Statistical and computational aspects of shape-constrained inference for covariance function estimation

Stephen Berg

Department of Statistics, Penn State University

Joint work with Hyebin Song (Stat, Penn State)

Introduction: MCMC for Bayesian statistics

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

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

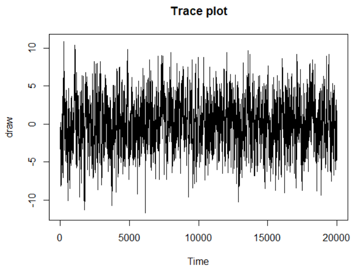
Application: autocovariance sequence estimation

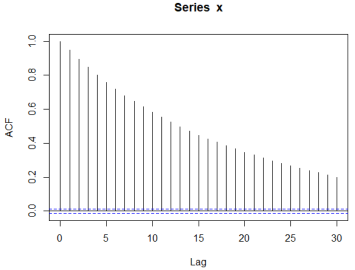
- The autocovariance sequence $\gamma = \{\gamma(k)\}_{k \in \mathbb{Z}}$, defined as

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

characterizes second order properties of a stationary time series $\{g(X_t)\}_{t \in \mathbb{Z}}$.

- Estimation of γ plays a key role in time series analysis and Markov Chain Monte Carlo (MCMC) simulation
  - E.g., informative diagnostic plot for convergence in MCMC simulation, spectral density estimation, etc.





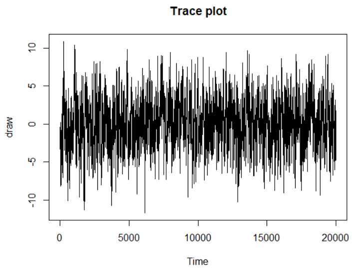
Introduction: autocovariance sequence estimation

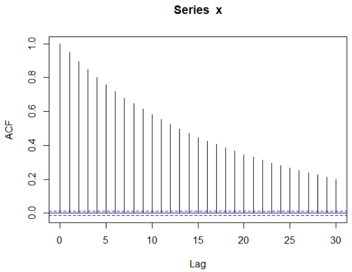
- For a given sample $\{g(X_i)\}_{i=0}^{M-1}$ of size M, the empirical autocovariance sequence $r_M = \{r_M(k)\}_{k\in\mathbb{Z}}$ defined as

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

is a natural estimator for $\gamma = \{\gamma(k)\}_{k\in\mathbb{Z}}$, where

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



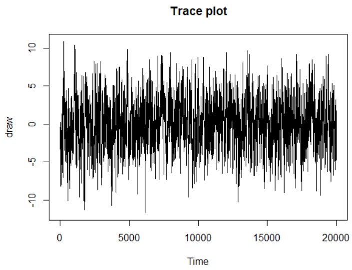


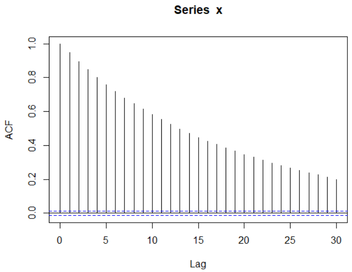
Introduction: autocovariance sequence estimation

Goal for today's talk: L2-consistent estimation of autocovariance sequence γ where

$$\gamma(k) = \mathrm{Cov}(g(X_0), g(X_k)), \quad \forall k$$

and $X_0, X_1, \ldots$ is a π-reversible Markov chain, using regularization based on shape constraints





Asymptotics

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

- Under mild conditions [e.g., Meyn and Tweedie [2009]], a central limit theorem can be established for $Y_M = \frac{1}{M}\sum_{t=0}^{M-1} g(X_t)$ such that

$$\sqrt{M}(Y_M - E_\pi[g]) \xrightarrow{d} N(0, \sigma^2)$$

  where $\sigma^2 = \sum_{k=-\infty}^{\infty} \gamma(k)$ and $\gamma(k) = \mathrm{Cov}_\pi(g(X_0), g(X_{|k|}))$.
- The asymptotic variance σ^2 quantifies the uncertainty of the estimate of $E_\pi[g]$ from an MCMC simulation.

Variance of empirical mean

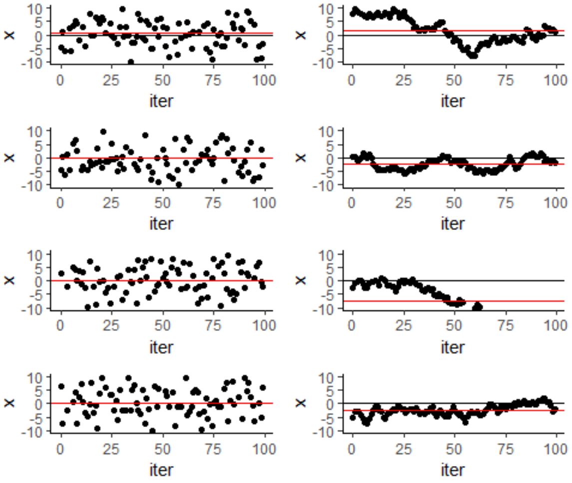


Figure: iid (left) and AR(1) (right) samples from the same $N(0,(1-0.99^2)^{-1})$ distribution

Asymptotic variance estimation

- Some natural estimators of σ^2 turn out to be inconsistent.
  - For example, simply summing the empirical autocovariances

$$\hat{\sigma}^2_{Emp} = \sum_{k=-\infty}^{\infty} r_M(k)$$

leads to an inconsistent estimator of σ^2.

- Several estimation methods proposed for estimating σ^2 with better statistical properties (e.g., consistency, $M^{1/3}$ convergence)
  - Spectral variance estimators [Anderson, 1971, Damerdji, 1991]:

$$\hat{\sigma}^2_{SV} = \sum_{k=-B_M}^{B_M} w_M(k) r_M(k)$$

for a properly chosen window function $w_M(k)$ such that $w_M(k) = 0$ for $k > B_M$.

  - Batch means and overlapping batch means estimators [Priestley, 1981, Flegal and Jones, 2010, Chakraborty et al., 2022]

$$\hat{\sigma}^2_{BM} = \frac{\lfloor M/B \rfloor}{B} \sum_{b=0}^{B-1} (\bar{Y}_b - \bar{Y}_M)^2$$

Initial sequence estimators

- Geyer [1992] introduces "initial sequence estimators" for estimating the asymptotic variance.

- The initial sequence estimators exploit positivity, monotonicity, and convexity constraints on certain summed autocovariances of reversible Markov chains. In particular, let

$$\Gamma(k) := \gamma(2k) + \gamma(2k+1) \qquad k = 0, 1, 2, \ldots$$

  - $\Gamma(k)$ are positive ($\Gamma(k) \geq 0$), monotone ($\Gamma(k) \geq \Gamma(k+1)$), and convex ($\Gamma(k) + \Gamma(k+2) \geq 2\Gamma(k+1)$) [Geyer, 1992]

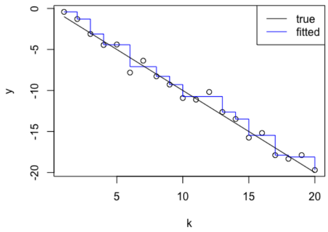
- The idea of Geyer [1992] is to estimate summed autocovariance sequences $\Gamma(k)$ by imposing these shape constraints.

Estimation with shape constraints

The work of Geyer [1992] can be considered as an example of shape-constrained inference. 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 $k = 1, \ldots, n$.

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



  - Single index model with monotonicity constraint [Kakade et al., 2011, Ganti et al., 2015, Dai et al., 2022]: $y_k = f(x_k^\top \beta) + \epsilon_k$, $f : \mathbb{R} \to \mathbb{R}$ is monotone

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]]

Connection with moment problems

Moment problem: given a sequence $m \in \mathbb{R}^{\mathbb{N}}$, is there any measure μ such that $m(k) = E_{X\sim\mu}[X^k]$, for all $k = 0, 1, 2, \ldots$?

- There is a definite answer for the moment problem.
- Moreover, turns out, some "shape constraints" of a sequence m are closely related to the properties of a representing measure μ for m

Theorem (Hausdorff moment theorem [Hausdorff, 1921])

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

In short, $[0, 1]$-moment sequence $\iff$ completely monotone

Connection with moment problems

- 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 $[-1+\delta, 1-\delta]$ for some $\delta > 0$ (true for e.g., an IID sample or a reversible chain with geometric ergodicity [Roberts and Rosenthal, 1997]).

Our approach

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

Our estimator (Moment LSE): for an input sequence r_M,

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

- projection onto ℓ_2 moment sequence set

Computation

Objective: minimize $L(\mu; r_M)$ over μ, where

$$L(\mu; r_M) = \sum_{k \in \mathbb{Z}} (r_M(k) - \int x^{|k|} \mu(dx))^2 \tag{3}$$

subject to μ a positive measure with $\mathrm{Supp}(\mu) \subseteq [-1+\delta, 1-\delta]$.

- For any input sequence r_M such that $|\{k; r_M(k) \neq 0\}| < \infty$, the representing measure for $\Pi_\delta(r_M)$ is discrete, and its support contains at most finite number of points [Berg and Song, 2023].
- A support reduction algorithm [Groeneboom et al., 2008] can be used for optimizing (3).

Computation

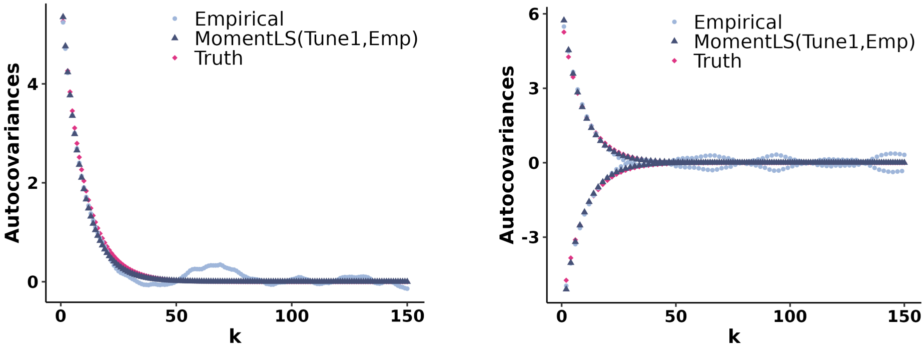
- For $r \in \ell_2(\mathbb{Z})$, define $\Pi(r;\Theta)$ as the projection of r onto set of Θ-moment sequences (moment sequence for a measure supported on Θ)
- approximate $\Pi(r;\Theta)$ by $\Pi(r;C)$ where $C = \{\alpha_1, ..., \alpha_s\} \subset \Theta$
  - C is a finely spaced "grid"
- turns projection problem into optimization over measures

$$\mu = \sum_{i=1}^{s} w_i \delta_{\alpha_i}$$

  where w_i are nonnegative
- computing $\Pi(r;C)$ is a quadratic programming problem similar to non-negative least squares:

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

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* $M = 8000$.

Assumptions

Consider a Markov chain $\{X_t\}$ on $(X, \mathcal{X})$ with a transition kernel $Q : X \times \mathcal{X} \to [0,1]$ and the stationary probability measure π. Let g be a function such that $\int g^2(x)\pi(dx) < \infty$. Let γ denote the autocovariance sequence of $g(X_t)$, i.e., $\gamma(k) = \mathrm{Cov}_\pi(g(X_0), g(X_k))$.

Assumptions:

1. (Assumptions on the chain) The kernel Q is ψ-irreducible, aperiodic, π-reversible, and geometrically ergodic.
2. (Assumptions on an input sequence r_M) r_M is an even function with a peak at 0 with a finite support, and $r_M^{\text{init}}(k) \underset{M\to\infty}{\to} \gamma(k)$ almost surely for each $k \in \mathbb{Z}$.

Statistical guarantee

Theorem ([Berg and Song, 2023])

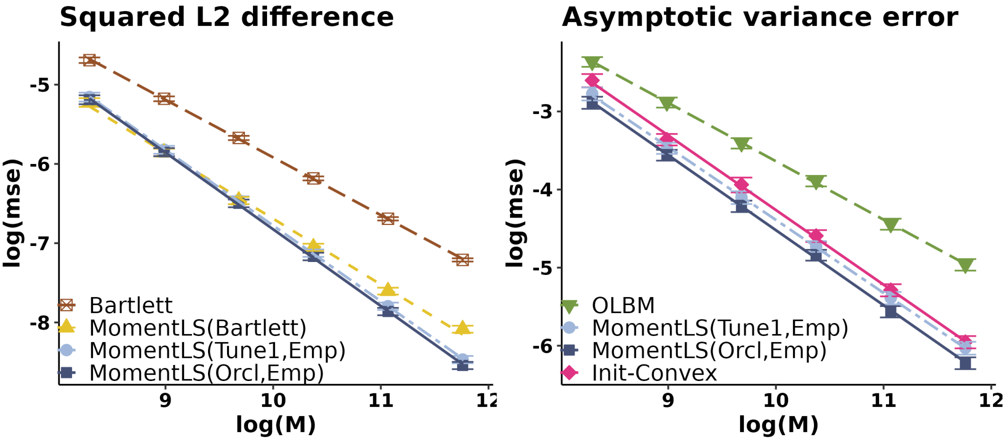
Consider a Markov chain $X_0, X_1, \ldots$ *and an input sequence* r_M *satisfying the aforementioned conditions. Let* F *denote the representing measure for* γ*. Suppose* $\delta > 0$ *is chosen so that* $0 < \delta \leq \Delta(F)$*. Then*

1. *(ℓ_2-consistency of the Moment LSE)* $\|\gamma - \Pi_\delta(r_M)\|^2 \underset{M\to\infty}{\to} 0, \;\; P_x$*-a.s.*
2. *(vague convergence of* $\hat{\mu}_{\delta,M}$*)* $P_x(\hat{\mu}_M \to F_g$ *vaguely, as* $M \to \infty) = 1$*, where* $\hat{\mu}_M$ *and* F *are the representing measures for* $\Pi_\delta(r_M)$ *and* γ*, and*
3. *(a.s. convergence of* $\hat{\sigma}^2$*)* $\sigma^2(\Pi_\delta(r_M)) \to \sigma^2(\gamma) \;\; P_x$*-a.s.*

for each initial condition $x \in \mathsf{X}$*, where we define* $\sigma^2(m) = \sum_{k\in\mathbb{Z}} m(k)$ *for a sequence* m *on* $\mathbb{Z}$*.*

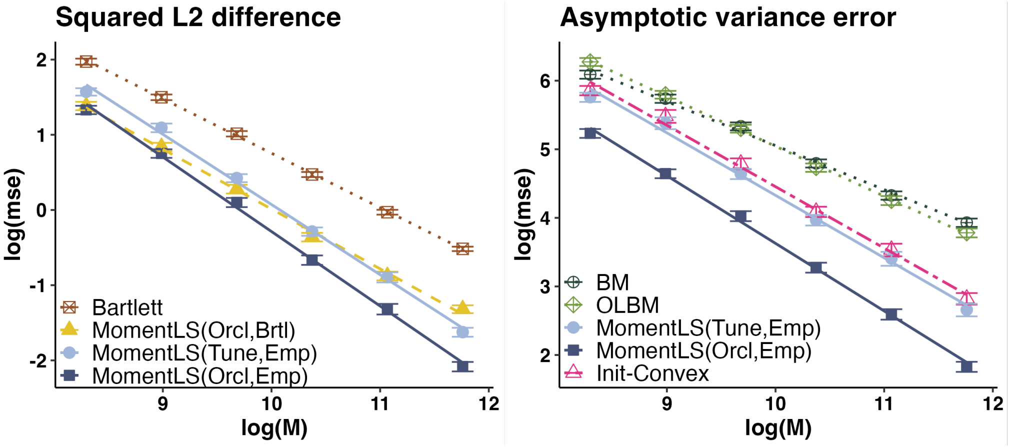
Empirical Studies

Metropolis-Hastings example:



Empirical Studies

AR(1) example with $\rho = 0.9$:



Weighted

- currently: covariance fitting objective $\|r_m - m\|^2 = \sum_{k \in \mathbb{Z}} (r_M(k) - m(k))^2$
- what about a weighted squared error loss function?
  - $Y = X\beta + \epsilon, \qquad Var(\epsilon) = \Sigma$
  - $\hat{\beta}_{OLS} = (X^\top X)^{-1} X^\top Y$
    $\hat{\beta}_{GLS} = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} Y$
- Covariances of the empirical autocovariances $r_M(k)$ are simpler on the Fourier transform scale:

$$\hat{\phi}_M(\omega) = \sum_{k=-(M-1)}^{M-1} r_M(k) \exp(-ik\omega)$$ Periodogram (sample)

$$\phi(\omega) = \sum_{k \in \mathbb{Z}} \gamma(k) \exp(-ik\omega)$$ Spectral density (population)

Periodogram asymptotics

- Periodogram at Fourier frequencies:

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

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

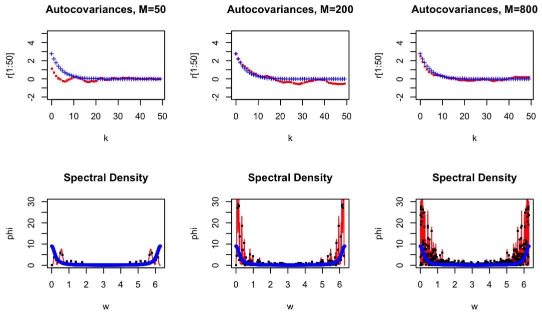


Figure: Autocovariances and spectral densities from an AR(1) process

Weighted loss function

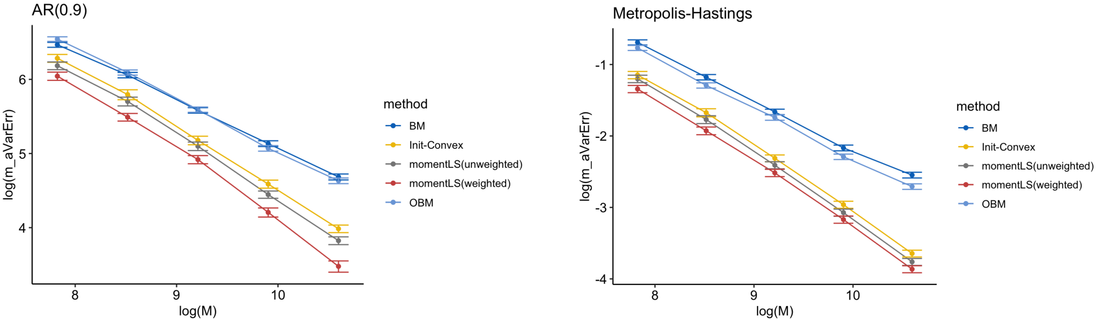
Objective function:

- unweighted covariance fitting objective $\sum_{k\in\mathbb{Z}}(r_M(k) - m(k))^2$
- by Parseval equality, $\|r_M - m\|^2 = (2\pi)^{-1}\int_{-\pi}^{\pi}(\hat{\phi}_M(\omega) - \hat{m}(\omega))^2\,d\omega$, where
  - $\hat{\phi}_M(\cdot)$ is the sample spectral density
  - $\hat{m}(\cdot)$ is the fitted spectral density (discrete time Fourier transform of m)
- suggests a weighted loss

$$\|r_M - m\|^2 = (2\pi)^{-1}\int_{-\pi}^{\pi}\left\{\frac{\hat{\phi}_M(\omega) - \hat{m}(\omega)}{\tilde{\phi}_M(\omega)}\right\}^2 d\omega$$

where $\tilde{\phi}_M(\omega)$ is a good estimate of true spectral density

Empirical performance



(a) AR(1), $\rho = 0.9$ (b) Metropolis-Hastings

Figure: *For an AR(1) and Metropolis-Hasting chain, a comparison of mean squared error for estimating the asymptotic variance* σ^2, *for* $M \in \{2500, 5000, ..., 40000\}$

Shape constraints for spatial covariance functions

How about "shape constraints" (or mixture representations) for covariance functions for a random field on $\mathbb{R}^d$?

- It is well known that the function $\gamma : [0, \infty) \to \mathbb{R}$ leads to a valid covariance function of the form $C(x_i, x_j) = \gamma(\|x_i - x_j\|)$ for $x \in \mathbb{R}^d$ in each dimension $d \geq 1$, if and only if the function γ admits a mixture representation of the form

$$\gamma(s) = \int_{[0,\infty)} \exp\left(-r^2 s^2\right) F(dr)$$

  [see, e.g., Gneiting 1999]
  - exploited in Choi et al. [2013] and Wang and Ghosh [2023]
  - Implies $\gamma(\sqrt{t})$ is completely monotone
- A parametric example: $C(x, y) = \gamma(\|x - y\|)$ is a function in Matern Kernel class [Stein, 1999]. Indeed, there exists a parametric $f(r; \rho, \nu)$ such that

$$\gamma(s) = k_M(s; \rho, \nu) = \int \exp(-r^2 s^2) f(r; \rho, \nu) dr$$

  [Tronarp et al., 2018]

Summary

- In this work, we propose a novel shape-constrained estimator of the autocovariance sequence resulting from a reversible Markov chain.
- The proposed estimator (MomentLSE) exploits the representability of the autocovariances of reversible Markov chains as the moments of a unique positive measure supported on $[-1, 1]$.
- We provide a theoretical analysis of the MomentLSE, in particular, we proved
  - a.s. ℓ_2 consistency of the momentLSE sequence,
  - a.s. vague convergence of the representing measure of the momentLSE sequence, and
  - a.s. consistency of the asymptotic variance estimator based on the momentLSE sequence for the true asymptotic variance σ^2.

Thank you!

References I

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. *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.

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.

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.

InKyung Choi, Bo Li, and Xiao Wang. Nonparametric estimation of spatial and space-time covariance function. *Journal of Agricultural, Biological, and Environmental Statistics*, 18:611–630, 2013.

Ran Dai, Hyebin Song, Rina Foygel Barber, and Garvesh Raskutti. Convergence guarantee for the sparse monotone single index model. *Electronic Journal of Statistics*, 16(2):4449–4496, 2022.

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.

References III

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.

Ravi Ganti, Nikhil Rao, Rebecca M Willett, and Robert Nowak. Learning single index models in high dimensions. *arXiv preprint arXiv:1506.08910*, 2015.

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

Tilmann Gneiting. Radial positive definite functions generated by Euclid's hat. *Journal of Multivariate Analysis*, 69(1):88–119, 1999.

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.

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

References IV

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

Sham M Kakade, Varun Kanade, Ohad Shamir, and Adam Kalai. Efficient learning of generalized linear and single index models with isotonic regression. *Advances in Neural Information Processing Systems*, 24, 2011.

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.

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.

References V

M.B. Priestley. *Spectral Analysis and Time Series*. Number Bd. 1-2 in Probability and mathematical statistics : a series of monographs and texbooks. Academic Press, 1981. ISBN 9780125649223.

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.

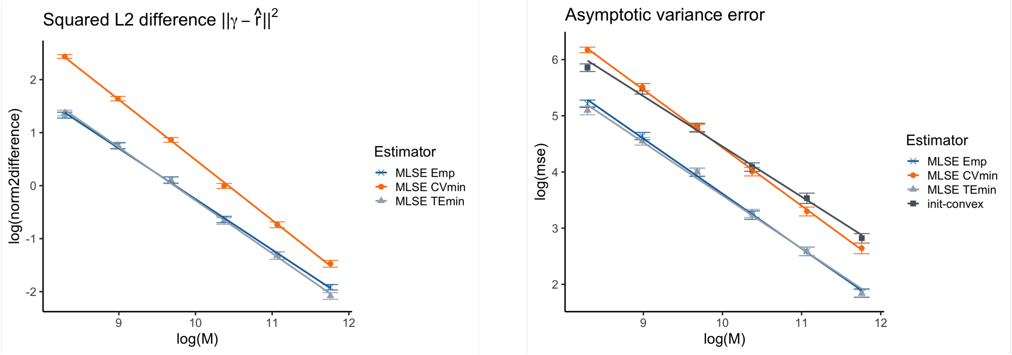
Michael L Stein. *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media, 1999.

Filip Tronarp, Toni Karvonen, and Simo Särkkä. Mixture representation of the matérn class with applications in state space approximations and bayesian quadrature. In *2018 IEEE 28th International Workshop on Machine Learning for Signal Processing (MLSP)*, pages 1–6. IEEE, 2018.

Yiming Wang and Sujit K Ghosh. Nonparametric estimation of isotropic covariance function. *Journal of Nonparametric Statistics*, 35(1):198–237, 2023.

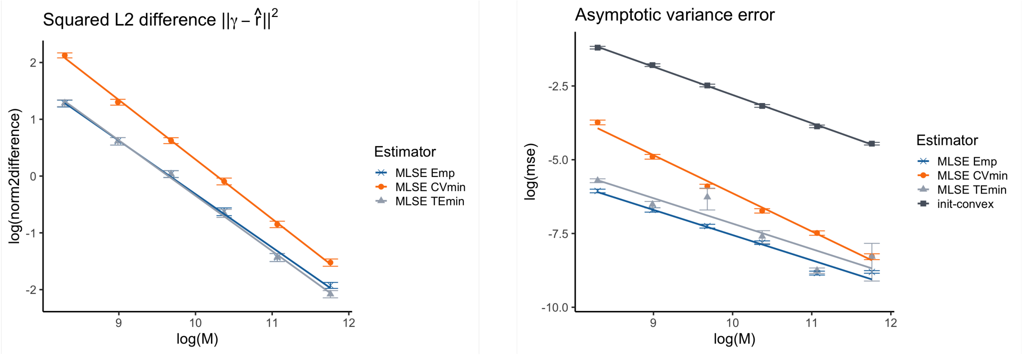
A few additional slides

AR(1) example ($\rho = .9$). Performance of Moment LSE oracle δ (Emp), δ chosen by minimizing estimated loss functions from 10-fold cross-validation (CVmin) and 10 independent chains (TEmin).



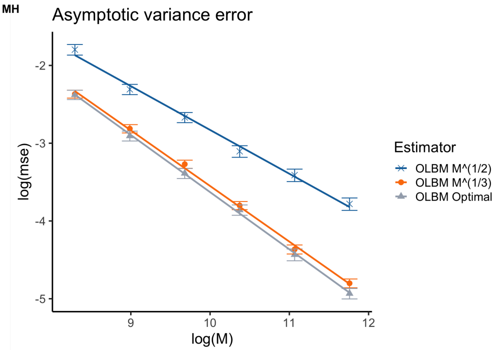
A few additional slides

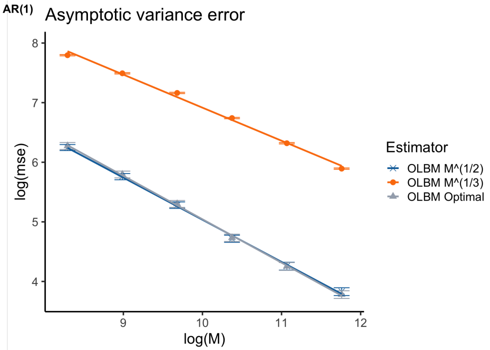
AR(1) example ($\rho = -.9$). Performance of Moment LSE oracle δ (Emp), δ chosen by minimizing estimated loss functions from 10-fold cross-validation (CVmin) and 10 independent chains (TEmin).



A few additional slides

Comparison of performance of OLBM estimators when batch size $= M^{1/3}$, $M^{1/2}$, and optimal batch size.





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

$$\max_{k=0,\ldots,M-1} |\hat{\rho}_M(k) - \rho(k)| = O_P(\sqrt{\log M/M}) \tag{4}$$

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

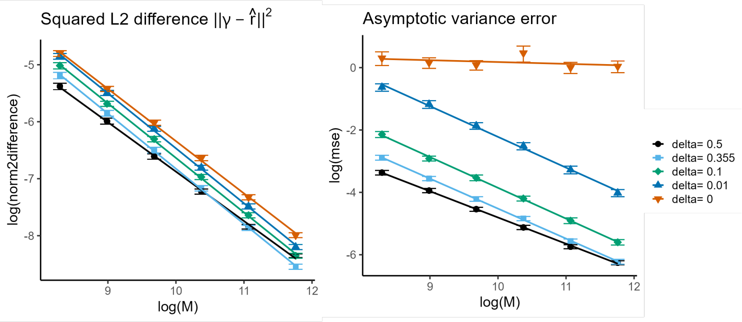


Figure: *Metropolis-Hastings example. The support of the representing measure for γ is contained in $[-.645, .645]$, i.e., the valid δ range is $0 < \delta \leq .355$.*

A few extra slides: Empirical Studies

Empirical illustration of the convergence properties of Moment LSEs

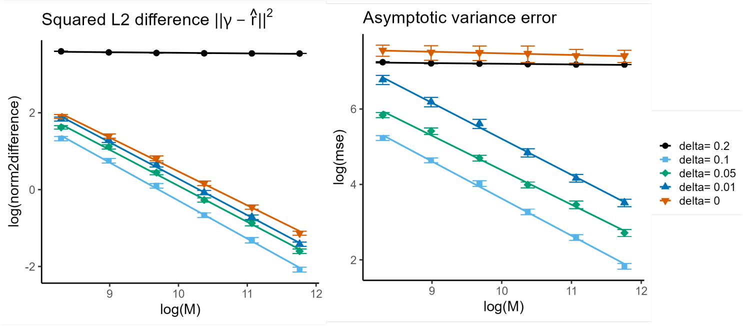


Figure: *AR(1) example with* $\rho = .9$. *The representing measure has a single support point at* $.9$. *The valid* δ *range is* $0 < \delta \leq .1$.

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 $\delta = 1 - \sup|\mathrm{Supp}(F)|$