Gibbs Reference Posterior distribution for Kriging parameters

Joseph Muré [1,2] Josselin Garnier [3]
Loic Le Gratiet [1] Anne Dutfoy [1]

[1]EDF R&D

[2]Université Paris Diderot

[3]École Polytechnique

June 12th 2017

Contents

Contents

The Model

Gaussian spatial process with **null** mean function and **stationary** covariance function/kernel $k \to$ Simple Kriging.

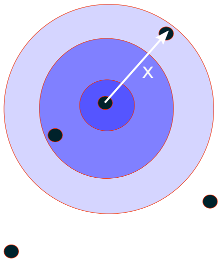
Matérn isotropic kernel with parameters $\sigma^2 > 0$, $\theta > 0$ and $\nu > 0$:

$$K(x) = \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{2\sqrt{\nu}|x|}{\theta} \right)^{\nu} B_\nu \left(\frac{2\sqrt{\nu}|x|}{\theta} \right),$$

Parameter interpretation :

Notations :

- B_ν : modified Bessel function of the second kind.



The Model

Gaussian spatial process with **null** mean function and **stationary** covariance function/kernel $k \rightarrow$ Simple Kriging.

Matérn isotropic kernel with parameters $\sigma^2 > 0$, $\theta > 0$ and $\nu > 0$:

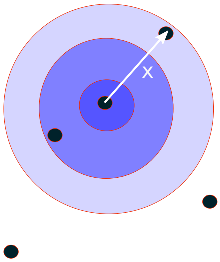
$$K(x) = \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{2\sqrt{\nu}|x|}{\theta} \right)^{\nu} B_{\nu} \left(\frac{2\sqrt{\nu}|x|}{\theta} \right),$$

Parameter interpretation :

- σ^2 : variance of the Gaussian Process. $K(0) = \lim_{x \to 0} K(x) = \sigma^2$

Notations :

- B_ν : modified Bessel function of the second kind.



The Model

Gaussian spatial process with **null** mean function and **stationary** covariance function/kernel $k \rightarrow$ Simple Kriging.

Matérn isotropic kernel with parameters $\sigma^2 > 0$, $\theta > 0$ and $\nu > 0$:

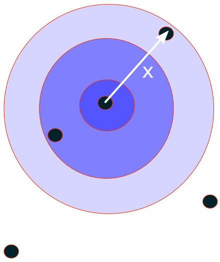
$$K(x) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{2\sqrt{\nu}|x|}{\theta} \right)^{\nu} B_{\nu} \left(\frac{2\sqrt{\nu}|x|}{\theta} \right),$$

Parameter interpretation :

- σ^2 : variance of the Gaussian Process.
  $K(0) = \lim_{x \rightarrow 0} K(x) = \sigma^2$
- θ : **correlation length** $\rightarrow$ a scaling parameter.
  $\theta \rightarrow 0$: different points in space uncorrelated.
  $\theta \rightarrow \infty$: GP expected to be nearly constant.

Notations :

- B_{ν} : modified Bessel function of the second kind.



The Model

Gaussian spatial process with **null** mean function and **stationary** covariance function/kernel $k \to$ Simple Kriging.

Matérn isotropic kernel with parameters $\sigma^2 > 0$, $\theta > 0$ and $\nu > 0$:

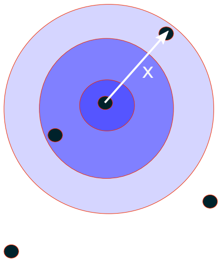
$$K(x) = \frac{\sigma^2}{\Gamma(\nu)2^{\nu-1}} \left(\frac{2\sqrt{\nu}|x|}{\theta} \right)^{\nu} B_{\nu} \left(\frac{2\sqrt{\nu}|x|}{\theta} \right),$$

Parameter interpretation :

- σ^2 : variance of the Gaussian Process. $K(0) = \lim_{x \to 0} K(x) = \sigma^2$
- θ : **correlation length** $\to$ a scaling parameter. $\theta \to 0$: different points in space uncorrelated. $\theta \to \infty$: GP expected to be nearly constant.
- ν : regularity. Assumed to be known.

Notations :

- B_{ν} : modified Bessel function of the second kind.



Parameter estimation problem

- Suppose we know $\boldsymbol{y} = (Y(\boldsymbol{x}^{(1)}), Y(\boldsymbol{x}^{(2)}), ..., Y(\boldsymbol{x}^{(n)}))$.
- Likelihood of parameters σ^2 and θ :

$$L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$$

Notations :

- $Y(\boldsymbol{x})$: value of GP Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

[a]Runze Li and Agus Sudjianto (2012). "Analysis of computer experiments using penalized likelihood in Gaussian Kriging models". In: *Technometrics*.

[b]Danilo Lopes (2011). "Development and implementation of Bayesian computer model emulators". PhD thesis. Duke University.

Parameter estimation problem

- Suppose we know $\boldsymbol{y} = (Y(\boldsymbol{x}^{(1)}), Y(\boldsymbol{x}^{(2)}), ..., Y(\boldsymbol{x}^{(n)}))$.
- Likelihood of parameters σ^2 and θ :

$$L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$$

- Maximum Likelihood Estimation **not robust** enough with few observation points :
  - the Likelihood may be rather **flat**[a] ;

Notations :

- $Y(\boldsymbol{x})$: value of GP Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

---

[a] Runze Li and Agus Sudjianto (2012). "Analysis of computer experiments using penalized likelihood in Gaussian Kriging models". In: *Technometrics*.

[b] Danilo Lopes (2011). "Development and implementation of Bayesian computer model emulators". PhD thesis. Duke University.

Parameter estimation problem

- Suppose we know $\boldsymbol{y} = (Y(\boldsymbol{x}^{(1)}), Y(\boldsymbol{x}^{(2)}), ..., Y(\boldsymbol{x}^{(n)}))$.
- Likelihood of parameters σ^2 and θ :

$$L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$$

- Maximum Likelihood Estimation **not robust** enough with few observation points :
  - the Likelihood may be rather **flat**[a] ;
  - its maximum may actually lie at $\theta = 0$ or ∞.[b]

[a] Runze Li and Agus Sudjianto (2012). "Analysis of computer experiments using penalized likelihood in Gaussian Kriging models". In: *Technometrics*.

[b] Danilo Lopes (2011). "Development and implementation of Bayesian computer model emulators". PhD thesis. Duke University.

Notations :

- $Y(\boldsymbol{x})$: value of GP Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

Contents

Bernardo's reference prior : brief introduction

Likelihood $L(y|\theta)$

Bernardo's reference prior : brief introduction

Likelihood $L(y|\theta) \rightarrow$ Likelihood of p independent realizations $L(y^{(p)}|\theta)$.

Bernardo's reference prior : brief introduction

Likelihood $L(y|\theta) \rightarrow$ Likelihood of p independent realizations $L(y^{(p)}|\theta)$.

For a given prior $\pi(\theta)$, define
$\pi\left(y^{(p)}, \theta\right) := L(y^{(p)}|\theta)\pi(\theta) =: \pi(\theta|y^{(p)})m_\pi\left(y^{(p)}\right)$.

Bernardo's reference prior : brief introduction

Likelihood $L(y|\theta) \rightarrow$ Likelihood of p independent realizations $L(y^{(p)}|\theta)$.

For a given prior $\pi(\theta)$, define
$\pi\left(y^{(p)}, \theta\right) := L(y^{(p)}|\theta)\pi(\theta) =: \pi(\theta|y^{(p)})m_\pi\left(y^{(p)}\right)$.
Expected intrinsic info $\mathcal{I}_\pi^{(p)} = \int m_\pi\left(y^{(p)}\right) \int \pi(\theta|y^{(p)}) \log \frac{\pi(\theta|y^{(p)})}{\pi(\theta)} d\theta dy^{(p)}$.

Bernardo's reference prior : brief introduction

Likelihood $L(y|\theta) \to$ Likelihood of p independent realizations $L(y^{(p)}|\theta)$.

For a given prior $\pi(\theta)$, define
$\pi\left(y^{(p)}, \theta\right) := L(y^{(p)}|\theta)\pi(\theta) =: \pi(\theta|y^{(p)}) m_\pi\left(y^{(p)}\right)$.

Expected intrinsic info $\mathcal{I}_\pi^{(p)} = \int m_\pi\left(y^{(p)}\right) \int \pi(\theta|y^{(p)}) \log \frac{\pi(\theta|y^{(p)})}{\pi(\theta)} d\theta dy^{(p)}$.

Bernardo's reference prior π^* : $\forall \pi, \lim_{p \to \infty} \left[\mathcal{I}_{\pi^*}^{(p)} - \mathcal{I}_\pi^{(p)}\right] \geqslant 0$.

Bernardo's reference prior : brief introduction

Likelihood $L(y|\theta) \to$ Likelihood of p independent realizations $L(y^{(p)}|\theta)$.

For a given prior $\pi(\theta)$, define
$\pi\left(y^{(p)},\theta\right) := L(y^{(p)}|\theta)\pi(\theta) =: \pi(\theta|y^{(p)})m_\pi\left(y^{(p)}\right)$.

Expected intrinsic info $\mathcal{I}_\pi^{(p)} = \int m_\pi\left(y^{(p)}\right)\int \pi(\theta|y^{(p)}) \log \frac{\pi(\theta|y^{(p)})}{\pi(\theta)} d\theta dy^{(p)}$.

Bernardo's reference prior π^* : $\forall \pi, \lim_{p\to\infty}\left[\mathcal{I}_{\pi^*}^{(p)} - \mathcal{I}_\pi^{(p)}\right] \geqslant 0$.

Under some regularity conditions, Bernardo's reference prior = Jeffreys' prior.

Bernardo's reference prior : brief introduction

Likelihood $L(y|\theta) \to$ Likelihood of p independent realizations $L(y^{(p)}|\theta)$.

For a given prior $\pi(\theta)$, define
$\pi\left(y^{(p)},\theta\right) := L(y^{(p)}|\theta)\pi(\theta) =: \pi(\theta|y^{(p)})m_\pi\left(y^{(p)}\right)$.

Expected intrinsic info $\mathcal{I}_\pi^{(p)} = \int m_\pi\left(y^{(p)}\right)\int \pi(\theta|y^{(p)})\log\frac{\pi(\theta|y^{(p)})}{\pi(\theta)}d\theta dy^{(p)}$.

Bernardo's reference prior π^* : $\forall\pi,\ \lim_{p\to\infty}\left[\mathcal{I}_{\pi^*}^{(p)} - \mathcal{I}_\pi^{(p)}\right] \geqslant 0$.

Under some regularity conditions, Bernardo's reference prior $=$ Jeffreys' prior.

Likelihood	Prior
$L(\cdot|\alpha,\beta,\gamma)$	$\pi^*(\alpha|\beta,\gamma)$
$L(\cdot|\beta,\gamma)$	$\pi^*(\beta|\gamma)$
$L(\cdot|\gamma)$	$\pi^*(\gamma)$

Table: Bernardo's reference prior with multiple parameters

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- Our **ordering** : $\sigma^2 \prec \theta$.

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- Our **ordering** : $\sigma^2 \prec \theta$. $\pi^0(\sigma^2 \mid \theta) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- Our **ordering** : $\sigma^2 \prec \theta$. $\pi^0(\sigma^2 \mid \theta) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \theta) \propto |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$ Improper likelihood ! Should be seen as the likelihood of $\boldsymbol{y}/\|\boldsymbol{y}\|$.

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- Our **ordering** : $\sigma^2 \prec \theta$. $\pi^0(\sigma^2 \mid \theta) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \theta) \propto |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$ Improper likelihood ! Should be seen as the likelihood of $\boldsymbol{y}/\|\boldsymbol{y}\|$.
- $\pi^1(\theta) \propto \sqrt{\mathrm{Tr}\left[\left(\frac{\partial}{\partial_\theta}\boldsymbol{\Sigma}_\theta\,(\boldsymbol{\Sigma}_\theta)^{-1}\right)^2\right] - \frac{1}{n}\,\mathrm{Tr}\left[\frac{\partial}{\partial_\theta}\boldsymbol{\Sigma}_\theta\,(\boldsymbol{\Sigma}_\theta)^{-1}\right]^2}$

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- Our **ordering** : $\sigma^2 \prec \theta$. $\pi^0(\sigma^2 \mid \theta) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \theta) \propto |\boldsymbol{\Sigma}_\theta|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_\theta^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$ Improper likelihood ! Should be seen as the likelihood of $\boldsymbol{y}/\|\boldsymbol{y}\|$.
- $\pi^1(\theta) \propto \sqrt{\mathrm{Tr}\left[\left(\frac{\partial}{\partial_\theta}\boldsymbol{\Sigma}_\theta \left(\boldsymbol{\Sigma}_\theta\right)^{-1}\right)^2\right] - \frac{1}{n}\,\mathrm{Tr}\left[\frac{\partial}{\partial_\theta}\boldsymbol{\Sigma}_\theta \left(\boldsymbol{\Sigma}_\theta\right)^{-1}\right]^2}$

π^1 is an **improper** distribution with respect to θ, but it leads to a **proper posterior** distribution.[a]

Notations :

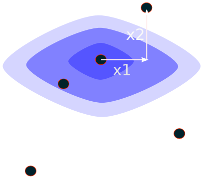
- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- θ : correlation length of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_\theta$

[a] James O Berger, Victor De Oliveira, and Bruno Sansó (2001). "Objective Bayesian analysis of spatially correlated data". In: *Journal of the American Statistical Association* 96.456, pp. 1361–1374.

Matérn anisotropic kernels

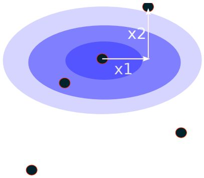
Matérn **tensorized** kernel with parameters $\sigma^2 > 0$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ and $\nu > 0$:

$$K_{\sigma^2,\boldsymbol{\theta},\nu}(\boldsymbol{x}) = \sigma^2 \prod_{i=1}^{n} K_{1,\boldsymbol{\theta}_i,\nu}(\boldsymbol{x}_i),$$



Matérn **geometric anisotropic** kernel with parameters $\sigma^2 > 0$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_n)$ and $\nu > 0$:

$$K_{\sigma^2,\boldsymbol{\theta},\nu}(\boldsymbol{x}) = K_{\sigma^2,1,\nu}\left(\sqrt{\sum_{i=1}^{n}\left(\frac{\boldsymbol{x}_i}{\boldsymbol{\theta}_i}\right)^2}\right),$$



Notations :

- σ^2 : variance of the Gaussian Process.
- ν : regularity.

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- $\pi^0(\sigma^2 \mid \boldsymbol{\theta}) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- $\boldsymbol{\theta}$: correlation lengths of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- $\pi^0(\sigma^2 \mid \boldsymbol{\theta}) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$
- Ordering of $\boldsymbol{\theta}_1,...,\boldsymbol{\theta}_r$? $\rightarrow$ none justifiable.

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- $\boldsymbol{\theta}$: correlation lengths of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- $\pi^0(\sigma^2 \mid \boldsymbol{\theta}) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$
- Ordering of $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_r$? $\rightarrow$ none justifiable.
- Solution : treat $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_r$ as a group.[a]

[a] Cuirong Ren, Dongchu Sun, and Sujit K Sahu (2013). "Objective Bayesian analysis of spatial models with separable correlation functions". In: *Canadian Journal of Statistics* 41.3, pp. 488–507.

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- $\boldsymbol{\theta}$: correlation lengths of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma_\theta}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma_\theta}^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- $\pi^0(\sigma^2 \mid \boldsymbol{\theta}) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma_\theta}|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma_\theta}^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$
- Ordering of $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_r$? $\rightarrow$ none justifiable.
- Solution : treat $\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_r$ as a group.

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- $\boldsymbol{\theta}$: correlation lengths of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma_\theta}$

Bernardo reference prior : application

- $L^0(\boldsymbol{y} \mid \sigma^2, \boldsymbol{\theta}) \propto \left(\frac{1}{\sigma^2}\right)^{\frac{n}{2}} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \exp\left\{-\frac{\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}}{2\sigma^2}\right\}$
- $\pi^0(\sigma^2 \mid \boldsymbol{\theta}) = \pi^0(\sigma^2) \propto \frac{1}{\sigma^2}$
- $L^1(\boldsymbol{y} \mid \boldsymbol{\theta}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-\frac{1}{2}} \cdot \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \boldsymbol{y}\right)^{-\frac{n}{2}}$
- Ordering of $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_r$? $\rightarrow$ none justifiable.
- Solution : treat $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_r$ as a group.
- Other solution : set $\pi^1(\boldsymbol{\theta}_i \mid \boldsymbol{\theta}_j \; \forall j \neq i) \propto$
$$\sqrt{\mathrm{Tr}\left[\left(\frac{\partial}{\partial \boldsymbol{\theta}_i} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right)^2\right] - \frac{1}{n} \mathrm{Tr}\left[\frac{\partial}{\partial \boldsymbol{\theta}_i} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right]^2}$$

Notations :

- $Y(\boldsymbol{x})$: value of Gaussian Process Y at point $\boldsymbol{x}$
- $\boldsymbol{y}$: vector of observations of Y at certain points
- σ^2 : variance of Y at any point
- $\boldsymbol{\theta}$: correlation lengths of Y
- $\mathbb{V}(\boldsymbol{y}) = \sigma^2 \boldsymbol{\Sigma}_{\boldsymbol{\theta}}$

Gibbs reference posterior

- Proper conditional prior :

$$\pi_i(\theta_i \mid \theta_j \ \forall j \neq i) \propto \sqrt{\mathrm{Tr}\left[\left(\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right)^2\right] - \frac{1}{n}\,\mathrm{Tr}\left[\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right]^2}$$

- Conditional posterior : $\pi_i(\theta_i \mid \boldsymbol{y}, \theta_j \ \forall j \neq i) \propto L^1(\boldsymbol{y} \mid \boldsymbol{\theta})\pi_i(\theta_i \mid \theta_j \ \forall j \neq i)$

Gibbs reference posterior

- Proper conditional prior :

$$\pi_i(\theta_i \mid \theta_j \ \forall j \neq i) \propto \sqrt{\mathrm{Tr}\left[\left(\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right)^2\right] - \frac{1}{n}\,\mathrm{Tr}\left[\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right]^2}$$

- Conditional posterior : $\pi_i(\theta_i | \ \boldsymbol{y}, \theta_j \ \forall j \neq i) \propto L^1(\boldsymbol{y} \mid \boldsymbol{\theta})\pi_i(\theta_i \mid \theta_j \ \forall j \neq i)$
- Problem : the conditional posteriors are **incompatible** !

Gibbs reference posterior

- Proper conditional prior :

$$\pi_i(\theta_i \mid \theta_j \; \forall j \neq i) \propto \sqrt{\mathrm{Tr}\left[\left(\frac{\partial}{\partial_{\theta_i}} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right)^2\right] - \frac{1}{n}\,\mathrm{Tr}\left[\frac{\partial}{\partial_{\theta_i}} \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right]^2}$$

- Conditional posterior : $\pi_i(\theta_i | \; \boldsymbol{y}, \theta_j \; \forall j \neq i) \propto L^1(\boldsymbol{y} \mid \boldsymbol{\theta}) \pi_i(\theta_i \mid \theta_j \; \forall j \neq i)$
- Problem : the conditional posteriors are **incompatible** !
- Optimal compromise = **Gibbs reference posterior** $\pi(\boldsymbol{\theta}|\boldsymbol{y})$: the stationary ergodic distribution for the Markov chain given by the Gibbs algorithm applied to the conditional posteriors.

for $k = 1$ to ∞ **do**
  Randomly choose i from the uniform distribution on $[\![1, n]\!]$.
  $\forall j \neq i$ set $\theta_j^{(k)} := \theta_j^{(k-1)}$.
  Sample $\theta_i^{(k)}$ from $\pi_i(\theta_i = \cdot \mid \theta_j^{(k)} \; \forall j \neq i)$.
end for

Gibbs reference posterior

- Proper conditional prior :

$$\pi_i(\theta_i \mid \theta_j \ \forall j \neq i) \propto \sqrt{\mathrm{Tr}\left[\left(\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right)^2\right] - \frac{1}{n}\,\mathrm{Tr}\left[\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right]^2}$$

- Conditional posterior : $\pi_i(\theta_i | \ \boldsymbol{y}, \theta_j \ \forall j \neq i) \propto L^1(\boldsymbol{y} \mid \boldsymbol{\theta})\pi_i(\theta_i \mid \theta_j \ \forall j \neq i)$
- Problem : the conditional posteriors are **incompatible** !
- Optimal compromise = **Gibbs reference posterior** $\pi(\boldsymbol{\theta}|\boldsymbol{y})$: the stationary ergodic distribution for the Markov chain given by the Gibbs algorithm applied to the conditional posteriors.
- Well defined for Matérn anisotropic kernels (with smoothness $\nu \notin \mathbb{N}$).

1: **for** $k = 1$ to ∞ **do**
2: Randomly choose i from the uniform distribution on $[\![1, n]\!]$.
3: $\forall j \neq i$ set $\theta_j^{(k)} := \theta_j^{(k-1)}$.
4: Sample $\theta_i^{(k)}$ from $\pi_i(\theta_i = \cdot \mid \theta_j^{(k)} \ \forall j \neq i)$.
5: **end for**

Existence of the Gibbs reference posterior

Proposition

For a given observation vector $\boldsymbol{y} \in \mathbb{R}^n$, if for any integer $i \in [\![1, r]\!]$, there exists a measurable positive function $m_{i,\boldsymbol{y}}$ on $(0, +\infty)$ such that

$$\forall \boldsymbol{\theta} \in (0, +\infty)^r, \pi_i(\theta_i | \boldsymbol{y}, \boldsymbol{\theta}_{-i}) \geqslant m_{i,\boldsymbol{y}}(\theta_i),$$

then the Gibbs reference posterior $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ (for the given $\boldsymbol{y}$) is well defined.

Existence of the Gibbs reference posterior

Proposition

For a given observation vector $\boldsymbol{y} \in \mathbb{R}^n$*, if for any integer* $i \in [\![1, r]\!]$*, there exists a measurable positive function* $m_{i,\boldsymbol{y}}$ *on* $(0, +\infty)$ *such that*

$$\forall \boldsymbol{\theta} \in (0, +\infty)^r, \pi_i(\theta_i | \boldsymbol{y}, \boldsymbol{\theta}_{-i}) \geqslant m_{i,\boldsymbol{y}}(\theta_i),$$

then the Gibbs reference posterior $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ *(for the given* $\boldsymbol{y}$*) is well defined.*

Definition

A design set $\left(\boldsymbol{x}^{(k)}\right)_{k \in [\![1;n]\!]}$ *is* ***coordinate-distinct*** *if* $\forall i \in [\![1, r]\!]$, $\forall k, k' \in [\![1, n]\!]$, $x_i^{(k)} = x_i^{(k')} \Rightarrow k = k'$.

Existence of the Gibbs reference posterior

Proposition

For a given observation vector $\boldsymbol{y} \in \mathbb{R}^n$, if for any integer $i \in [\![1, r]\!]$, there exists a measurable positive function $m_{i,\boldsymbol{y}}$ on $(0, +\infty)$ such that

$$\forall \boldsymbol{\theta} \in (0, +\infty)^r, \pi_i(\theta_i | \boldsymbol{y}, \boldsymbol{\theta}_{-i}) \geqslant m_{i,\boldsymbol{y}}(\theta_i),$$

then the Gibbs reference posterior $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ (for the given $\boldsymbol{y}$) is well defined.

Definition

*A design set $\left(\boldsymbol{x}^{(k)}\right)_{k \in [\![1;n]\!]}$ is **coordinate-distinct** if $\forall i \in [\![1, r]\!]$, $\forall k, k' \in [\![1, n]\!]$, $x_i^{(k)} = x_i^{(k')} \Rightarrow k = k'$.*

Theorem

For coordinate-distinct design sets $(n > 3r + 1)$ and Matérn anisotropic geometric and tensorized kernels, whatever the true values of the parameters σ_0^2 and $\boldsymbol{\theta}_0$ (with $\nu \notin \mathbb{N}$ known), the observation vector $\boldsymbol{y}$ is almost surely such that the Gibbs reference posterior is well defined.

Sketch of proof

$$\pi_i(\theta_i|\boldsymbol{y},\boldsymbol{\theta}_{-i})^2 \propto \left|\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right| \left(\boldsymbol{y}^\top \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{y}\right)^{-n} \left(\mathrm{Tr}\left[\left(\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right)^2\right] - \frac{1}{n}\mathrm{Tr}\left[\frac{\partial}{\partial_{\theta_i}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\left(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}\right)^{-1}\right]^2\right) \tag{1}$$

Basic idea : $\forall i \in [\![1,r]\!]$, we need to control $\pi_i(\theta_i|\boldsymbol{y},\boldsymbol{\theta}_{-i})$ when

1. $\theta_i \to 0$;
2. $\min\{\theta_j, j \in [\![1,r]\!]\} \to +\infty$.

1. Coordinate-distinct design set $\Rightarrow \boldsymbol{\Sigma}_{\boldsymbol{\theta}} \underset{\theta_i \to 0}{\to} \boldsymbol{I}_n$.
2. Reparametrization $\mu_j := 1/\theta_j$ and continuity argument (TCVD).

Optimal compromise : searching a definition

Given: a set of conditionals $\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})$.

Optimal compromise : searching a definition

Given: a set of conditionals $\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})$.

Complete them with "**hypermarginals**" ν^i: $\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})\nu^i(\boldsymbol{\alpha}_{-i}) =: \pi_i\nu^i(\boldsymbol{\alpha})$.

Optimal compromise : searching a definition

Given: a set of conditionals $\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})$.

Complete them with "**hypermarginals**" ν^i: $\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})\nu^i(\boldsymbol{\alpha}_{-i}) =: \pi_i\nu^i(\boldsymbol{\alpha})$.

If the conditionals are compatible, $\exists \left(\nu^i\right)_{i\in[\![1,r]\!]}$ st $\forall i, j \in [\![1, r]\!]$, $\pi_i\nu^i = \pi_j\nu^j$.

Optimal compromise : searching a definition

Given: a set of conditionals $\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})$.

Complete them with "**hypermarginals**" ν^i: $\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})\nu^i(\boldsymbol{\alpha}_{-i}) =: \pi_i\nu^i(\boldsymbol{\alpha})$.

If the conditionals are compatible, $\exists \left(\nu^i\right)_{i\in[\![1,r]\!]}$ st $\forall i,j \in [\![1,r]\!],\ \pi_i\nu^i = \pi_j\nu^j$.
If not, weaken the notion of compatibility : apply it to the hypermarginals.

Definition (Compatibility of hypermarginals version 1)

*A set of hypermarginals $\left(\nu^i\right)_{i\in[\![1,r]\!]}$ is **compatible** with $(\alpha_i)_{i\in[\![1,r]\!]}$ if*

$$\forall i,j,k \in [\![1,r]\!], \quad \int \pi_i\nu^i(\boldsymbol{\alpha})d\alpha_k = \int \pi_j\nu^j(\boldsymbol{\alpha})d\alpha_k$$

Definition (Compromise)

*A joint distribution is a **compromise** if its hypermarginals are **compatible**.*

Definition (Compatibility of hypermarginals version 1)

A set of hypermarginals $\left(\nu^i\right)_{i \in [\![1,r]\!]}$ *is* ***compatible*** *with* $(\alpha_i)_{i \in [\![1,r]\!]}$ *if*

$$\forall i, j, k \in [\![1, r]\!], \quad \int \pi_i \nu^i(\boldsymbol{\alpha}) d\alpha_k = \int \pi_j \nu^j(\boldsymbol{\alpha}) d\alpha_k$$

Definition (Compromise)

A joint distribution is a ***compromise*** *if its hypermarginals are* ***compatible****.*

Definition (Compatibility of hypermarginals version 1)

A set of hypermarginals $\left(\nu^i\right)_{i\in[\![1,r]\!]}$ *is* ***compatible*** *with* $(\alpha_i)_{i\in[\![1,r]\!]}$ *if*

$$\forall i, j, k \in [\![1, r]\!], \quad \int \pi_i \nu^i(\boldsymbol{\alpha}) d\alpha_k = \int \pi_j \nu^j(\boldsymbol{\alpha}) d\alpha_k$$

Definition (Compromise)

A joint distribution is a ***compromise*** *if its hypermarginals are* ***compatible****.*

This works... if $r = 2$. If $r > 2$, usually no compromise exists !

Definition (Compatibility of hypermarginals version 1)

A set of hypermarginals $\left(\nu^i\right)_{i\in[\![1,r]\!]}$ *is* ***compatible*** *with* $(\alpha_i)_{i\in[\![1,r]\!]}$ *if*

$$\forall i,j,k \in [\![1,r]\!], \quad \int \pi_i \nu^i(\boldsymbol{\alpha}) d\alpha_k = \int \pi_j \nu^j(\boldsymbol{\alpha}) d\alpha_k$$

Definition (Compromise)

A joint distribution is a ***compromise*** *if its hypermarginals are* ***compatible****.*

This works... if $r = 2$. If $r > 2$, usually no compromise exists !

Definition (Compatibility of hypermarginals version 2)

A set of hypermarginals $\left(\nu^i\right)_{i\in[\![1,r]\!]}$ *is* ***compatible*** *with* $(\alpha_i)_{i\in[\![1,r]\!]}$ *if*

$$\forall i,j \in [\![1,r]\!], \quad \nu^i(\boldsymbol{\alpha}_{-i}) = \frac{1}{r}\sum_{j=1}^{r} \int \pi_j \nu^j(\boldsymbol{\alpha}) d\alpha_i$$

Compatibility (version 1) $\Rightarrow$ Compatibility (version 2)

Definition (Compatibility of hypermarginals version 2)

A set of hypermarginals $\left(\nu^i\right)_{i\in[\![1,r]\!]}$ *is* ***compatible*** *with* $(\alpha_i)_{i\in[\![1,r]\!]}$ *if*

$$\forall i,j \in [\![1,r]\!], \quad \nu^i(\boldsymbol{\alpha}_{-i}) = \frac{1}{r}\sum_{j=1}^{r}\int \pi_j \nu^j(\boldsymbol{\alpha}) d\alpha_i$$

Definition (Compromise)

A joint distribution P is a ***compromise*** *if its hypermarginals are* ***compatible****.*

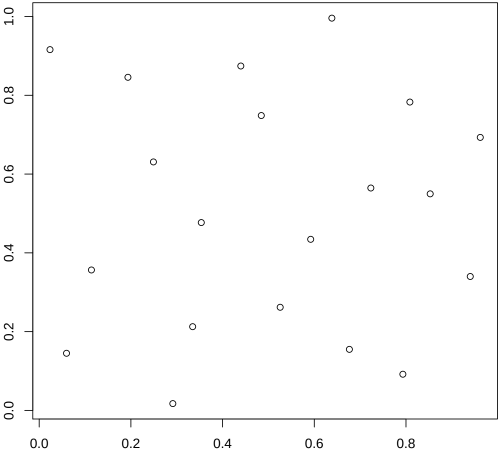
Definition (Optimal compromise)

A compromise is ***optimal*** *if it minimizes, among all compromises, the functional*

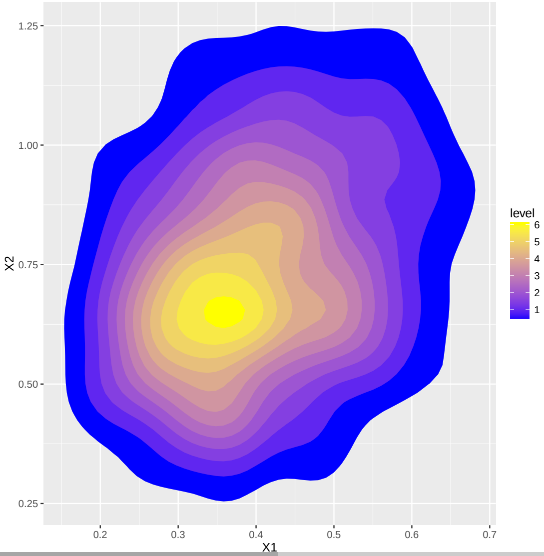
$$E(P) = \sum_{i=1}^{r}\int \left[\pi_i(\alpha_i|\boldsymbol{\alpha}_{-i})P^i(\boldsymbol{\alpha}_{-i}) - P(\boldsymbol{\alpha})\right]^2 d\boldsymbol{\alpha}$$

The invariant distribution of the Gibbs algorithm is the optimal compromise.

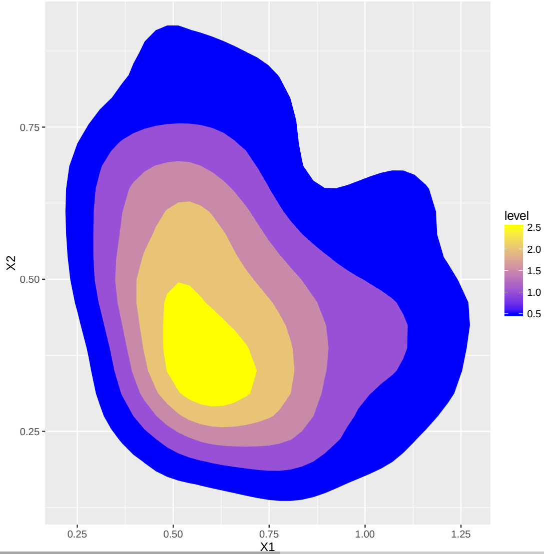
Design set



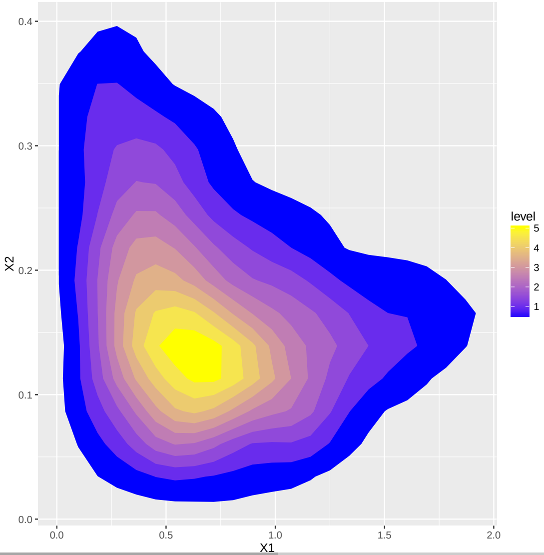
A posteriori distribution on θ (Corr. length 0.3 & 0.6)



A posteriori distribution on θ (Corr. length 0.6 & 0.3)



A posteriori distribution on θ (Corr. length 0.8 & 0.2)



Contents

Distance between sets of correlation lengths

For a given design set, the **distance** between two sets of correlation lengths $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ is the following Frobenius norm :

$$\| \operatorname{argtanh}(\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}) - \operatorname{argtanh}(\boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \|$$

$$\left\| \begin{pmatrix} \dots & \dots & \dots \\ \dots & \operatorname{argtanh}[K_{\hat{\boldsymbol{\theta}}}(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)})] & \dots \\ \dots & \dots & \dots \end{pmatrix} - \begin{pmatrix} \dots & \dots & \dots \\ \dots & \operatorname{argtanh}[K_{\boldsymbol{\theta}}(\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)})] & \dots \\ \dots & \dots & \dots \end{pmatrix} \right\|$$

This distance takes into account errors made when estimating near-1 correlation coefficients as much as errors made when estimating near-0 correlation coefficients.

Robustness gain of the Maximum A Posteriori estimator

RMSE (Root Mean Square Error) of correlation parameter estimators computed over **varying** realizations of the **Gaussian Process** and **randomly drawn** 30-point **design sets** :

Corr. lengths	MLE	MAP	– (%)
0.4 – 0.8 – 0.2	3.49	2.97	15
0.5 – 0.5 – 0.5	4.00	3.46	13
0.7 – 1.3 – 0.4	4.02	3.64	9
0.8 – 0.3 – 0.6	3.75	3.26	13
0.8 – 1.0 – 0.9	4.65	4.18	10

Table: RMSE of the estimators, computed over varying 30-point design sets following the uniform probability distribution and varying realizations of the Gaussian Process with variance 1 and the given set of correlation lengths

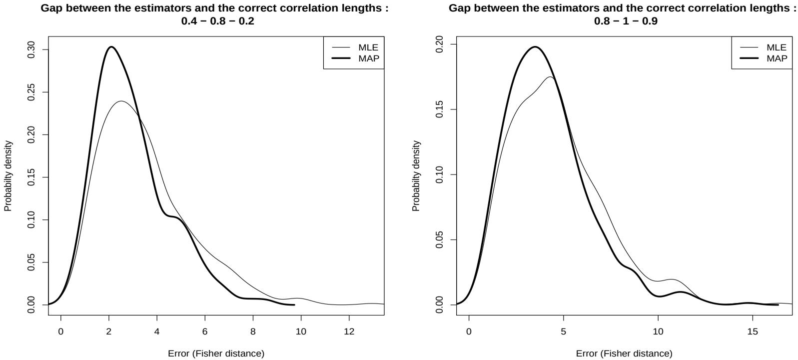


Figure: Estimated Probability density of the error of the MLE and MAP estimators with respect to a 30-point design set following the uniform distribution and a Gaussian Process with variance parameter 1, regularity parameter 5/2 and correlation lengths 0.4 – 0.8 – 0.2 (left) and 0.8 – 1.0 – 0.9 (right).

Prediction intervals

We consider **95% prediction intervals** : the lower bound is the 2.5% quantile and the upper bound the 97.5% quantile of the following predictive distributions :

- when the **true** value of the set of **parameters** is known
- when we assume the **MLE** estimate to be the true value (plug-in)
- when we assume the **MAP** estimate to be the true value (plug-in)
- after averaging over the **posterior distribution** on the set of parameters

We **average** the indicator function of the event "the prediction interval contains the true parameter value" over :

- the **sample space** (*i.e.* all points where prediction is to be performed)
- all realizations of the **Gaussian Process**
- all random uniformly drawn 30-point **design sets**

Corr. lengths	True	MLE	MAP	FPD
0.4 – 0.8 – 0.2	0.95	0.88	0.91	0.95
0.5 – 0.5 – 0.5	0.95	0.89	0.90	0.94
0.7 – 1.3 – 0.4	0.95	0.90	0.92	0.95
0.8 – 0.3 – 0.6	0.95	0.89	0.91	0.95
0.8 – 1.0 – 0.9	0.95	0.90	0.92	0.94

Table: Average across randomly drawn design sets and realizations of the Gaussian Process (with variance parameter 1 and regularity parameter 5/2) of the coverage of 95% Prediction Intervals across the sample space.

Corr. lengths	True	MLE	MAP	FPD
0.4 – 0.8 – 0.2	2.23	2.05 (-8)	2.13 (-4)	2.59 (+16)
0.5 – 0.5 – 0.5	1.69	1.55 (-8)	1.58 (-6)	1.84 (+9)
0.7 – 1.3 – 0.4	1.09	1.02 (-6)	1.07 (-2)	1.21 (+11)
0.8 – 0.3 – 0.6	1.63	1.51 (-7)	1.56 (-4)	1.82 (+12)
0.8 – 1.0 – 0.9	0.71	0.66 (-7)	0.69 (-3)	0.76 (+8)

Table: Average across randomly drawn design sets and GP realizations (with variance parameter 1 and regularity parameter 5/2) of the mean length across the sample space of 95% Prediction Intervals.

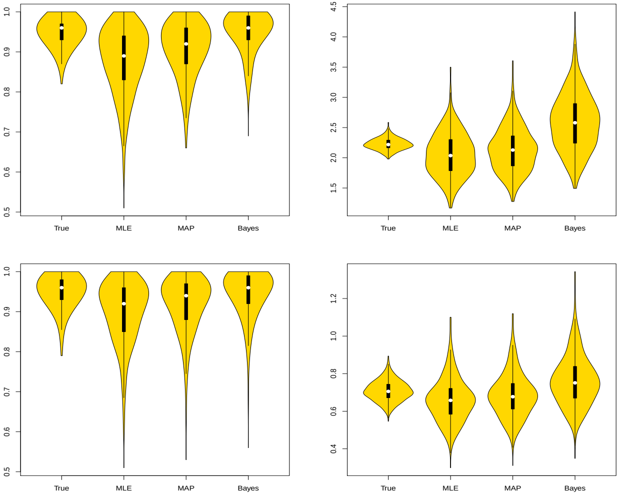


Figure: Coverage (left) and mean length (right) across the sample space of Prediction Intervals for random design sets and GP realizations (variance 1, regularity 5/2, correlation lengths 0.4 – 0.8 – 0.2 (top) and 0.8 – 1.0 – 0.9 (bottom)).

Conclusion

- A proper "objective" distribution on correlation lengths $\boldsymbol{\theta}$ with good frequentist properties.

Conclusion

- A proper "objective" distribution on correlation lengths $\boldsymbol{\theta}$ with good frequentist properties.
- Valid for a large class of correlation kernels.

Conclusion

- A proper "objective" distribution on correlation lengths $\boldsymbol{\theta}$ with good frequentist properties.
- Valid for a large class of correlation kernels.
- Can be extended to the Universal Kriging case.