergence seems to hold even with $\delta = 0$
- But convergence of the estimated asymptotic variance is lost with $\delta = 0$

In Berg and Song [2023] we suggest a rule for tuning δ, based on a modification of a batch-size estimation procedure from Politis [2003].
Under the assumption

$$\max_{k=0,\ldots,M-1} |\hat{\rho}_M(k) - \rho(k)| = O_P(\sqrt{\log M/M}) \tag{3}$$

on the sample autocorrelations, we show our rule leads to a conservative (not too large) choice of δ.

Empirical Studies

1. Empirical illustration of the convergence properties of Moment LSEs

- Recall that the Moment LSE resulting from an input sequence r_M is the projection $\Pi_\delta(r_M)$ of r_M onto the set $\mathscr{M}_\infty([-1+\delta, 1-\delta]) \cap \ell_2(\mathbb{Z})$.
- We proved the a.s. convergence of the autocovariance sequence (in L2 sense) and a.s. convergence of the asymptotic variance estimate of the moment LS estimators $\Pi_\delta(r_M)$ for any choice of $\delta > 0$ such that $\delta > 0$ and $\mathrm{Supp}(F) \subseteq [-1+\delta, 1-\delta]$.
- 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 $[-1+\delta, 1-\delta]$.

A few extra slides: Empirical Studies

Empirical illustration of the convergence properties of Moment LSEs

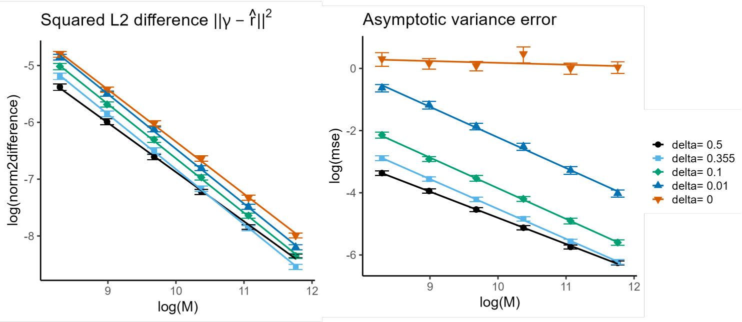


Figure: *Metropolis-Hastings example. The support of the representing measure for γ is contained in $[-.645, .645]$, i.e., the valid δ range is $0 < \delta \leq .355$.*

A few extra slides: Empirical Studies

Empirical illustration of the convergence properties of Moment LSEs

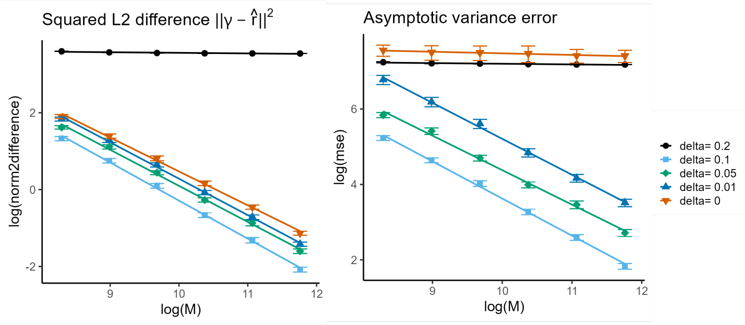


Figure: *AR(1) example with* $\rho = .9$. *The representing measure has a single support point at* $.9$. *The valid* δ *range is* $0 < \delta \leq .1$.

A few extra slides: Empirical Studies

Comparison with other state-of-the-art estimators

For the Bartlett windowed estimators, BM, OBM, and Moment LSEs, hyperparameters are required. We used oracle hyperparameter settings:

- From Flegal and Jones [2010], for the BM and OLBM methods, the mean-squared-error optimal batch sizes for estimating $\sigma^2(\gamma)$ are

$$b_M^{(\mathrm{BM})} = \left(\frac{\Gamma^2 M}{\sigma^2(\gamma)}\right)^{1/3} = C_2 M^{1/3} \quad \text{and} \quad b_M^{(\mathrm{OLBM})} = \left(\frac{8\Gamma^2 M}{3\sigma^2(\gamma)}\right)^{1/3} = C_3 M^{1/3}$$

  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 OLBM [Damerdji, 1991], we let $C_1 = C_3$.
- For the choice of oracle δ, we let $\delta = 1 - \sup|\mathrm{Supp}(F)|$