Title

Fixed-Domain Asymptotics Under Vecchias Approximation of Spatial Process Likelihoods.

Permalink
https://escholarship.org/uc/item/47p289r3

Journal
Statistica Sinica, 34(4)

ISSN
1017-0405

Authors
Zhang, Lu
Tang, Wenpin
Banerjee, Sudipto

Publication Date
2024-10-01

DOI
10.5705/ss.202021.0428

Peer reviewed

Published in final edited form as:
Stat Sin. 2024 October ; 34(4): 1863–1881. doi:10.5705/ss.202021.0428.

Fixed-Domain Asymptotics Under Vecchia's Approximation of Spatial Process Likelihoods

Lu Zhang,
Wenpin Tang,
Sudipto Banerjee
University of Southern California, Los Angeles, Columbia University, New York, University of California, Los Angeles

Abstract

Statistical modeling for massive spatial data sets has generated a substantial literature on scalable spatial processes based upon Vecchia's approximation. Vecchia's approximation for Gaussian process models enables fast evaluation of the likelihood by restricting dependencies at a location to its neighbors. We establish inferential properties of microergodic spatial covariance parameters within the paradigm of fixed-domain asymptotics when they are estimated using Vecchia's approximation. The conditions required to formally establish these properties are explored, theoretically and empirically, and the effectiveness of Vecchia's approximation is further corroborated from the standpoint of fixed-domain asymptotics.

Key words and phrases:

Fixed-domain asymptotics; Gaussian processes; Matérn covariance function; Microergodic parameters; Spatial statistics

1. Introduction

Geostatististical data are often modeled by treating observations as partial realizations of a spatial random field. We customarily model the random field $\{Y(s): s \in \mathscr{D}\}$ over a bounded region $\mathscr{D} \in \mathbb{R}^d$ as a Gaussian process (GP), denoted $Y(s) \sim GP(\mu_\beta(s), K_\theta(\cdot\,,\,\cdot))$, with mean $\mu_\beta(s)$ and covariance function $K_\theta(s, s') = \mathrm{cov}(y(s_i), y(s_j))$. The probability law for a finite set $\chi = \{s_1, s_2, \ldots, s_n\}$ is given by $y \sim N(\mu_\beta, K_\theta)$, where $y = (y(s_i))$ and $\mu_\beta = (\mu_\beta(s_i))$ are $n \times 1$ vectors with elements $y(s_i)$ and $\mu_\beta(s_i)$, respectively, and $K_\theta = (K_\theta(s_i, s_j))$ is the $n \times n$ spatial covariance matrix whose (i, j)th element is the value of the covariance function $K_\theta(s_i, s_j)$. We consider the widely employed stationary Matérn covariance function [Matérn, 1986, Stein, 1999b] given by

lzhang63@usc.edu .
Division of Biostatistics, Department of Population and Public Health Sciences, University of Southern California, Los Angeles, CA 90089, U.S.A.
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027, U.S.A.
Department of Biostatistics, University of California, Los Angeles, CA 90095 U.S.A.
Equally contributing authors.

$$K_{\theta}(s, s') := \frac{\sigma^2(\phi \parallel h \parallel)^{\nu}}{\Gamma(\nu)2^{\nu - 1}}\mathcal{K}_{\nu}(\phi \parallel h \parallel), \ \parallel h \parallel \ \geq 0, \tag{1.1}$$

where $h = s - s'$, $\sigma^2 > 0$ is called the partial sill or spatial variance, $\phi > 0$ is the scale or decay parameter, $\nu > 0$ is a smoothness parameter, $\Gamma(\cdot)$ is the Gamma function, $\mathcal{K}_{\nu}(\cdot)$ is the modified Bessel function of order ν [Abramowitz and Stegun, 1965, Section 10] and $\theta = \left\{\sigma^2, \phi, \nu\right\}$. The spectral density corresponding to (1.1), which we will need later, is

$$f\left(\left|u\right|\right) = C\frac{\sigma^2\phi^{2\nu}}{\left(\phi^2 + u^2\right)^{\nu + \frac{d}{2}}} \text{ for some } C > 0\,. \tag{1.2}$$

Likelihood-based inference for θ will require matrix computations in the order of $\sim n^3$ floating point operations (flops) and can become impracticable when the number of spatial locations, n, is very large. Writing $Y_n = (y_1, y_2, \ldots, y_n)^{\top}$, where $y_i := y(s_i)$ for $i = 1,2, \ldots, n$ are the n sampled measurements, we write the joint density $p(Y_n \mid \theta) := N(Y_n; \mu_{\beta}, K_{\theta})$ as

$$p(Y_n \mid \theta) = p(y_1; \theta)\prod_{i = 2}^{n} p(y_i \mid y_{(i - 1)}; \theta), \tag{1.3}$$

where $y_{(i)} = (y_1, \ldots y_i)$. Vecchia [1988] suggested a simple approximation to (1.3) based upon the notion that it may not be critical to use all components of $y_{(i - 1)}$ in $p(y_i \mid y_{(i - 1)}; \theta)$. Instead, the joint density $p(Y_n \mid \theta)$ in (1.3) is approximated by

$$\widetilde{p}(Y_n \mid \theta) = p(y_1 \mid \theta)\prod_{i = 2}^{n} p(y_i \mid S_{(i - 1)}; \theta), \tag{1.4}$$

where $S_{(i)}$ is a subvector of $y_{(i)}$ for $i = 1, \ldots, n$. The density $\widetilde{p}(Y_n \mid \theta)$ in (1.4) is called Vecchia's approximation and can be regarded as a quasi- or composite likelihood [Zhang, 2012, Eidsvik et al., 2014, Bachoc and Lagnoux, 2020]. Vecchia's approximation has attracted a remarkable amount of attention in recent times, already too vast to be comprehensively reviewed here [see, e.g., Stein et al., 2004, Datta et al., 2016a,b, Guinness, 2018, Katzfuss et al., 2020, Katzfuss and Guinness, 2021, Peruzzi et al., 2022]. Algorithmic developments in Bayesian and frequentist settings [Finley et al., 2019, Zhang et al., 2019, Katzfuss et al., 2020] have enabled scalability to massive data sets (with $n \sim 10^7$ locations) and (1.4) lies at the core of several methods that tackle “big data” problems in geospatial analysis [Sun et al., 2012, Banerjee, 2017, Heaton et al., 2019].

The Vecchia approximation has recently garnered substantial attention in the spatial statistics literature as an edifice for building massively scalable Gaussian process models. While substantial methodological innovation has been generated by this approach, developing a theoretical understanding regarding the inference and identifiability of the spatial process parameters has remained largely unaddressed. This is because the Vecchia approximation distorts the stationarity of the parent process and, hence, the theoretical tractability of the spatial processes are lost. Our current approach is an original first attempt based upon Zhang [2012] to formally introduce methods that can study the asymptotic properties of inference from Vecchia’s approximation. While a completely rigorous development is available only in the one-dimensional setting, we emphasize that the approach we develop is novel and should generate subsequent theoretical research in two-dimensions. Therefore, we limit the formal theory to one-dimension but present some insightful numerical experiments in two-dimensions to show that the inferential behavior secured over the real line will be expected to carry over to spatial domains.

Following the fixed-domain (infill) asymptotic paradigm for spatial inference [Stein, 1999a, Zhang and Zimmerman, 2005] we discuss inferential properties for the parameters in (1.1). In this setting, Zhang [2004] showed that not all parameters in θ admit consistent maximum likelihood estimators from the full Gaussian likelihood in (1.3) constructed with a stationary Matérn covariance function, but certain *microergodic* parameters are consistently estimated. Du et al. [2009, Theorem 5] formally established the asymptotic distributions of these microergodic parameters. Kaufman and Shaby [2013] addressed jointly estimating the decay and the variance parameters in the Matérn family and the effect of a prefixed decay on inference when having relatively small sample size. All of the aforementioned work has been undertaken using (1.3). Here, we formally establish the inferential properties for the estimates of microergodic parameters obtained from Vecchia’s approximate likelihood in (1.4). We build our work on a brief but insightful discussion in Section 10.5.3 of Zhang [2012], regarding the inferential behaviour arising from (1.4). To the best of our knowledge such explorations have not hitherto been formally undertaken. Following the aforementioned works in spatial asymptotics, we will restrict attention to the infill or fixed domain setting and focus on the inferential properties of the microergodic parameters for any given value of the smoothness parameter. More specifically, we explore the criteria for asymptotic normality for the maximum likelihood estimates of the microergodic parameters obtained from Vecchia’s approximation. In this regard, our work follows the paradigm laid out in Zhang [2012] in that we can no longer assume that the conditioning set is bounded. We provide conditions under which the inference under the Vecchia approximations of the Matérn process will be asymptotically equivalent to the full model. This distinguishes our intended contribution from that in Bachoc and Lagnoux [2020], where bounded conditional sets are exploited to establish consistency results for some selected values of the smoothness parameter. On the other hand, we show that a different set of conditions can yield a closed form asymptotic distribution for any given value of the smoothness parameter. For the subsequent development, it suffices to assume that $\mu_\beta(s) = 0$, i.e., the data has been detrended. Hence, we work with a zero-centered stationary Gaussian process with the Matérn covariance function in (1.1), a fixed smoothness parameter ν and with the sampling locations χ_n restricted to a bounded region.

The balance of this article is arranged as follows. One of our key results, Theorem 1, is presented in Section 2, providing general criteria for asymptotic normality of maximum likelihood estimates of microergodic parameters obtained from Vecchia's approximation. In Section 3, we demonstrate that these general criteria are implied by a condition on the conditioning size which grows much slower than the sample size. We numerically check the conclusions for one-dimensional cases and extend the discussion for two-dimensional cases in Section 4.

2. Infill Asymptotics for Vecchia's approximation

2.1 Microergodic parameters

Identifiability and consistent estimation of θ in (1.1) relies upon the equivalence and orthogonality of Gaussian measures. Two probability measures P_1 and P_2 on a measurable space $(\Omega, \mathscr{F})$ are said to be *equivalent*, denoted $P_1 \equiv P_2$, if they are absolutely continuous with respect to each other. Thus, $P_1 \equiv P_2$ implies that for all $A \in \mathscr{F}$, $P_1(A) = 0$ if and only if $P_2(A) = 0$. On the other hand, P_1 and P_2 are orthogonal, denoted $P_1 \perp P_2$, if there exists $A \in \mathscr{F}$ for which $P_1(A) = 1$ and $P_2(A) = 0$. While measures may be neither equivalent nor orthogonal, Gaussian measures are in general one or the other. For a Gaussian probability measure P_θ indexed by a set of parameters θ and κ, a function of θ, we say that $\kappa(\theta)$ is *microergodic* if $\kappa(\theta_1) \neq \kappa(\theta_2)$ implies $P_{\theta_1} \perp P_{\theta_2}$ [see, e.g., Stein, 1999b, Zhang, 2012]. Two Gaussian probability measures defined by Matérn covariance functions $K_{\theta_1}(h)$ and $K_{\theta_2}(h)$, where $\theta_1 = \left\{\sigma_1^2, \phi_1, \nu\right\}$ and $\theta_2 = \left\{\sigma_2^2, \phi_2, \nu\right\}$ are equivalent if and only if $\sigma_1^2\phi_1^{2\nu} = \sigma_2^2\phi_2^{2\nu}$ [Theorem 2 in Zhang, 2004]. Consequently, one cannot consistently estimate σ^2 or ϕ [Corollary 1 in Zhang, 2004] from full Gaussian process likelihood functions, $\sigma^2\phi^{2\nu}$ is a microergodic parameter that can be consistently estimated.

If the oracle (data generating) values of ϕ and σ^2 are ϕ_0 and σ_0^2, respectively, then for any fixed value of the decay $\phi = \phi_1$, we know from Du et al. [2009, Theorem 5] that

$$\sqrt{n}\left(\hat{\sigma}_n^2\phi_1^{2\nu} - \sigma_0^2\phi_0^{2\nu}\right) \xrightarrow{\mathscr{L}} N\left(0, 2\left(\sigma_0^2\phi_0^{2\nu}\right)^2\right), \tag{2.5}$$

where $\hat{\sigma}_n^2$ is the maximum likelihood estimator from the full likelihood (1.3).

2.2 Parameter estimation

Let $\hat{\sigma}_{n,vecch}^2$ be the maximum likelihood estimate of the variance σ^2,

$$\hat{\sigma}_{n,vecch}^2 = \text{argmax}_{\sigma^2}\left\{\tilde{p}\left(Y_n \mid \phi_1, \sigma^2\right), \sigma^2 \in \mathbb{R}^+\right\},$$

(2.6)

where $\widetilde{p}(\cdot)$ is the density (1.4). We develop the asymptotic equivalence of $\hat{\sigma}^2_{n,vecch}$ with $\hat{\sigma}^2_n$. To proceed further, we introduce some notations. Assume that the target process $y(s) \sim GP(0, K_\theta(\cdot))$, where $K_\theta(h)$ is defined in (1.1) with a fixed ν. Let P_j, $j = 0, 1$ denote probability measures for $y(s) \sim GP(0, K_{\theta_i})$ with $\theta_j = \{\sigma^2_j, \phi_j, \nu\}$. Assume that $\sigma^2_1 = \sigma^2_0 \phi_0^{2\nu} / \phi_1^{2\nu}$ and let $E_j(\cdot)$ denote the expectation with respect to probability measure P_j, $j = 0, 1$. We define

$$\begin{gathered} e_{0,j} := y_1, \mu_{i,j} := E_j\left(y_i \mid y_{(i-1)}\right), e_{i,j} := y_i - \mu_{i,j}, i = 2, \ldots, n \\ \widetilde{e}_{0,j} := y_1, \widetilde{\mu}_{i,j} := E_j\left(y_i \mid S_{(i-1)}\right), \widetilde{e}_{i,j} := y_i - \widetilde{\mu}_{i,j}, i = 2, \ldots, n\,. \end{gathered} \tag{2.7}$$

In Lemma 1 we derive a useful expression for $\hat{\sigma}^2_{n,vecch}$ using the quantities in (2.7).

Lemma 1. *The estimate of* σ^2 *from Vecchia's likelihood approximation with fixed* ν, *and* $\phi = \phi_1$ *can be expressed as*

$$\hat{\sigma}^2_{n,vecch} = \frac{\sigma^2_1}{n} \sum_{i=1}^{n} \frac{\widetilde{e}^2_{i,1}}{E_1 \widetilde{e}^2_{i,1}}\,. \tag{2.8}$$

Proof. In Vecchia's approximation (1.4) with fixed ν, $\phi = \phi_1$ and unknown σ^2 in $K_\theta(\cdot)$, $p\left(y_i \mid S_{(i-1)}\right)$ is Gaussian with mean $\widetilde{\mu}_{i,1} = \widetilde{\Sigma}^{12}_{i,1}\left(\widetilde{\Sigma}^{22}_{i,1}\right)^{-1} S_{(i-1)}$ and variance $\widetilde{\Sigma}_{i,1} := \widetilde{\Sigma}^{11}_{i,1} - \widetilde{\Sigma}^{12}_{i,1}\left(\widetilde{\Sigma}^{22}_{i,1}\right)^{-1}\widetilde{\Sigma}^{21}_{i,1}$, where $\begin{pmatrix} \widetilde{\Sigma}^{11}_{i,1} & \widetilde{\Sigma}^{12}_{i,1} \\ \widetilde{\Sigma}^{21}_{i,1} & \widetilde{\Sigma}^{22}_{i,1} \end{pmatrix}$ is the covariance matrix of $\begin{pmatrix} y_i \\ S_{(i-1)} \end{pmatrix}$ under $\widetilde{p}\left(\cdot\,; \phi_1, \sigma^2\right)$. Since $\widetilde{\mu}_{i,1}$ does not depend on σ^2, and $\widetilde{\Sigma}_{i,1}$ can be expressed as $\sigma^2 \widetilde{\Sigma}^{\dagger}_{i,1}$, where $\widetilde{\Sigma}^{\dagger}_{i,1}$ does not depend upon σ^2, the conditional distributions $p\left(y_i \mid S_{(i-1)}\right)$ under Vecchia's approximation is

$$p\left(y_i \mid S_{(i-1)}\right) = \frac{1}{\sqrt{2\pi\sigma^2 \widetilde{\Sigma}^{\dagger}_{i,1}}} \exp\left(-\frac{\widetilde{e}^2_{i,1}}{2\sigma^2 \widetilde{\Sigma}^{\dagger}_{i,1}}\right).$$

A direct computation of (2.6) with any fixed ϕ_1 yields (2.8), where we have used the fact $E_1 \widetilde{e}^2_{i,1} = \sigma^2_1 \widetilde{\Sigma}^{\dagger}_{i,1}$ on the right hand side of (2.8). □

Our main result builds on the discussion in Section 10.5.3 of Zhang [2012] to establish the following theorem that explores the asymptotic distribution of $\hat{\sigma}^2_{n,vecch}$.

Theorem 1. *Assume that either of the following conditions holds:*

$$\sum_{i=1}^{n} \frac{E_0\left(\widetilde{e}_{i,1} - e_{i,0}\right)^2}{E_1 \widetilde{e}^2_{i,1}} = \mathcal{O}\left(1\right) \text{ and } \sum_{i=1}^{n} \left(\frac{E_0 e^2_{i,0}}{E_1 \widetilde{e}^2_{i,1}} - 1\right)^2 = \mathcal{O}\left(1\right).$$

(2.9)

or

$$\sum_{i=1}^{n} \frac{E_1(\tilde{e}_{i,1} - e_{i,0})^2}{E_0 e_{i,0}^2} = \mathcal{O}\left(1\right) \textit{ and } \sum_{i=1}^{n} \left(\frac{E_1 \tilde{e}_{i,1}^2}{E_0 e_{i,0}^2} - 1\right)^2 = \mathcal{O}\left(1\right).$$

(2.10)

Then

$$\sqrt{n}\left(\hat{\sigma}_{n,vecch}^2 \phi_1^{2\nu} - \sigma_0^2 \phi_0^{2\nu}\right) \xrightarrow{\mathscr{L}} N\left(0, 2\left(\sigma_0^2 \phi_0^{2\nu}\right)^2\right).$$

(2.11)

Before presenting the proof of Theorem 1, we state and prove the following lemma.

Lemma 2. *The assumptions in* (2.9) *imply that*

$$E_0\left[\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right] = o\left(\sqrt{n}\right),$$

(2.12)

Proof. We first prove that (2.9) implies (2.12). Note that

$$\begin{aligned} \left|E_0\left[\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right]\right| &= \left|\sum_{i=1}^{n} \left(\frac{E_0(\tilde{e}_{i,1} - e_{i,0} + e_{i,0})^2}{E_1 \tilde{e}_{i,1}^2} - 1\right)\right| \\ &= \left|\sum_{i=1}^{n} \frac{E_0(\tilde{e}_{i,1} - e_{i,0})^2}{E_1 \tilde{e}_{i,1}^2} + \sum_{i=1}^{n} \left(\frac{E_0 e_{i,0}^2}{E_1 \tilde{e}_{i,1}^2} - 1\right)\right| \\ &\le \sum_{i=1}^{n} \frac{E_0(\tilde{e}_{i,1} - e_{i,0})^2}{E_1 \tilde{e}_{i,1}^2} + \sum_{i=1}^{n} \left|\frac{E_0 e_{i,0}^2}{E_1 \tilde{e}_{i,1}^2} - 1\right|, \end{aligned}$$

where the second equality follows from the fact that $\tilde{e}_{i,1} - e_{i,0}$ and $e_{i,0}$ are independent under P_0. By the first condition in (2.9), we get $\sum_{i=1}^{n} E_0(\tilde{e}_{i,1} - e_{i,0})^2 / E_1 \tilde{e}_{i,1}^2 = \mathcal{O}\left(1\right) = o\left(\sqrt{n}\right)$. Fix $\varepsilon > 0$. By the second condition in (2.9), there is $M > 0$ such that $\sum_{i > M} \left(E_0 e_{i,0}^2 / E_1 \tilde{e}_{i,1}^2 - 1\right)^2 < \varepsilon$. So for $n > M$, we can use the Cauchy–Schwarz inequality to obtain

$$\begin{aligned} \sum_{i=1}^{n} \left|\frac{E_0 e_{i,0}^2}{E_1 \tilde{e}_{i,1}^2} - 1\right| &= \sum_{i=1}^{M} \left|\frac{E_0 e_{i,0}^2}{E_1 \tilde{e}_{i,1}^2} - 1\right| + \sum_{i=M+1}^{n} \left|\frac{E_0 e_{i,0}^2}{E_1 \tilde{e}_{i,1}^2} - 1\right| \\ &\le \sum_{i=1}^{M} \left|\frac{E_0 e_{i,0}^2}{E_1 \tilde{e}_{i,1}^2} - 1\right| + \sqrt{(n - M)\varepsilon}. \end{aligned}$$

Dividing both sides by $\sqrt{n}$ reveals that $\limsup_{n \to \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left| E_0 e_{i,0}^2 / E_1 \tilde{e}_{i,1}^2 - 1 \right| \le \sqrt{\varepsilon}$. Since $\varepsilon > 0$ is arbitrary, it follows that $\sum_{i=1}^{n} \left| E_0 e_{i,0}^2 / E_1 \tilde{e}_{i,1}^2 - 1 \right| = o\left(\sqrt{n}\right)$ and we obtain (2.12).

□

We now present a proof of Theorem 1.

Proof of Theorem 1. Recall that $\sum_{i=1}^{n} e_{i,0}^2 / \left(E_0 e_{i,0}^2\right) = Y_n^\top V_{n,0}^{-1} Y_n$, where $V_{n,0}$ is the covariance matrix of Y_n under P_0. Using (2.8) that was derived in Lemma 1, we obtain

$$\sqrt{n}\left(\hat{\sigma}_{n,vecch}^2 / \sigma_1^2 - 1\right) = \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right] + \sqrt{n}\left(\frac{1}{n} Y_n^\top V_{n,0}^{-1} Y_n - 1\right). \tag{2.13}$$

By the central limit theorem, we have $\sqrt{n}\left(\frac{1}{n} Y_n^\top V_{n,0}^{-1} Y_n - 1\right) \xrightarrow{\mathscr{L}} N\left(0, 2\right)$.

We next show that the condition (2.9) implies that

$$\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2} = o\left(\sqrt{n}\right). \tag{2.14}$$

To prove this, it will be sufficient to show that

$$\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2} - E_0\left[\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right] = O(1). \tag{2.15}$$

The result in (2.14) will then follow from Lemma 2. We turn to proving (2.15). Our argument relies on the equivalence of Gaussian sequences. Let $\widetilde{P}_{1,n}$ be the probability distribution corresponding to $\widetilde{p}\left(Y_n; \phi_1, \sigma_1^2\right)$, and let $\rho_n := \widetilde{p}\left(Y_n; \phi_1, \sigma_1^2\right) / p\left(Y_n; \phi_0, \sigma_0^2\right)$ be the Radon-Nikodym derivative of $\widetilde{P}_{1,n}$ with respect to P_0 on the realization Y_n for a given n. Write $\widetilde{P}_{1,\infty}$ (respectively, $P_{0,\infty}$) for the probability distribution corresponding to $\widetilde{p}\left(\cdot\,; \phi_1, \sigma_1^2\right)$ (respectively, $p\left(\cdot\,; \phi_0, \sigma_0^2\right)$) on the infinite sequence $(Y_1, Y_2, \ldots)$. By Kakutani’s dichotomy, $\widetilde{P}_{1,\infty}$ and $P_{0,\infty}$ are either equivalent or mutually singular to each other. If $\widetilde{P}_{1,\infty}$ is equivalent to $P_{0,\infty}$, then $\lim_{n \to \infty} \rho_n = \rho_\infty =: d\widetilde{P}_{1,\infty} / dP_{0,\infty}$ with P_0-probability 1 [see, e.g., Ibragimov and Rozanov, 1978, Section III.2.1]. Also, $\mathbb{P}_0(0 < \rho_\infty < \infty) = 1$ and $-\infty < E_0(\log \rho_\infty) < \infty$. As a consequence,

$$\log \rho_n = -\frac{1}{2} \log \frac{\det \widetilde{V}_{n,1}}{\det V_{n,0}} - \frac{1}{2}\left(\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1 \tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right) = \mathscr{O}(1),$$

$$E_0(\log\rho_n) = -\frac{1}{2}\log\frac{\det\widetilde{V}_{n,1}}{\det V_{n,0}} - \frac{1}{2}E_0\left[\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0e_{i,0}^2}\right] = \mathcal{O}(1),$$

where $\widetilde{V}_{n,1}$ (respectively, $V_{n,0}$) is the covariance matrix of Y_n under $\widetilde{P}_{1,\infty}$ (respectively, P_0). By taking the difference of the above two equations, we get (2.15) under the condition that $\widetilde{P}_{1,\infty}$ is equivalent to $P_{0,\infty}$. Using Theorem 5, Section VII.6 of Shiryaev [1996], we can conclude that

$$\begin{aligned}\widetilde{P}_{1,\infty}\,\text{is equivalent to}\,P_{0,\infty} &\Leftrightarrow \sum_{i=1}^{\infty}\left[\frac{E_0(\tilde{e}_{i,1}-e_{i,0})^2}{E_1\tilde{e}_{i,1}^2} + \left(\frac{E_0e_{i,0}^2}{E_1\tilde{e}_{i,1}^2} - 1\right)^2\right] < \infty\\ &\Leftrightarrow \sum_{i=1}^{\infty}\left[\frac{E_1(\tilde{e}_{i,1}-e_{i,0})^2}{E_0e_{i,0}^2} + \left(\frac{E_1\tilde{e}_{i,1}^2}{E_0e_{i,0}^2} - 1\right)^2\right] < \infty\,.\end{aligned} \tag{2.16}$$

Since first equivalence in (2.16) is simply a reformulation of (2.9), we have established that the condition (2.9) implies (2.15) and, hence, the result in (2.14).

The proof from condition (2.10) to (2.14) is established following the proof from condition (2.9) to (2.14). We now break the quantity $\sum_{i=1}^{n}\tilde{e}_{i,1}^2/E_1\tilde{e}_{i,1}^2 - \sum_{i=1}^{n}e_{i,0}^2/E_0e_{i,0}^2$ in (2.14) into

$$\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0e_{i,0}^2} - E_1\left[\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0e_{i,0}^2}\right]$$
$$+E_1\left[\sum_{i=1}^{n}\frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^{n}\frac{e_{i,0}^2}{E_0e_{i,0}^2}\right],$$

and replace the left hand-side of (2.15) and (2.12) by the two quantities respectively. Then the equivalence of P_1 and P_0 along with lemma 2 shows that the replaced (2.12) holds. The proof that (2.10) implies the replaced (2.15) remains the same except that we now use the second equivalence in (2.16). Thus we complete the proof of Theorem 1. □

Turning to the connection between Theorem 1 and predictive consistency of Vecchia's approximation in the sense of Kaufman and Shaby [2013, p.478], note that $e_{i,0}$ and $\tilde{e}_{i,1}$ are the predictive errors for y_i under the full model with correct parameters and under (1.4) with possibly incorrectly specified parameters $(\phi,\sigma) = (\phi_1,\sigma_1)$. A consequence of (2.9) or (2.10) is that $E_1\tilde{e}_{i,1}^2/E_0e_{i,0}^2 \to 1$ as i (and hence n) is large. Hence, (2.9) or (2.10) implies asymptotic normality of estimates as well as predictive consistency.

3. Infill Asymptotics for Vecchia's Approximation on the line

It is possible to obtain further insights into Theorem 1 when considering the asymptotic normality of $\hat{\sigma}_{n,vecch}$ for Matérn models with observations on the real line. Whilst the conditions (2.9) and (2.10) are, in general, analytically intractable due to the presence of $E_0\tilde{e}_{i,1}^2$, we will show that (2.10) holds for Matérn models on $\mathbb{R}$.

To simplify the presentation, we consider the fixed domain $\mathscr{D} = [0, 1]$, and the sampled locations $\chi = \{i/n : 0 \le i \le n\}$. Denote $\delta = 1/n$ for the spacing of χ, and $y_i = y(i\delta), 0 \le i \le n$ for the observations. We define $S_{(i)} = S_{(i)}[k] := (y_i, y_{i-1}, \ldots, y_{i-k+1})$ for a positive integer k, where $S_{(i)}[k]$ is the vector of k consecutive observations backward from y_i. The integer k is capped by i since $S_{(i)}[k]$ is a subvector of $y_{(i)}$.

Assumption 1. Let $\mathscr{D} = [0, 1]$, and $\chi = \{i\delta : 0 \le i \le n\}$ with $\delta = 1/n$. Then

$$\sum_{i=1}^{n} \frac{E_1(e_{i,1} - e_{i,0})^2}{E_1 e_{i,1}^2} = \mathcal{O}\left(1\right). \tag{3.17}$$

Before stating the main result on $\mathbb{R}$, we demonstrate why this assumption is reasonable. We empirically investigate $\sum_{i=1}^{n} E_1(e_{i,1} - e_{i,0})^2 / E_1 e_{i,1}^2$ for increasing values of n. Figure 1 plots the values of $\sum_{i=1}^{n} E_1(e_{i,1} - e_{i,0})^2 / E_1 e_{i,1}^2$ with $\chi = \{i\delta : 0 \le i \le n\}$ for $\nu = 0.25, 0.5, 1.0, 1.5, 2.0$, and n ranging from 100 to 1200. As n increases, the plot tends to flatten as is suggested by the assumption.

Some additional explanation is also possible from a theoretical view-point. Since P_0 and P_1 are equivalent, Corollary 3.1 of Stein [1990a] implies that $E_1(e_{i,1} - e_{i,0})^2 / E_1 e_{i,1}^2 \to 0$ as $n, i \to \infty$. By stationarity and symmetry of the Matérn model, $e_{i,j}$ is distributed as the error of the least square estimate of $y_0 := y(0)$ given observations $y_{(i)} = (y_1, \ldots, y_i)^\top$. Hence, $e_{i,j}$ can be realized as $y_0 - E_j(y_0 \mid y_{(i)})$. Similarly, $\tilde{e}_{i,j}$ can be realized as $y_0 - E_j(y_0 \mid S_{(i-1)})$. Now consider the infinitely sampled locations $\{i\delta : i \ge 0\}$, and extend the finite sample $Y_n := (y_0, \ldots, y_n)^\top$ to $Y := (y_0, y_1, \ldots)^\top$ with $y_i := y(i\delta)$. The sampled locations of Y form an infinite grid on $[0, \infty)$, and $\delta = 1/n$ is determined based on the sample size of Y_n. Let $e_{\infty,j}$ be the error $y_0 - E_j(y_0 \mid y_1, \ldots)$ for the infinite sequence Y. For f_0 (resp. f_1) the spectral density under P_0(resp . P_1), it is easily seen that $f_0, f_1 \sim C\sigma_0^2\phi_0^\nu u^{-2\nu-1}$ as $u \to \infty$, and $(f_1 - f_0)/f_0 \asymp u^{-2}$. Therefore, by Theorem 2 of Stein [1999b],

$$\frac{E_1(e_{\infty,1} - e_{\infty,0})^2}{E_1 e_{\infty,1}^2} = O\left(\delta^{\min(2\nu+1,4)} \log\left(\delta^{-1}\right)^{1(\nu = 3/2)}\right). \tag{3.18}$$

Intuitively, $e_{i,j} \approx e_{\infty,j}$ for large i. It will not be unreasonable to speculate a stronger result where (3.18) still holds by replacing $e_{\infty,j}$ with $e_{i,j}$ for large i, which would imply (3.17). As indicated on p.138 of Stein [1999a], obtaining the rate of $E_1(e_{i,1} - e_{i,0})^2 / E_1 e_{i,1}^2$ for any bounded domain $\mathscr{D}$ is a highly non-trivial task. The only known results are obtained in Stein [1990b, 1999b] for $\nu = \frac{1}{2}, \frac{3}{2}, \ldots$, which also imply (3.17). In fact, we need that $E_1 e_{i,1}^2 = E_1 e_{\infty,1}^2 \left(1 + \mathcal{O}\left(i^{-\kappa}\right)\right)$ for any $\kappa > 0$. Hence, the only missing piece in the above heuristics is an estimate of $E_1 e_{i,0}^2 / E_1 e_{\infty,0}^2$, which we do not explore further here. Our result is stated as follows.

Theorem 2. *Let* $\mathscr{D} = [0, 1]$, $\chi = \{i\delta : 0 \le i \le n\}$ *and* $S_{(i)} = S_{(i)}[n^{\epsilon}]$ *for* $\epsilon \in (0, 1)$. *If* (3.17) *holds, then* (2.10) *also holds. Consequently,* (2.11) *holds for the Matérn model* $(\nu > 0)$.

We make a few remarks before presenting a proof. Theorem 2 states that the asymptotic normality of the microergodic parameter $\sigma^2 \phi^{2\nu}$ still holds under Vecchia's approximation in a neighborhood of at most size $k = n^{\epsilon} \ll n$ (sample size), where the computation of $\hat{\sigma}_{n,vecch}^2$ is much cheaper. This justifies the validity of Vecchia's approximation for Matérn models from a fixed-domain perspective. The range n^{ϵ} may not be optimal, and it might be possible to improve to $k = \mathcal{O}(n)$. However, we do not pursue this direction here from a theoretical standpoint. A simulation study is provided for the case of $k = \mathcal{O}(n)$ in Section 4.

An interesting situation arises with $\nu = 1/2$, in which the process reduces to the Ornstein-Uhlenbeck process, and $p(y_i \mid y_{(i-1)}) = p(y_i \mid y_{i-1})$. Therefore, (2.11) trivially holds for Vechhia's approximation with a neighbor of size $k = 1 = \mathcal{O}(1)$. It is, therefore, natural to enquire whether the asymptotic normality of (2.11) holds under Vecchia's approximation within a range $k = \mathcal{O}(1)$. If this is true, computational efforts can be reduced further. Unfortunately, this need not be the case. For $\nu < 1/4$ and $k = \mathcal{O}(1)$, $n^{2\nu}\left(\hat{\sigma}_{n,vecch}^2 \phi_1^{2\nu} - \sigma_0^2 \phi_0^{2\nu}\right)$ converges to a non-Gaussian distribution [Bachoc and Lagnoux, 2020]. The cases for $\nu \ge 1/4$ remain unresolved.

Now we turn to the proof of Theorem 2. The key to this analysis is the following proposition, which relies on a result on the bound of $e_{\infty,j} - e_{i,j}$, i.e. the difference between the errors of the finite and the infinite least square estimates [Baxter, 1962]. The study dates back to the work of Kolmogorov [1941], see also Grenander and Szegö [1958], Ibragimov [1964], Dym and McKean [1970, 1976], Ginovian [1999] for related discussions.

Proposition 1. *Let* $\kappa > 0$. *There exist* $C_0, C_\kappa > 0$ *such that for* $\delta < 1$ *and* $j = 0, 1$,

$$E_j e_{\infty,j}^2 \sim C_0 \delta^{2\nu}, \tag{3.19}$$

$$E_j \left(e_{\infty,j} - e_{i,j}\right)^2 \le C_\kappa \delta^{2\nu} i^{-\kappa}. \tag{3.20}$$

Proof. From the discussion below (3.17) that $e_{i,j}$ and $e_{\infty,j}$ can be realized as $e_{i,j} = y_0 - E_j(y_0 \mid y_1, \ldots, y_i)$ and $e_{\infty,j} = y_0 - E_j(y_0 \mid y_1, y_2, \ldots)$, where $y_i := y(i\delta)$ is indexed by nonnegative integers, we know from Stein [1999a, p.77] that the spectral density of Y under P_j is

$$\bar{f}_j^{\delta}(u) = \frac{1}{\delta}\sum_{\ell = -\infty}^{\infty} f_j\left(\frac{u + 2\pi\ell}{\delta}\right) \text{for } u \in (-\pi, \pi],\ j = 0, 1,$$

where f_j is the spectral density defined by (1.2) corresponding to P_j. For $j = 0, 1$, $f_j(u) \sim C\sigma_0^2\phi_0^{2\nu}u^{-2\nu - 1}$ as $u \to \infty$. From Stein [1999a, p.80, (17)], we obtain

$$E_j e_{\infty, j}^2 \sim 2\pi C\sigma_0^2\phi_0^{2\nu}\delta^{2\nu}\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\Sigma_{\ell = -\infty}^{\infty}|u + 2\pi\ell|^{-2\nu - 1}\right)du\right),$$

which implies (3.19) with

$$C_0 = 2\pi C\sigma_0^2\phi_0^{2\nu}\exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi}\log\left(\Sigma_{\ell = -\infty}^{\infty}|u + 2\pi\ell|^{-2\nu - 1}\right)du\right).$$

Turning to (3.20), we know from Baxter [1962, p.142, (15)] that

$$E_j(e_{i, j} - e_{\infty, j})^2 = E_j e_{\infty, j}^2 E_j e_{i, j}^2 \sum_{m = i}^{\infty}|\phi_{m, j}(0)|^2, \tag{3.21}$$

where $\phi_{m, j}(\cdot)$ are the Szegö polynomials associated with the spectral $\bar{f}_j^{\delta}$ [see Section 2.1 of Grenander and Szegö, 1958, for background]. Note that $E_j e_{\infty, j}^2 \sim C_0\delta^{2\nu}$ and $E_j e_{i, j}^2 \le E_j e_{1, j}^2 \to 0$ as $\delta \to 0$. It will be sufficient to establish

$$\sum_{m = 0}^{\infty} m^{\kappa}|\phi_{m, j}(0)| \le D_\kappa \text{ for some } D_\kappa > 0, \tag{3.22}$$

in which case the identity (3.21) will imply (3.20). The key observation of Baxter [1962] (Theorem 2.3) is that (3.22) holds if the κ^{th} moment of the Fourier coefficients associated with $\bar{f}_j^{\delta}$ is bounded from above by D'_κ for some $D'_\kappa > 0$, i.e.

$$\sum_{m = 0}^{\infty} m^{\kappa}|c_{m, j}| < D'_\kappa, \text{ where } c_{m, j} := \frac{1}{2\pi}\int_{-\pi}^{\pi}\bar{f}_j^{\delta}(u)e^{-inu}du.$$

A sufficient condition for the latter to hold is that the κ^{th} derivative of $\bar{f}_j^{\delta}$ is integrable, and $\int_{-\pi}^{\pi}\left|\frac{d^\kappa}{du^\kappa}\bar{f}_j^{\delta}(u)\right|du \le D''_\kappa$, for some $D''_\kappa > 0$ which does not depend on $\delta < 1$. Breaking the sum of $\bar{f}_j^{\delta}$ according to $\ell = 0$ and $\ell \neq 0$ produces

$$\int_{-\pi}^{\pi}\left|\frac{d^{K}}{du^{K}}\bar{f}_{j}^{\delta}(u)\right|du \le A\int_{-\infty}^{\infty}\left(1+u^{2\nu+1+\kappa}\right)^{-1}du + A'\delta^{2\nu}\sum_{\ell\ne 0} l^{-2\nu-1-\kappa}, \tag{3.23}$$

where A, $A' > 0$ are numerical constants. Hence, the right hand side of (3.23) is bounded by $D_{\kappa}'' = A\int_{-\infty}^{\infty}\left(1+u^{2\nu+1+\kappa}\right)^{-1}du + A'\sum_{\ell\ne 0} l^{-2\nu-1-\kappa}$, which depends only on κ. □

Proof of Theorem 2. We fix $\kappa = 2/\epsilon$, and use $C_{\kappa}^{(1)}, C_{\kappa}^{(2)}, \ldots$ to denote constants depending only on κ. Note that $E_0e_{i,0}^2 = E_0e_{\infty,0}^2 + E_0(e_{\infty,0} - e_{i,0})^2$, since $e_{\infty,0}$ and $e_{\infty,0} - e_{i,0}$ are independent under P_0. By Proposition 1, we get $E_0e_{i,0}^2 \sim C_0\delta^{2\nu}\left(1 + C_{\kappa}^{(1)}i^{-K}\right)$. Similarly, $E_1\tilde{e}_{i,1}^2 = E_1e_{\infty,1}^2 + E_1(\tilde{e}_{i,1} - e_{\infty,1})^2 \sim C_0\delta^{2\nu}\left(1 + C_{\kappa}^{(2)}\min(i, n^{\epsilon})^{-K}\right)$, because $\tilde{e}_{i,1}$ is realized as $y_0 - E_1(y_0 \mid S_{(i-1)}[k])$ with $k = \min(i, n^{\epsilon})$. Therefore,

$$\sum_{i=1}^{n}\left(\frac{E_1\tilde{e}_{i,1}^2}{E_0e_{i,0}^2} - 1\right)^2 \le \sum_{i \ge n^{\epsilon}}\left(\frac{C_{\kappa}^{(3)}n^{-\epsilon K}}{1 + C_{\kappa}^{(1)}i^{-K}}\right)^2 + \sum_{i=1}^{n}\left(\frac{C_{\kappa}^{(4)}i^{-K}}{1 + C_{\kappa}^{(1)}i^{-K}}\right)^2$$
$$\le \frac{C_{\kappa}^{(5)}}{n^3} + C_{\kappa}^{(6)}\sum_{i=1}^{n} i^{-2K} \le C_{\kappa}^{(7)}.$$

Moreover, we have

$$\frac{E_1(\tilde{e}_{i,1} - e_{i,0})^2}{E_0e_{i,0}^2} \le 3\frac{E_1(\tilde{e}_{i,1} - e_{\infty,1})^2}{E_0e_{i,0}^2} + 3\frac{E_1(e_{i,1} - e_{\infty,1})^2}{E_0e_{i,0}^2} + 3\frac{E_1(e_{i,1} - e_{i,0})^2}{E_1e_{i,1}^2}\frac{E_1e_{i,1}^2}{E_0e_{i,0}^2}.$$

By the same argument as above, for the first two terms:

$$\sum_{i=1}^{n}\frac{E_1(\tilde{e}_{i,1} - e_{\infty,1})^2}{E_0e_{i,0}^2} < C_{\kappa}^{(8)}\text{and} \sum_{i=1}^{n}\frac{E_1(e_{i,1} - e_{\infty,1})^2}{E_0e_{i,0}^2} < C_{\kappa}^{(9)}.$$

For the last term,

$$\sum_{i=1}^{n}\frac{E_1(e_{i,1} - e_{i,0})^2}{E_1e_{i,1}^2}\frac{E_1e_{i,1}^2}{E_0e_{i,0}^2} \le C_{\kappa}^{(10)}\sum_{i=1}^{n}\frac{E_1(e_{i,1} - e_{i,0})^2}{E_1e_{i,1}^2},$$

which converges because of (3.17). This establishes (2.10) and, hence, (2.11) follows. □

4. Simulations

Based on (2.12) and (2.15) provided in Theorem 1, Theorem 2 has proved that

$$c_n(\phi_1, \phi_0, k) = \frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n} \frac{\tilde{e}_{i,1}^2}{E_1\tilde{e}_{i,1}^2} - \sum_{i=1}^{n} \frac{e_{i,0}^2}{E_0 e_{i,0}^2}\right] = o(1) \tag{4.24}$$

when $k = n^{\epsilon}$ for $\epsilon \in (0, 1)$. The equation (4.24) induces the critical condition (2.14), resulting in the convergence in law in (2.11). Looking into the more challenging case $k = \mathcal{O}(\log(n))$, we extend the discussion in Theorem 2 via investigating the behaviour of $c_n(\phi_1, \phi_0, k)$ in (4.24) for a sequence of datasets with increasing sample sizes. Our experiments involve two data generation schemes. The first scheme considers the study domains $\mathcal{D}_1 = [0, 1]$ with n observations on the grid $\chi_1 = \{i/(n-1) : 0 \le i \le n-1\}$ and $\mathcal{D}_2 = [0, 1]^2$ with $n = n_s^2$ observations on the grid $\chi_2 = \{(i/(n_s - 1), j/(n_s - 1)) : 0 \le i, \le n_s - 1, 0 \le j \le n_s - 1\}$. With this scheme, we generate a sequence of datasets with increasing sample size on increasingly finer grids on the study domains. The second scheme generates data on a “disturbed grid”. On $\mathcal{D}_1 = [0, 1]$ with n observations, χ_1 comprises locations randomly sampled by $\mathrm{N}(i/(n+2), 0.15/(n+2))$ for $i = 1, \ldots, n$. On $\mathcal{D}_2 = [0, 1]^2$ with $n = n_s^2$ observations, locations in χ_2 are generated by $\{N(i/(n_s + 2), 0.15/(n_s + 2)), N(i/(n_s + 2), 0.15/(n_s + 2))\}$ for $i, j = 1, \ldots, n_s$. With this scheme, we first generate simulations with the largest sample size, and then randomly select successively larger subsets from the same dataset to examine the tendency of $c_n(\phi_1, \phi_0, k)$ with an increasing n. The first scheme matches the setup of our proofs in the preceding sections, and the second scheme serves as a more directly informative regime for simulation studies about asymptotics. In practice, estimation using Vecchia’s approximation (1.4) is complicated by the fixed ordering of locations. Guinness [2018] has provided excellent practical insights into this issue that can considerably improve finite sample behaviour in certain settings. In this study, we test two different orderings of locations, maximin ordering and sorted coordinate ordering. The sorted coordinate ordering orders locations on χ_2 first based on the second coordinate and then break ties based on the associated first coordinate. We take $S_{(i)}[k]$ as the at most k nearest neighbors of y_{i+1}. In both studies on $\mathcal{D}_1$ and $\mathcal{D}_2$, we fix $\sigma^2 = 1.0$ and consider 5 different smoothness values $\nu \in \{0.25, 0.5, 1.0, 1.5, 2\}$. We choose different decay parameters ϕ_0 for different ν so that $K_\theta(h) = 0.05$ when $h = 0.2$ and 0.5 for the study on $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively.

For each fixed value of $\theta = \{\sigma^2, \phi, \nu\}$, we generate 100 datasets with Y_n being the realization from $y(s) \sim GP(0, K_\theta(\cdot))$ and calculate $c_n(\phi_1, \phi_0, k)$ with k being the closest integer to $3\log(n)$, and $\phi_1 = 1.2\phi_0$ and $1.1\phi_0$ for $\mathcal{D}_1$ and $\mathcal{D}_2$, respectively. Then, we record the mean and standard deviation of the 100 values of $c_n(\phi_1, \phi_0, k)$. We repeat this process for different values of n ranging from $2^6 = 64$ to $2^{12} = 4096$ in the study on $\mathcal{D}_1$. The study on $\mathcal{D}_2$ follows the study on $\mathcal{D}_1$ with n_s ranging from 9 to 81. The code for this simulation study is available on https://github.com/LuZhangstat/vecchia_consistency. Figure 2a & 2b summarize the study results on $\mathcal{D}_1$ under the two data generation schemes. Each figure presents 10 different graphs, one for each value of ν and each ordering, showing the mean and standard deviation of $c_n(\phi_1, \phi_0, k)$ for different values of n.

The value of $c_n(\phi_1, \phi_0, k)$, as seen in Figure 2, decreases rapidly as the sample size increases, supporting the main conclusion in Theorem 2. We do not observe a strong impact of ordering and data generation scheme on the results. The corresponding graphs for the study on $\mathscr{D}_2$ are presented in Figure 3. These graphs also reveal decreasing trends, but with more gentle slopes as compared to Figure 2. The results under the second data generation scheme are slightly better than those under the first scheme. When the smoothness ν is small, the standard deviation decreases faster with maximin ordering than with sorted coordinate ordering. Meanwhile, the standard deviation doesn't decrease significantly with the increase of n when ν is large. To explore further, we reproduce the study on $\mathscr{D}_2$ with k being the closet integer to $\sqrt{n}$, and we illustrate the results in Figure 4. We observe that the standard deviation decreases rapidly as n increases in all cases.

We have also seen, from the proof in Theorem 1, that $c_n(k, \phi_1, \phi_0) = \sqrt{n}\left(\hat{\sigma}^2_{n,vecch}/\sigma_1^2 - \hat{\sigma}^2_{0,n}/\sigma_0^2\right)$ where $\hat{\sigma}^2_{0,n} = \text{argmax}_{\sigma^2}\left\{p\left(y; \phi_0, \sigma^2\right), \sigma^2 \in \mathbb{R}^+\right\}$ is the maximum likelihood estimator from (1.3) when fixing $\phi_1 = \phi_0$. Hence, $c_n(k, \phi_1, \phi_0)$ also measures the discrepancy between $\hat{\sigma}^2_{n,vecch}/\sigma_1^2$ and $\hat{\sigma}^2_{0,n}/\sigma_0^2$, and the decreasing trend of $c_n(\phi_0, \phi_1, k)$ indicates that the inference based on Vecchia's approximation approaches the inference based on the full likelihood as sample size increases. This phenomenon reveals that Vecchia's approximation is still efficient when the neighbor size k is substantially smaller than the sample size.

5. Conclusions and Future work

We have developed insights into inference based on GP likelihood approximations by Vecchia [1988] under fixed domain asymptotics for geostatistical data analysis. We have formally established the sufficient conditions for such approximations to have the same asymptotic efficiency as a full GP likelihood in estimating parameters in Matérn covariance function. The insights obtained here will enhance our understanding of identifiability of process parameters and can also be useful for developing priors for the microergodic parameters in Bayesian modeling. The results derived here will also offer insights into formally establishing posterior consistency of process parameters for a number of Bayesian models that have emerged from (1.4) [Datta et al., 2016a,b, Katzfuss and Guinness, 2021, Peruzzi et al., 2022].

We anticipate the current manuscript to generate further research in variants of geostatistical models. For example, it is conceivable that these results will lead to asymptotic investigations of covariance-tapered models [see, e.g., Wang et al., 2011] and in adapting some results, such as Theorems 2 and 3 in Kaufman and Shaby [2013] where ϕ is estimated, to the approximate likelihoods presented here. Another direction of research can lead to formal developments regarding the inferential consistency of the "nugget" or the variance of measurement error when the spatial process has a discontinuity at 0 arising white noise [Tang et al., 2021]. Finally, there is scope to specifically investigate fixed domain inference for other likelihood approximations that extend or generalize (1.4) [see, e.g., Katzfuss and Guinness, 2021, Peruzzi et al., 2022].

Acknowledgements

The work of the first and third author was supported, in part, by funding from the National Science Foundation grants NSF/DMS 1916349 and NSF/IIS 1562303, and by the National Institute of Environmental Health Sciences (NIEHS) under grants R01ES030210 and 5R01ES027027.

References

Abramowitz M and Stegun A. Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables. Dover, 1965.

Bachoc F and Lagnoux A. Fixed-domain asymptotic properties of maximum composite likelihood estimators for Gaussian processes. J. Statist. Plann. Inference, 209:62–75, 2020.

Banerjee Sudipto. High-dimensional bayesian geostatistics. Bayesian Analysis, 12:583–614, 2017. [PubMed: 29391920]

Baxter Glen. An asymptotic result for the finite predictor. Math. Scand, 10:137–144, 1962.

Datta A, Banerjee S, Finley AO, and Gelfand AE. Hierarchical Nearest-Neighbor Gaussian Process Models for Large Geostatistical Datasets. Journal of the American Statistical Association, 111:800–812, 2016a. URL 10.1080/01621459.2015.1044091. [PubMed: 29720777]

Datta A, Banerjee S, Finley AO, Hamm NAS, and Schaap M. Non-separable Dynamic Nearest-Neighbor Gaussian Process Models for Large spatio-temporal Data With an Application to Particulate Matter Analysis. Annals of Applied Statistics, 10:1286–1316, 2016b. URL 10.1214/16-A0AS931. [PubMed: 29657659]

Du J, Zhang H, and Mandrekar VS. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. The Annals of Statistics, 37 (6A):3330–3361, 2009.

Dym H and McKean HP. Extrapolation and interpolation of stationary Gaussian processes. Ann. Math. Statist, 41:1817–1844, 1970.

Dym H and McKean HP. Gaussian processes, function theory, and the inverse spectral problem. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1976. Probability and Mathematical Statistics, Vol. 31.

Eidsvik Jo, Shaby Benjamin A., Reich Brian J., Wheeler Matthew, and Niemi Jarad. Estimation and prediction in spatial models with block composite likelihoods. Journal of Computational and Graphical Statistics, 23(2):295–315, 2014. doi: 10.1080/10618600.2012.760460. URL 10.1080/10618600.2012.760460.

Finley Andrew O, Datta Abhirup, Cook Bruce C, Morton Douglas C, Andersen Hans E, and Banerjee Sudipto. Efficient algorithms for Bayesian Nearest Neighbor Gaussian Processes. Journal of Computational and Graphical Statistics, 28(2):401–414, 2019. [PubMed: 31543693]

Ginovian MS. Asymptotic behavior of the prediction error for stationary random sequences. Izv. Nats. Akad. Nauk Armenii Mat, 34(1):18–36 (2000), 1999.

Grenander Ulf and Szegö Gabor. Toeplitz forms and their applications. California Monographs in Mathematical Sciences. University of California Press, Berkeley-Los Angeles, 1958.

Guinness Joseph. Permutation and grouping methods for sharpening Gaussian process approximations. Technometrics, 60(4):415–429, 2018. [PubMed: 31447491]

Heaton Matthew J, Datta Abhirup, Finley Andrew O, Furrer Reinhard, Guinness Joseph, Guhaniyogi Rajarshi, Gerber Florian, Gramacy Robert B, Hammerling Dorit, Katzfuss Matthias, et al. A case study competition among methods for analyzing large spatial data. Journal of Agricultural, Biological and Environmental Statistics, 24(3):398–425, 2019. [PubMed: 31496633]

Ibragimov IA. On the asymptotic behavior of the prediction error. Theory of Probability & Its Applications, 9(4):627–634, 1964.

Ibragimov Ildar Abdullovich and Rozanov YA. Gaussian random processes, volume 9 of Applications of Mathematics. Springer-Verlag, New York-Berlin, 1978. Translated from the Russian by A. B. Aries.

Katzfuss Matthias and Guinness Joseph. A general framework for vecchia approximations of gaussian processes. Statist. Sci, 36(1):124–141, 02 2021. doi: 10.1214/19-STS755. URL 10.1214/19-STS755.

Katzfuss Matthias, Guinness Joseph, Gong Wenlong, and Zilber Daniel. Vecchia approximations of Gaussian-process predictions. Journal of Agricultural, Biological and Environmental Statistics, 25(3):383–414, 2020.

Kaufman CG and Shaby BA. The role of the range parameter for estimation and prediction in geostatistics. Biometrika, 100(2):473–484, 2013.

Kolmogorov A. Interpolation und Extrapolation von stationären zufälligen Folgen. Bull. Acad. Sci. URSS Sér. Math [Izvestia Akad. Nauk. SSSR], 5:3–14, 1941.

Matérn B. Spatial Variation. Springer-Verlag, 1986.

Peruzzi Michele, Banerjee Sudipto, and Finley Andrew O.. Highly scalable bayesian geostatistical modeling via meshed gaussian processes on partitioned domains. Journal of the American Statistical Association, 117(538):969–982, 2022. doi: 10.1080/01621459.2020.1833889. URL 10.1080/01621459.2020.1833889. [PubMed: 35935897]

Shiryaev AN. Probability, volume 95 of Graduate Texts in Mathematics. Springer-Verlag, New York, second edition, 1996.

Stein ML. Interpolation of Spatial Data: Some Theory for Kriging. Springer-Verlag, 1999a.

Stein Michael L.. Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. Ann. Statist, 18 (2):850–872, 1990a.

Stein Michael L.. Bounds on the efficiency of linear predictions using an incorrect covariance function. Ann. Statist, 18(3):1116–1138, 1990b.

Stein Michael L.. Predicting random fields with increasing dense observations. Ann. Appl. Probab, 9(1):242–273, 1999b.

Stein Michael L, Chi Zhiyi, and Welty Leah J. Approximating likelihoods for large spatial data sets. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66(2):275–296, 2004.

Sun Ying, Li Bo, and Genton Marc G. Geostatistics for large datasets. In Advances and challenges in space-time modelling of natural events, pages 55–77. Springer, 2012.

Tang Wenpin, Zhang Lu, and Banerjee Sudipto. On identifiability and consistency of the nugget in Gaussian spatial process models. J. R. Stat. Soc. Ser. B. Stat. Methodol, 83(5):1044–1070, 2021.

Vecchia AV. Estimation and model identification for continuous spatial processes. Journal of the Royal Statistical society, Series B, 50:297–312, 1988.

Wang Daqing, Loh Wei-Liem, et al. On fixed-domain asymptotics and covariance tapering in gaussian random field models. Electronic Journal of Statistics, 5:238–269, 2011.

Zhang Hao. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. Journal of the American Statistical Association, 99(465):250–261, 2004.

Zhang Hao. Asymptotics and computation for spatial statistics. In Advances and Challenges in Space-time Modelling of Natural Events, pages 239–252. Springer, 2012.

Zhang Hao and Zimmerman Dale L. Towards reconciling two asymptotic frameworks in spatial statistics. Biometrika, 92(4):921–936, 2005.

Zhang Lu, Datta Abhirup, and Banerjee Sudipto. Practical bayesian modeling and inference for massive spatial datasets on modest computing environments. Statistical Analysis and Data Mining: The ASA Data Science Journal, 12(3):197–209, 2019. doi: 10.1002/sam.11413. URL 10.1002/sam.11413.

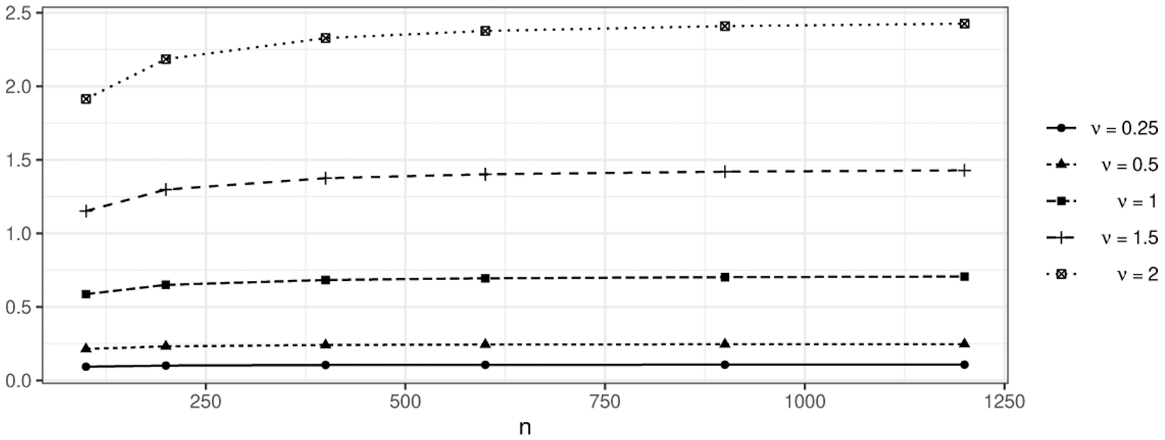
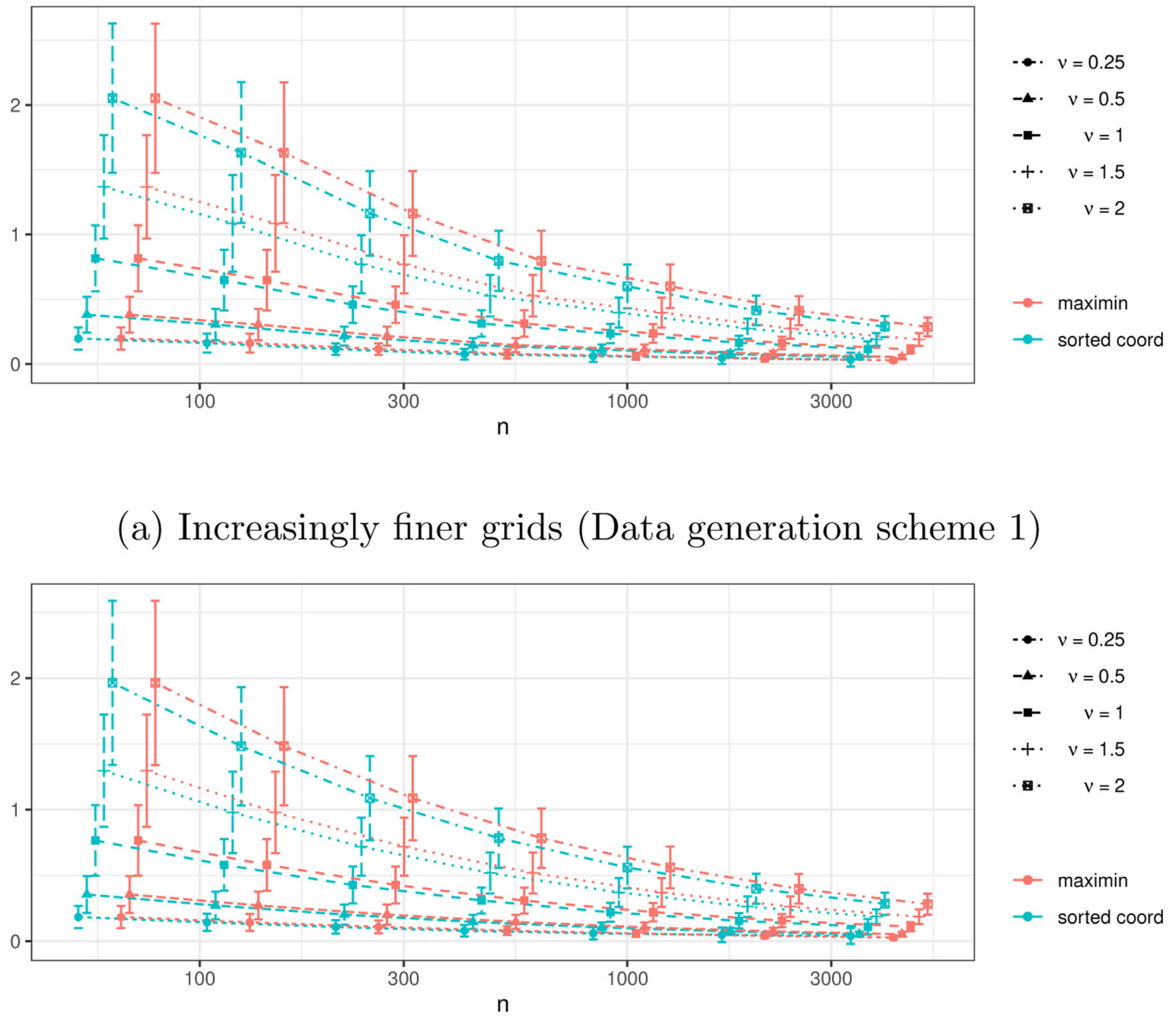


Figure 1:

Trend of $\sum_{i=1}^{n} E_1(e_{i,1} - e_{i,0})^2 / E_1 e_{i,1}^2$ for Matérn model when χ is a regular grid $\chi = \{i\delta : 0 \leq i \leq n\}$. Parameter σ^2 in Matérn covariogram equals 1.0 and decay ϕ for different ν are set to make the correlation of two points equals 0.05 when their distance reaches 0.2.

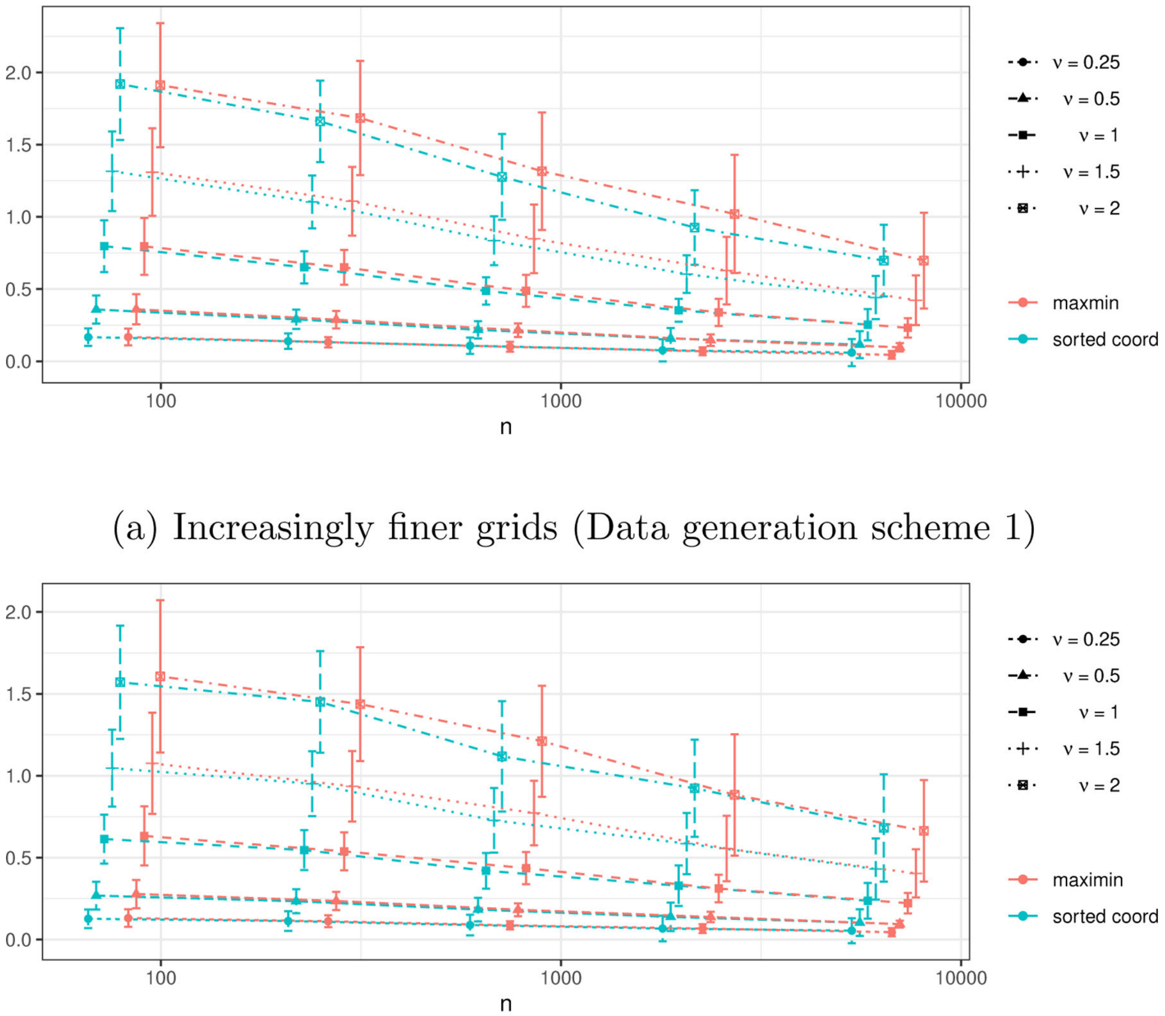


(a) Increasingly finer grids (Data generation scheme 1)

(b) Successively increasing datasets (Data generation scheme 2)

Figure 2:

The mean of $c_n(\phi_1, \phi_0, k)$ of 100 simulations on $\mathscr{D}_1 = [0, 1]$. The error bars represent one standard deviation. The sample size n take on values in 64, 128, 256, 512, 1024, 2048 and 4096. The graphs in red and blue show the results using maximin ordering and sorted coordinate ordering, respectively.

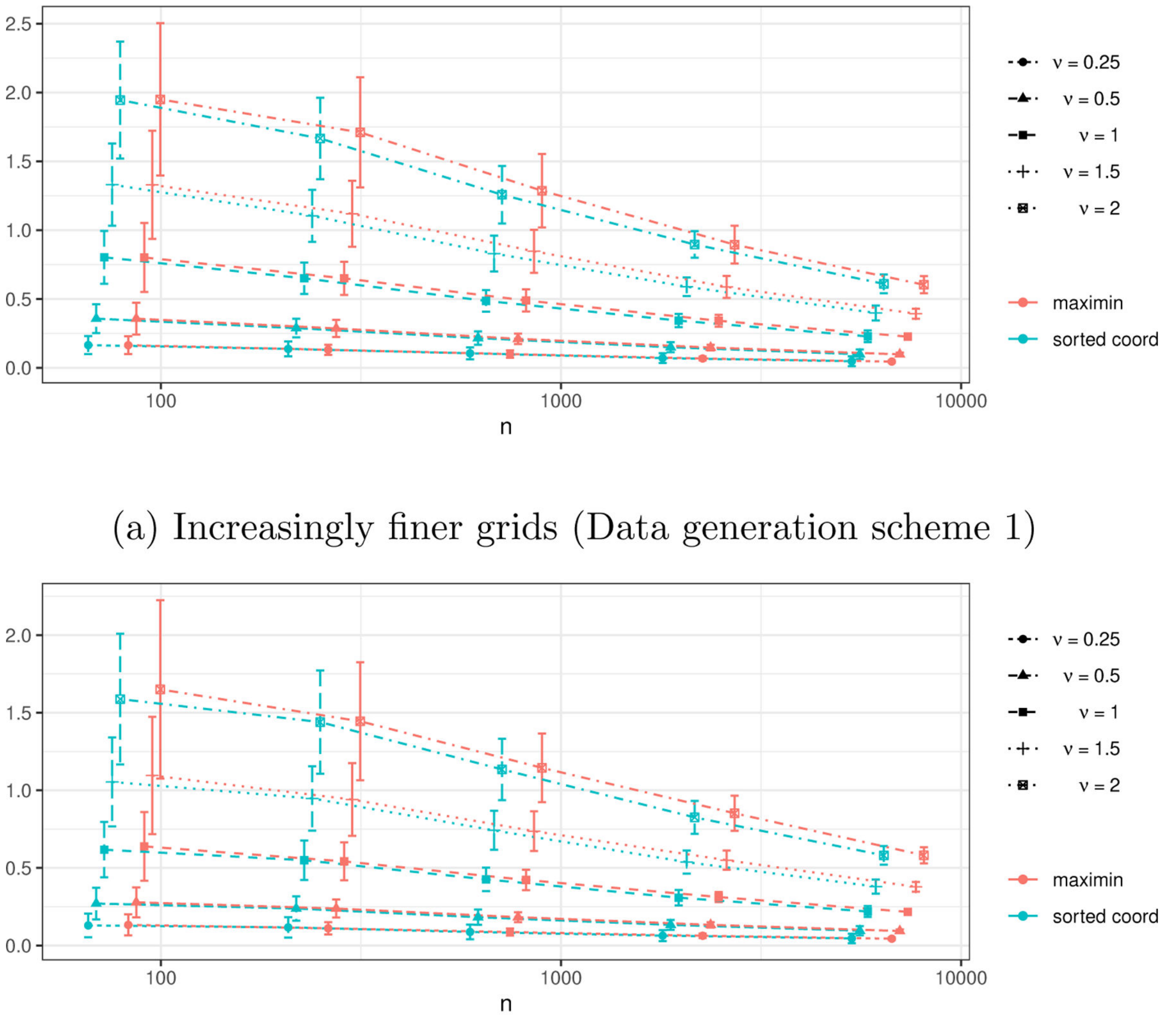


(a) Increasingly finer grids (Data generation scheme 1)

(b) Successively increasing datasets (Data generation scheme 2)

Figure 3:

The mean of $c_n(\phi_1, \phi_0, k)$ of 100 simulations on $\mathscr{D}_2 = [0, 1]^2$. The error bars represent one standard deviation. The sample size n take on values in 81, 256, 729, 2209 and 6561. The graphs in red and blue show the results using maximin ordering and sorted coordinate ordering, respectively.



(a) Increasingly finer grids (Data generation scheme 1)

(b) Successively increasing datasets (Data generation scheme 2)

Figure 4:

The mean of $c_n(\phi_1, \phi_0, k)$ of 100 simulations on $\mathscr{D}_2 = [0, 1]^2$. The error bars represent one standard deviation. The sample size n take on values in 81, 256, 729, 2209 and 6561. The graphs in red and blue show the results using maximin ordering and sorted coordinate ordering, respectively.