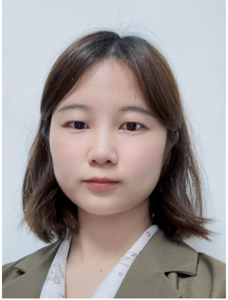
Bayesian Fixed-domain Asymptotics for Covariance Parameters in Spatial Gaussian Process Models

Institute for Mathematical Science, NUS
November, 2023

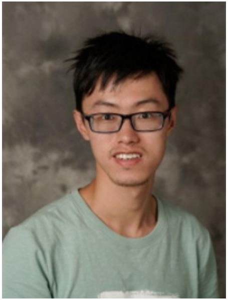
Cheng LI

National University of Singapore
Department of Statistics and Data Science
Faculty of Science

Joint Work With



Saifei Sun
Assistant Professor
City University of Hong Kong



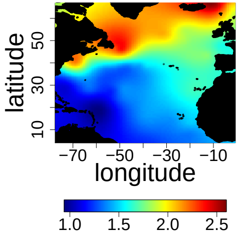
Yichen Zhu
Postdoctoral Researcher
Bocconi University

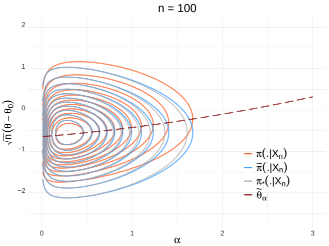
Introduction

Based on two papers:

- **Li, C.** (2022) Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model. *AoS*, arXiv:2010.02126.
- **Li, C.**, S. Sun, and Y. Zhu (2023+) Fixed-domain posterior contraction rates for spatial Gaussian process model with nugget. *JASA*, arXiv:2207.10239.

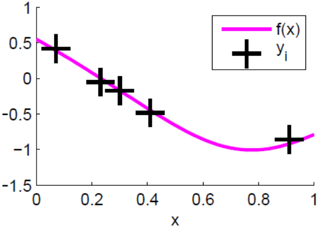
This research is at the interface of two fields: **spatial statistics** and **Bayesian asymptotics**.

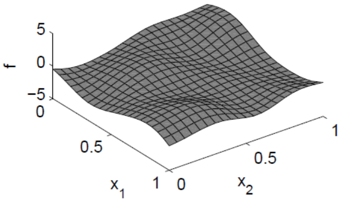




Gaussian Process

- **Gaussian Processes (or Gaussian random fields)** is widely used for interpolation in machine learning, spatial statistics, and computer models.





Gaussian Process

- $X \sim \text{GP}(\mu, K)$ means that $X(\cdot)$ is a Gaussian process on a spatial domain $\mathcal{S} \subseteq \mathbb{R}^d$, with mean function $\mu(s) : \mathcal{S} \to \mathbb{R}$ and covariance function $K(s, s') : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$.

- For any collection of distinct $s_1, \ldots, s_n \in \mathcal{S}$, the random vector $X_n \sim \mathcal{N}(\mu_n, K_n)$, where

$$X_n = \begin{pmatrix} X(s_1) \\ \vdots \\ X(s_n) \end{pmatrix}, \ \mu_n = \begin{pmatrix} \mu(s_1) \\ \vdots \\ \mu(s_n) \end{pmatrix}, \ K_n = \begin{pmatrix} K(s_1, s_1) & \ldots & K(s_1, s_n) \\ \vdots & \ddots & \vdots \\ K(s_n, s_1) & \ldots & K(s_n, s_n) \end{pmatrix}.$$

- Given the observation X_n, we can interpolate the function value at a new location s^* as $X(s^*)|X_n \sim \mathcal{N}(\mu^*, k^*)$, where

$$\begin{aligned} \mu^* &= \mu(s^*) + k_n(s^*)^\top K_n^{-1}(X_n - \mu_n), \\ k^* &= K(s^*, s^*) - k_n(s^*)^\top K_n^{-1} k_n(s^*), \\ k_n(s^*)^\top &= \begin{pmatrix} K(s_1, s^*) & \ldots & K(s_n, s^*) \end{pmatrix}. \end{aligned}$$

Spatial Gaussian Process Regression

- Gaussian process (GP) regression models in spatial statistics:

Universal Kriging Model

$$Y(s_i) = m(s_i)^\top \beta + X(s_i), \quad i = 1, \ldots, n;$$
$$X(\cdot) \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,\nu}).$$

Model with Nugget

$$Y(s_i) = m(s_i)^\top \beta + X(s_i) + \epsilon(s_i), \quad i = 1, \ldots, n;$$
$$X(\cdot) \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,\nu}), \quad \epsilon(\cdot) \sim N(0, \tau).$$

- $m(\cdot)$: p-dimensional spatially referenced predictors; coefficients $\beta \in \mathbb{R}^p$
- $\sigma^2 K_{\alpha,\nu}(\cdot,\cdot)$: Matérn covariance function
- $\epsilon(\cdot)$: measurement error / nugget

Spatial Gaussian Process Regression

Recent decades have seen an increasing volume of massive spatial and spatiotemporal data. One example is the remote sensing data in Geographic Information Systems (GIS).

- **Parameter estimation**: Estimation of $\beta, \sigma^2, \alpha, \nu, \tau$, etc.
  Identification; Interpretation; Large sample properties, ...
- **Prediction**: Predicting $Y(\cdot)$ at a new location s^*
- **Spatial correlation:** Make sense of $\text{Cov}(Y(s), Y(s'))$ for $s \neq s'$?
  Related to the covariance estimation in GP.

Our work focuses on the **parameter estimation** and **prediction**.

Covariance Function

- We focus on $X(\cdot) \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$, where $\sigma^2 K_{\alpha,\nu}$ is the **isotropic Matérn covariance function**

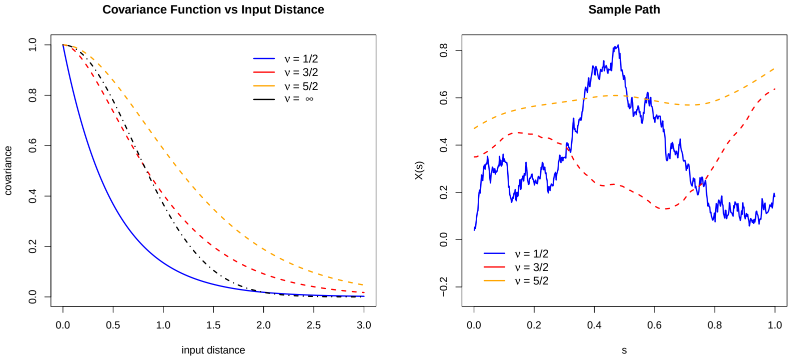
$$\sigma^2 K_{\alpha,\nu}(s-t) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\alpha\|s-t\|)^{\nu} \mathcal{K}_{\nu}(\alpha\|s-t\|)$$

  for any $s, t \in \mathcal{S}$, where $\mathcal{K}_{\nu}(\cdot)$ is the modified Bessel function of the second kind, and $\|\cdot\|$ is the Euclidean norm.

- $\nu > 0$ is the **smoothness parameter**:
  - $\nu = 1/2$ and $d = 1$: $\sigma^2 K_{\alpha,\nu}(s-t) = \sigma^2 \exp(-\alpha|s-t|)$, the Ornstein-Uhlenbeck (OU) process; sample path continuous but not differentiable.
  - $\nu \geq m + \frac{d}{2}$: $X \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$ is m times mean square differentiable.
  - $\nu \to \infty$: becomes the squared exponential covariance function $c_1 \exp(-c_2\|s-t\|^2)$.

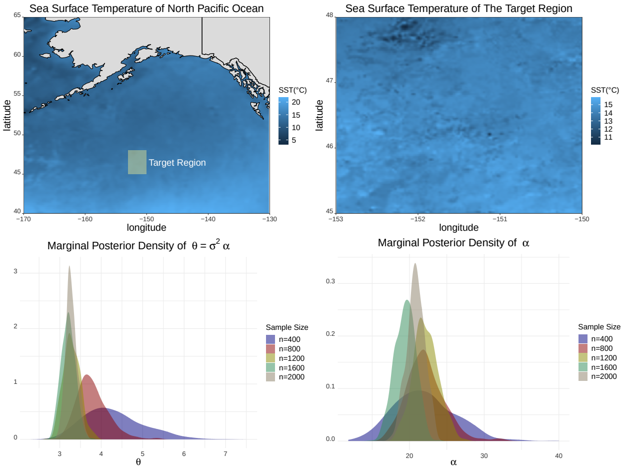
- $\sigma^2 > 0$ is the variance (or partial sill) parameter, and $\alpha > 0$ is the inverse range (or length-scale) parameter. They control the vertical and horizontal scaling of the covariance function.

Matérn Covariance Function



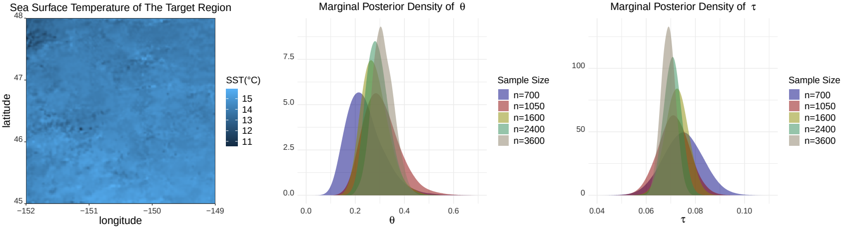
Example: Sea Surface Temperature Data

- When we fit the universal kriging model $Y(\cdot) = m(\cdot)^\top \beta + X(\cdot)$;
- $m(\cdot)$: 10 monomials of latitude and longitude up to degree 3;
- $X(\cdot)$: Matérn with $\nu = 1/2$ (exponential covariance function).



Example: Sea Surface Temperature Data

- When we fit the model with nugget $Y(\cdot) = m(\cdot)^\top \beta + X(\cdot) + \epsilon(\cdot)$;
- $m(\cdot)$: 3 monomials of latitude and longitude up to degree 1;
- $X(\cdot)$: Matérn with $\nu = 1/2$ (exponential covariance function).



Parameter Estimation in Covariance Function

- The main focus of this work is on the estimation of the **covariance parameters** σ^2 and α in $\mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$, and also the **nugget parameter** τ in the model with nugget.
- The smoothness parameter ν is assumed to be known and fixed.
- Estimation of ν is important and technically challenging, with some recent progress in the frequentist literature (Wu, Lim and Xiao '13 JMVA, Loh '15 AOS, Wu and Lim '16 Stat. Sin., Loh, Sun and Wen '21 AOS, Loh and Sun '23 Bernoulli).
- Our work is on the Bayesian large sample properties (asymptotics).

Parameter Estimation in Covariance Function

- Our overall model setup is **different** from **Bayesian nonparametric regression using Gaussian process priors**, such as van der Vaart and van Zanten ('08 AOS, '09 AOS, '11 JMLR).
- In **Bayesian nonparametric regression**, they assume that
  - $Y(\cdot)$ is a **true deterministic function** $(m(\cdot)^\top \beta + X(\cdot))$ plus some i.i.d. noise $(\epsilon(\cdot))$.
  - The GP model on $X(\cdot)$ is merely a **prior**. The **true** $X(\cdot)$ does not need to be a sample path from $\text{GP}(0, \sigma^2 K_{\alpha,\nu})$ – it only needs to be well approximated by the GP.
  - There are **no "true" parameters**. All GP parameters (σ^2, α, ν) are merely tuning parameters.
- In contrast, we assume that $X(\cdot) \sim \text{GP}(0, \sigma^2 K_{\alpha,\nu})$ is the **true** model. Therefore, $m(\cdot)^\top \beta + X(\cdot)$ is a random function, not a deterministic function. We assume that there are **true** parameters of σ^2, α, ν.

Bayesian Setup for Universal Kriging

- We first study the universal kriging model (in Li 2022):

$$Y(s_i) = m(s_i)^\top \beta + X(s_i), \quad i = 1, \dots, n;$$
$$X(\cdot) \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,\nu}).$$

- $m(\cdot) = (m_1(\cdot), \dots, m_p(\cdot))^\top$ is a vector of p known functions.
- We observe $Y_n = (Y(s_1), \dots, Y(s_n))^\top$ and M_n, the stacked obs of $m(\cdot)$.
- If $X_n = (X(s_1), \dots, X(s_n))^\top$, then the model is $Y_n = M_n \beta + X_n$.
- We impose the normal prior $\beta \mid \sigma^2, \alpha \sim \mathcal{N}\left(0_p, \sigma^2 \Omega_\beta^{-1}\right)$.
- Ω_β can be $\mathbf{0}_{p\times p}$, leading to a noninformative (improper) prior.
- The posterior of β conditional on (σ^2, α) is

$$\beta \mid \sigma^2, \alpha, Y_n, M_n \sim \mathcal{N}\left(\widetilde{\beta}_\alpha, \sigma^2 \left(M_n^\top R_\alpha^{-1} M_n + \Omega_\beta\right)^{-1}\right),$$

where $\widetilde{\beta}_\alpha = \left(M_n^\top R_\alpha^{-1} M_n + \Omega_\beta\right)^{-1} M_n^\top R_\alpha^{-1} Y_n$.

Bayesian Setup for Universal Kriging

- The log-likelihood function is

$$\mathcal{L}_n(\beta, \sigma^2, \alpha) = -\frac{n}{2}\log\sigma^2 - \frac{1}{2}\log|R_\alpha| - \frac{1}{2\sigma^2}(Y_n - M_n\beta)^\top R_\alpha^{-1}(Y_n - M_n\beta),$$

  where R_α is the $n \times n$ **Matérn correlation matrix**, whose (i,j)-entry is $K_{\alpha,\nu}(s_i - s_j)$, for $1 \le i, j \le n$. $|R_\alpha|$ is the determinant of R_α. So $X_n \sim \mathcal{N}(0, \sigma^2 R_\alpha)$.

- If we integrate out β from the posterior and then maximize over σ^2, we obtain the **REML** $\widetilde{\sigma}^2_\alpha$

$$\widetilde{\sigma}^2_\alpha = \frac{Y_n^\top \left[R_\alpha^{-1} - R_\alpha^{-1}M_n\left(M_n^\top R_\alpha^{-1}M_n + \Omega_\beta\right)^{-1}M_n^\top R_\alpha^{-1}\right] Y_n}{n - p}.$$

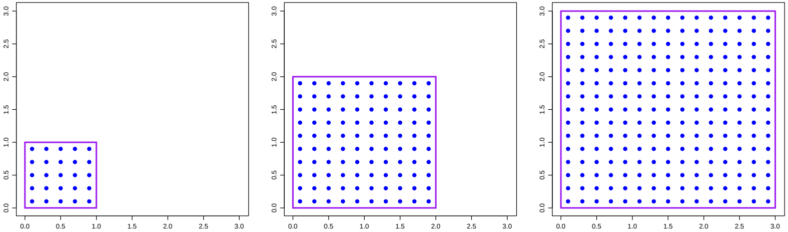
  If $p = 0$ and $\Omega_\beta = \mathbf{0}_{p\times p}$, then this is the **MLE** of σ^2 given α.

- There is no closed-form REML or MLE for the range parameter α.

Two Asymptotic Regimes

Estimation of (σ^2, α) falls into two asymptotic regimes:

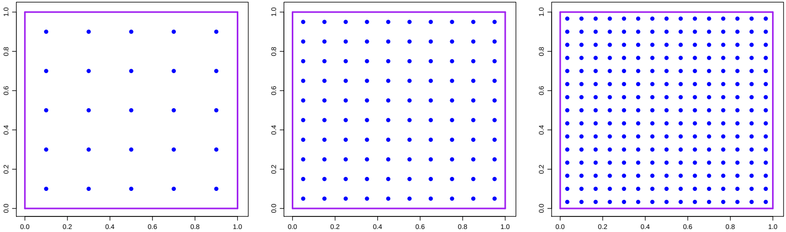
- **Increasing-domain asymptotics**: The domain $\mathcal{S}$ increases as the sample size *n* increases.
- The adjacent points have a minimum distance apart $\Longrightarrow$ weak dependence.



Two Asymptotic Regimes

Estimation of (σ^2, α) falls into two asymptotic regimes:

- **Fixed-domain asymptotics**: Also known as infill asymptotics. The domain $\mathcal{S}$ remains fixed and bounded as *n* increases.
- The adjacent points getting closer and closer $\Longrightarrow$ increasingly strong dependence.



Challenge in Fixed-Domain Asymptotics

- Spatial applications mostly have locations in a spatial or a spatiotemporal domain.

- This implies that the dimension of the location index $s \in \mathcal{S} \subseteq \mathbb{R}^d$ is $d = 1, 2, 3$, and $\mathcal{S}$ is a fixed and bounded domain, such as $[0, 1]^d$.

- **A negative result:** For dimension $d = 1, 2, 3$, Zhang (2004, JASA) has shown that there exists no consistent estimator for σ^2 and α in the isotropic Matérn covariance function.

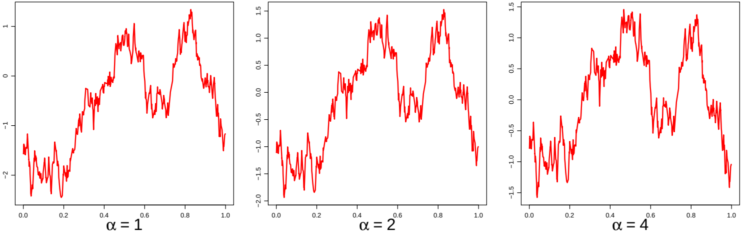
- For two sets of parameters (σ_1^2, α_1) and (σ_2^2, α_2) (with the same ν), the measures induced by the two Gaussian processes $\text{GP}(0, \sigma_1^2 K_{\alpha_1,\nu})$ and $\text{GP}(0, \sigma_2^2 K_{\alpha_2,\nu})$ are **equivalent** to each other, if and only if

$$\sigma_1^2 \alpha_1^{2\nu} = \sigma_2^2 \alpha_2^{2\nu}.$$

  Otherwise, if $\sigma_1^2 \alpha_1^{2\nu} \neq \sigma_2^2 \alpha_2^{2\nu}$, then the two Gaussian measures are orthogonal.

Challenge in Fixed-Domain Asymptotics

- For two **equivalent** Gaussian processes, it is impossible to estimate the parameters (σ^2, α) based on a single sample path:



- However, since the equivalence relation is totally determined by the product $\sigma^2\alpha^{2\nu}$, it is possible to consistently estimate $\theta = \sigma^2\alpha^{2\nu}$.

- θ is called the **microergodic parameter** (Stein 1999).

- The microergodic parameter θ is crucial for the prediction (kriging) performance.

Bayesian Fixed-Domain Asymptotics

- We reparametrize the model by replacing σ^2 with $\theta = \sigma^2\alpha^{2\nu}$. This leads to a model with parameters (θ, α).

- We assume that there are true parameters $\beta_0, \sigma_0^2, \alpha_0$, such that for all $s \in \mathcal{S}$,

$$Y(s) = m(s)^\top \beta_0 + X(s), \text{ and } X \sim \mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu}).$$

- **Assumption**: $m_1(\cdot), \ldots, m_p(\cdot)$ have $(\nu + d/2)$-times bounded derivatives on $\mathcal{S}$.

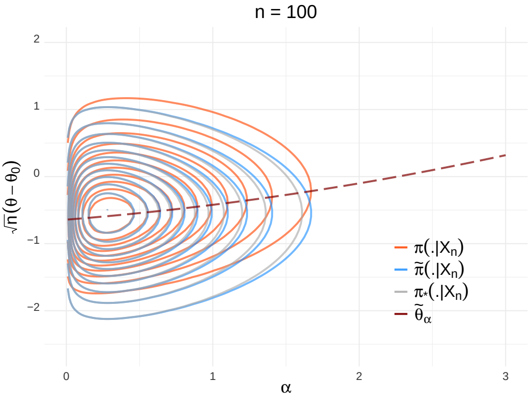
- With a prior distribution $\pi(\beta, \theta, \alpha)$, the posterior of (β, θ, α) given the data (Y_n, M_n) can be written as

$$\pi(\beta, \theta, \alpha | Y_n, M_n) = \frac{\exp\left\{\mathcal{L}_n(\beta, \theta/\alpha^{2\nu}, \alpha)\right\} \pi(\beta|\theta,\alpha)\pi(\theta|\alpha)\pi(\alpha)}{\int \exp\left\{\mathcal{L}_n(\beta', \theta'/\alpha'^{2\nu}, \alpha')\right\} \pi(\beta'|\theta',\alpha')\pi(\theta'|\alpha')\pi(\alpha')\mathrm{d}\alpha'\mathrm{d}\theta'\mathrm{d}\beta'}.$$

- Question: Does the posterior of θ and α converge or not?
  (Yes for θ; No for α.)

Properties of REML

- The REML $\widetilde{\theta}_\alpha$ (dashed line) is an increasing function in α.
- As $n \to \infty$, $\widetilde{\theta}_\alpha$ is close to $\widetilde{\theta}_{\alpha_0}$ uniformly over $\alpha \in [n^{-\underline{\kappa}}, n^{\overline{\kappa}}]$.
- The Bernstein-von Mises (BvM) theorem seems to hold for θ, but not α.



Limiting Posterior Distribution in Universal Kriging

The profile restricted log-likelihood (after integrating out β from the posterior) is

$$\widetilde{\mathcal{L}}_n(\alpha) = -\frac{n-p}{2}\log\frac{Y_n^\top\left[R_\alpha^{-1} - R_\alpha^{-1}M_n\left(M_n^\top R_\alpha^{-1}M_n + \Omega_\beta\right)^{-1}M_n^\top R_\alpha^{-1}\right]Y_n}{n-p}$$
$$-\frac{1}{2}\log|R_\alpha| - \frac{1}{2}\log\left|M_n^\top R_\alpha^{-1}M_n + \Omega_\beta\right| - \frac{n-p}{2}.$$

Theorem 1 (**L. 2022** for Universal Kriging)

Under some mild assumptions, the posterior distributions of θ and α are asymptotically independent, and the joint posterior of (θ,α) satisfies

$$\left\|\Pi(\mathrm{d}\theta,\mathrm{d}\alpha|Y_n,M_n) - \mathcal{N}\left(\mathrm{d}\theta \mid \widetilde{\theta}_{\alpha_0}, \frac{2\theta_0^2}{n}\right)\times\widetilde{\Pi}(\mathrm{d}\alpha|Y_n,M_n)\right\|_{\mathsf{TV}} \to 0,$$

as $n\to\infty$ almost surely, where $\widetilde{\Pi}(\mathrm{d}\alpha|Y_n,M_n)$ is profile posterior distribution with the density

$$\widetilde{\pi}(\alpha|Y_n,M_n) = \frac{\exp\left\{\widetilde{\mathcal{L}}_n(\alpha)\right\}\pi(\alpha|\theta_0)}{\int_0^\infty \exp\left\{\widetilde{\mathcal{L}}_n(\alpha')\right\}\pi(\alpha'|\theta_0)\mathrm{d}\alpha'},$$

Explicit Profile Posterior for 1d OU Process

Consider the following special case:

- $p = 0$ (no regression terms $m(\cdot)^\top \beta$);
- $d = 1$, $\mathcal{S} = [0, 1]$;
- $\nu = 1/2$, $Y = X \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,1/2})$ (exponential covariance function);
- The sampling points $\mathcal{S}_n = \{s_1, \ldots, s_n\}$ are the equispaced grid with $s_i = i/n$ for $i = 1, \ldots, n$.
- This gives the 1-d Ornstein-Uhlenback (OU) process.

Explicit Profile Posterior for 1d OU Process

Theorem 2 (**L. 2022** Limiting Posterior for 1d OU process)

Under some relaxed assumptions on the prior of (θ, α)*,*

$$\left\| \Pi(\mathrm{d}\theta, \mathrm{d}\alpha | Y_n) - \mathcal{N}\left(\mathrm{d}\theta \middle| \widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n\right) \times \Pi_*(\mathrm{d}\alpha | Y_n) \right\|_{\mathsf{TV}} \to 0,$$

as $n \to \infty$ *in probability, where* $\Pi_*(\mathrm{d}\alpha | Y_n)$ *has the density*

$$\pi_*(\alpha | Y_n) \propto \sqrt{\alpha} \exp\left\{ -\frac{(\alpha - u_*)^2}{2v_*} \right\} \cdot \pi(\alpha | \theta_0), \text{ for all } \alpha \in \mathbb{R}_+,$$

$$\text{with} \quad u_* = \frac{n(A_1 - A_2)}{A_1}, \quad v_* = \frac{n(A_1 - 2A_2 + A_3)}{A_1},$$

$$A_1 = \sum_{i=2}^{n-1} Y(s_i)^2, \quad A_2 = \sum_{i=1}^{n-1} Y(s_i) Y(s_{i+1}), \quad A_3 = \sum_{i=1}^{n} Y(s_i)^2.$$

Furthermore, $v_* > 0$ *and* $v_* \asymp 1$ *as* $n \to \infty$ *in probability. Therefore, the posterior of range parameter* α*,* $\pi(\alpha | Y_n)$*, does not converge to any point mass.*

A Simulation Study

- Consider the 1d OU process (isotropic Matérn with $\nu = 1/2$), without regression terms $m(\cdot)^\top \beta$.
- The true parameters are $\sigma_0^2 = 2$, $\alpha_0 = 1$, and $\theta_0 = \sigma_0^2 \alpha_0^{2\nu} = 2$.
- For the $d = 1$ case, $\mathcal{S} = [0, 1]$, sampling locations $s_i = \frac{2i-1}{2n}$ for $i = 1, \ldots, n$. Sample size $n = 25, 50, 100, 200, 400$.
- We directly observe $Y_n = X_n$ from $\mathrm{GP}(0, \sigma_0^2 K_{\alpha_0,\nu})$ on the sampling locations.
- Independent gamma priors on θ and α, with the same shape parameter 1.1 and rate parameter 0.1.
- Random walk Metropolis algorithm (RWM): draw 5000 samples after 1000 burnins from the true joint posterior $\Pi(\mathrm{d}\theta, \mathrm{d}\alpha | X_n)$, the limiting posteriors $\mathcal{N}\left(\mathrm{d}\theta \middle| \widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n\right) \times \widetilde{\Pi}(\mathrm{d}\alpha | X_n)$ in Theorem 1, and the limiting posterior $\mathcal{N}\left(\mathrm{d}\theta \middle| \widetilde{\theta}_{\alpha_0}, 2\theta_0^2/n\right) \times \Pi_*(\mathrm{d}\alpha | X_n)$ in Theorem 2.
- We calculate the the Wasserstein-2 (W_2) distance between the true posterior and our approximations.

Posterior Means and Variances

All numbers are averaged over 100 macro replications. The true parameter values are $\theta_0 = 2$ and $\alpha_0 = 1$.

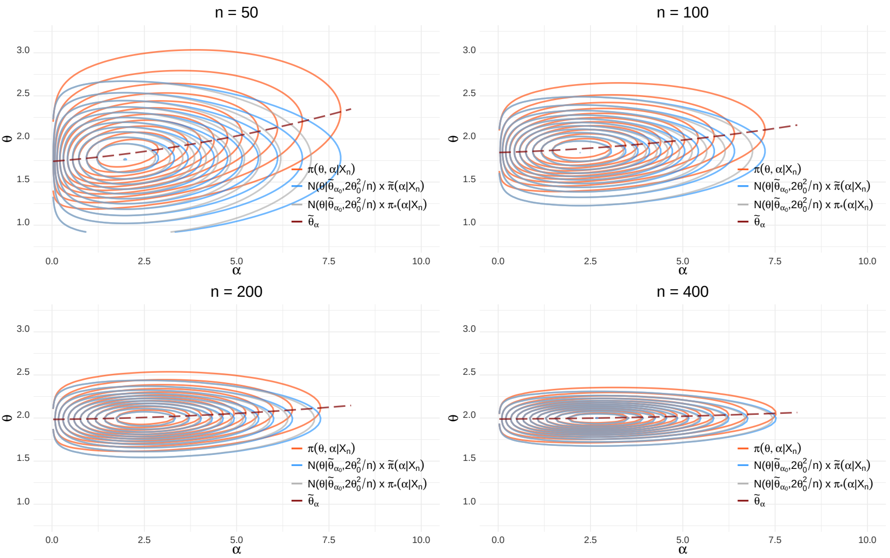
$d = 1$	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
$\mathrm{E}(\theta\|X_n)$	2.6795	2.1932	2.1467	2.0740	2.0320
$\mathrm{Var}(\theta\|X_n)$	0.9825	0.2441	0.1031	0.0455	0.0212
$\widetilde{\mathrm{E}}(\theta\|X_n)$	2.0404	1.9357	2.0214	2.0130	2.0028
$\widetilde{\mathrm{Var}}(\theta\|X_n)$	0.3197	0.1599	0.0798	0.0399	0.0200
$\mathrm{E}(\alpha\|X_n)$	3.1924	2.9803	2.7392	2.9947	2.5075
$\mathrm{Var}(\alpha\|X_n)$	5.3673	4.0441	2.9987	3.7074	2.5080
$\widetilde{\mathrm{E}}(\alpha\|X_n)$	2.9717	2.8767	2.6941	2.9534	2.5012
$\widetilde{\mathrm{Var}}(\alpha\|X_n)$	4.5474	3.7045	2.9094	3.6840	2.4664
$\mathrm{E}_*(\alpha\|X_n)$	2.5267	2.6534	2.5873	2.9105	2.4933
$\mathrm{Var}_*(\alpha\|X_n)$	2.5207	2.7894	2.5783	3.3733	2.4291

W_2 Distances

All numbers are averaged over 100 macro replications. Standard errors are in the parentheses.

$d = 1$	$n = 25$	$n = 50$	$n = 100$	$n = 200$	$n = 400$
$W_2\left(\Pi(\mathrm{d}\theta \mid X_n), \mathcal{N}\left(\widetilde{\theta}_{\alpha_0}, \frac{2\theta_0^2}{n}\right)\right)$	0.8051 (0.0326)	0.3000 (0.0101)	0.1449 (0.0042)	0.0706 (0.0024)	0.0335 (0.0010)
$W_2(\Pi(\mathrm{d}\alpha \mid X_n), \widetilde{\Pi}(\mathrm{d}\alpha \mid X_n))$	0.3175 (0.0290)	0.1807 (0.0183)	0.1260 (0.0086)	0.1303 (0.0099)	0.1073 (0.0077)
$W_2(\Pi(\mathrm{d}\alpha \mid X_n), \Pi_*(\mathrm{d}\alpha \mid X_n))$	0.8972 (0.0874)	0.4259 (0.0504)	0.2131 (0.0211)	0.1583 (0.0160)	0.1095 (0.0075)

Contour Plots of Two Approximations



Model with Nugget

- Based on results from Li, Sun and Zhu (2023+) JASA paper.
- We further consider the universal kriging model with nugget $Y(s) = m(s)^\top \beta + X(s) + \epsilon(s)$, where $X \sim \mathrm{GP}(0, \sigma^2 K_{\alpha,\nu})$ and $\epsilon \sim N(0, \tau)$.
- Similar to the model without nugget, when $d = 1, 2, 3$, the microergodic parameter $\theta = \sigma^2 \alpha^{2\nu}$ can be consistently estimated, but σ^2 and α cannot.
- In general, when $d = 1, 2, 3$, β, σ^2, α do not have consistent estimators under fixed-domain asymptotics, and hence, no posterior consistency.
- We focus on the Bayesian posterior distribution of the microergodic parameter θ and the nugget parameter τ.

Impact from Nugget

The nugget brings a big difference to the fixed-domain asymptotic theory.

- The nugget parameter τ itself can be estimated with the parametric rate of $\mathcal{O}(n^{-1/2})$.
- However, the microergodic parameter $\theta = \sigma^2\alpha^{2\nu}$ no longer has the $\mathcal{O}(n^{-1/2})$ rate.
- Chen et al. '00 Stat. Sin. shows that for the 1d OU process example, the convergence rate of the MLE of θ **deteriorates** from $\mathcal{O}(n^{-1/2})$ in the model without nugget, to $\mathcal{O}(n^{-1/4})$ in the model with nugget.
- Tang et al. '21 JRSSB shows that under some mathematical assumptions, for Y_n on the equispaced grids, for isotropic Matérn with general ν, the MLE of θ converges at the rate $\mathcal{O}\left(n^{-\frac{1}{2(2\nu/d+1)}}\right)$.
- The convolution of the GP X with the Gaussian noise ϵ makes the estimation more difficult. Furthermore, the MLEs of θ and τ have no closed forms or even first-order approximation.

Posterior Contraction Rates for Isotropic Matérn

Theorem 3 (**L.**, Sun, Zhu 2023+ for Model with Nugget)

Under some mild assumptions, for a general class of stratified sampling designs $\mathcal{S}_n$ in $[0,1]^d$, the posterior distribution of θ and τ satisfies that for any positive sequence $L_n \to \infty$ as $n \to \infty$,

$$\Pi\left(\left|\frac{\theta}{\theta_0} - 1\right| < L_n n^{-\frac{1}{2(4\nu/d+1)}+\varrho} \log n, \right.$$

$$\left. \text{and } \left|\frac{\tau}{\tau_0} - 1\right| < L_n n^{-\frac{1}{2}} \log n \;\middle|\; Y_n, M_n \right) \to 1,$$

as $n \to \infty$ almost surely, where $\varrho > 0$ is a fixed number related to the prior and can be arbitrarily small.

New Proof Techniques

The main challenge for the model with nugget is that the likelihood function is difficult to work with – no closed-form MLE, not even closed-form first-order Taylor expansions for θ and τ.

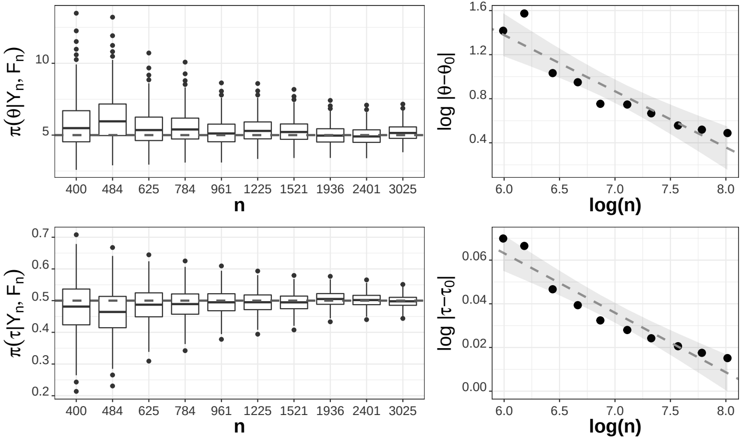
We solve this parametric problem in a nonparametric way:

- We take the idea from the **Schwarz's posterior consistency theorem** in Bayesian nonparametrics, which needs two components:
  - (i) A consistent frequentist estimator that satisfies some concentration inequalities with exponentially small tail probabilities;
  - (ii) An evidence lower bound that decays only polynomially in n as $n \to \infty$.
- We derive a new evidence lower bound for strongly dependent data Y_n that satisfies (ii).
- We use the **higher-order quadratic variation estimators** of θ and τ, developed in Loh '15 AOS, Loh, Sun and Wen '21 AOS, Loh and Sun '23 Bernoulli. Our posterior contraction rates are inherited from the rates of these estimators.

A Simulation Study

- Consider the 2d isotropic Matérn with $\nu = 1/2$ and $\nu = 1/4$; $m(\cdot)$ includes 6 monomials up to degree 2.
- The true parameters are $\theta_0 = 5, \alpha_0 = 1, \tau_0 = 0.5$, $\beta_0 = (1, -1.5, -1.5, 2, 1, 2)^\top$.
- Domain $\mathcal{S} = [0, 1]^2$; Sampling locations in $\mathcal{S}_n$ are the regular grid $((2i - 1)/(2m), (2j - 1)/(2m))$ for $i, j = 1, \ldots, m$.
- Sample size $n = m^2 \approx 400 \times 1.25^{k-1}$ for $k = 1, \ldots, 10$.
- We draw observations Y_n from $\mathrm{GP}(m(\cdot)^\top \beta_0, \sigma_0^2 K_{\alpha_0,\nu})$ on $\mathcal{S}_n$.
- Prior: $\beta \sim N(0, 10^6)$, $\theta \sim \mathrm{InvGamma}(0.1, 0.1)$, $\tau \sim \mathrm{InvGamma}(0.1, 0.1)$, $\alpha \sim \mathrm{InvGaussian}(1, 1)$.
- We check how the marginal posterior distributions of θ and τ change with n, as well as the posterior preditive MSEs.

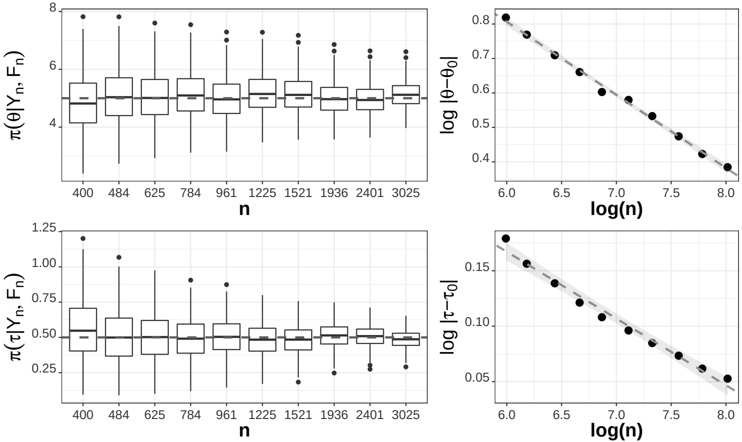
Posterior Contraction for $\nu = 1/2$



Left column: Boxplots for the marginal posterior densities of θ and τ versus the increasing sample size n.

Right column: Posterior means of $|\theta - \theta_0|$ and $|\tau - \tau_0|$ versus the increasing sample size n, on the logarithm scale.

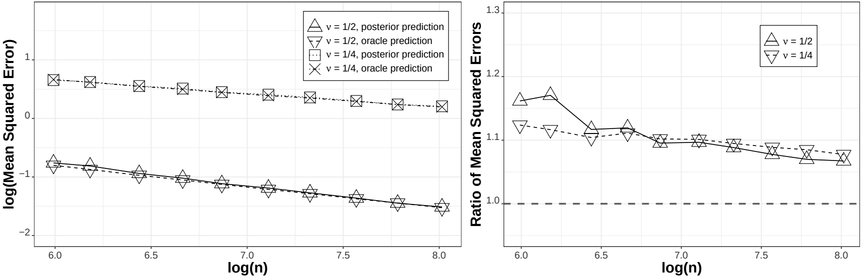
Posterior Contraction for $\nu = 1/4$



Left column: Boxplots for the marginal posterior densities of θ and τ versus the increasing sample size n.

Right column: Posterior means of $|\theta - \theta_0|$ and $|\tau - \tau_0|$ versus the increasing sample size n, on the logarithm scale.

Prediction MSE for $\nu = 1/2$ and $\nu = 1/4$



Left panel: The prediction mean squared errors under both the Bayesian posterior prediction and the oracle prediction based on the true parameters.

Right panel: Ratios of the Bayesian prediction mean squared error and the oracle prediction mean squard error.

Conclusion

- We have explored the Bayesian large sample properties for the covariance parameters in the universal kriging model without and with nugget, under fixed-domain asymptotics.
- There is no posterior contraction for the individual parameters σ^2 and α with the domain dimension $d = 1, 2, 3$.
- Posterior contraction rates are derived for the microergodic parameter θ and the nugget τ.
- Bayesian posterior prediction performance remains asymptotically efficient with misspecified α, theoretically and empirically.

The End

This work is supported by the Singapore Ministry of Education Academic Research Funds Tier 1 Grants R-155-000-201-114 and A-0004822-00-00.

- Li. C. (2022) Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model. *Annals of Statistics* 50(6): 3334–3363, arXiv:2010.02126.
- Li, C., S. Sun, and Y. Zhu (2023+) Fixed-domain posterior contraction rates for spatial Gaussian process model with nugget. *Journal of the American Statistical Association*, forthcoming, arXiv:2207.10239.

Thank you! & Questions?

NUS National University of Singapore | Department of Statistics and Data Science Faculty of Science