Gaussian processes with inequality constraints

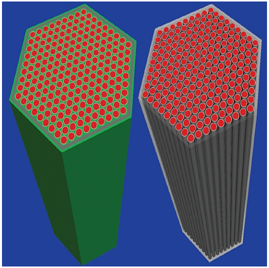
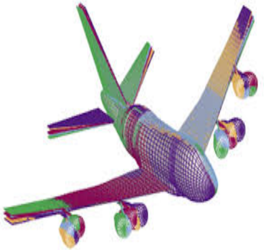
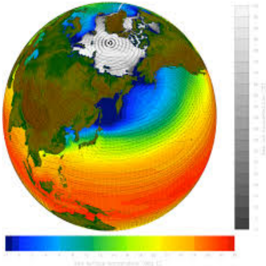
François Bachoc

Institut de Mathématiques de Toulouse
Université Paul Sabatier

Joint work with **Nicolas Durrande** (Prowler), **Andrès Felipe López Lopera** (Mines Saint Etienne) **Agnès Lagnoux** (Institut de Mathématiques de Toulouse) and **Olivier Roustant** (Mines Saint Etienne)

Laboratoire ERIC
February 2019

Computer models have become essential in science and industry !

For clear reasons : cost reduction, possibility to explore hazardous or extreme scenarios...

Computer models as expensive functions

A computer model can be seen as a deterministic function

$$\begin{aligned} f : \mathbb{X} \subset \mathbb{R}^d &\to \mathbb{R} \\ x &\mapsto f(x) \end{aligned}$$

- x : tunable simulation parameter (e.g. geometry)
- $f(x)$: scalar quantity of interest (e.g. energetic efficiency)

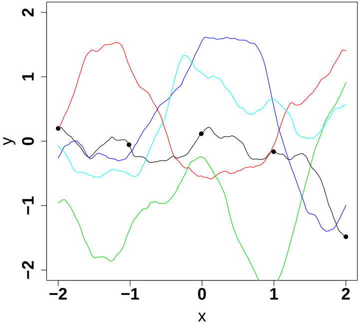
The function f is usually

- continuous (at least)
- non-linear
- only available through evaluations $x \mapsto f(x)$

$\Longrightarrow$ black box model

Gaussian processes (Kriging model)

Modeling the **black box function** as a **single realization** of a Gaussian process $x \to \xi(x)$ on the domain $\mathbb{X} \subset \mathbb{R}^d$



Usefulness

Predicting the continuous realization function, from a finite number of **observation points**

Definition

A stochastic process $\xi : \mathbb{X} \to \mathbb{R}$ is Gaussian if for any $x_1, ..., x_n \in \mathbb{X}$, the vector $(\xi(x_1), ..., \xi(x_n))$ is a Gaussian vector

Mean and covariance functions

The distribution of a Gaussian process is characterized by

- Its mean function : $x \mapsto m(x) = \mathbb{E}(\xi(x))$ Can be any function $\mathbb{X} \to \mathbb{R}$
- Its covariance function $(x_1, x_2) \mapsto k(x_1, x_2) = Cov(\xi(x_1), \xi(x_2))$

The covariance function

- The function $k : \mathbb{X}^2 \to \mathbb{R}$, defined by $k(x_1, x_2) = cov(\xi(x_1), \xi(x_2))$

In most classical cases :

- Stationarity : $k(x_1, x_2) = k(x_1 - x_2)$
- Continuity : $k(x)$ is continuous ' $\Rightarrow'$ Gaussian process realizations are continuous
- Decrease : $k(x)$ decreases with $||x||$ and $\lim_{||x|| \to +\infty} k(x) = 0$

The covariance function

$$k : (x_1, x_2) \to k(x_1, x_2) = cov(\xi(x_1), \xi(x_2))$$

k must me symmetric non-negative definite

$$\forall n \in \mathbb{N}, \forall x_1, ..., x_n \in \mathbb{R}^d, \forall \lambda_1, ..., \lambda_n \in \mathbb{R} : \quad \sum_{i,j=1}^{n} \lambda_i \lambda_j k(x_i, x_j) \geq 0$$

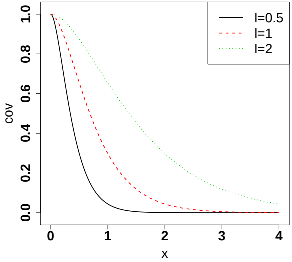
$\Longrightarrow$ the covariance matrix $[k(x_i, x_j)]_{i,j=1,\ldots,n}$ must be non-negative definite
$\Longrightarrow$ Many possibilities on $\mathbb{R}^d$

The Matérn $\frac{3}{2}$ covariance function, for a Gaussian process on $\mathbb{R}$ is parameterized by

- A variance parameter $\sigma^2 > 0$
- A correlation length parameter $\ell > 0$

It is defined as

$$k_{\sigma^2,\ell}(x_1, x_2) = \sigma^2 \left(1 + \sqrt{6}\frac{|x_1 - x_2|}{\ell}\right) e^{-\sqrt{6}\frac{|x_1 - x_2|}{\ell}}$$



Interpretation

- Stationarity, continuity, decrease
- σ^2 corresponds to the order of magnitude of the functions that are realizations of the Gaussian process
- ℓ corresponds to the speed of variation of the functions that are realizations of the Gaussian process

$\Rightarrow$ Natural generalization on $\mathbb{R}^d$

Conditional distribution

Gaussian process ξ observed at $x_1, ..., x_n$

Notation

- $y = (\xi(x_1), ..., \xi(x_n))^\top$
- R is the $n \times n$ matrix $[k(x_i, x_j)]$
- $r(x) = (k(x, x_1), ..., k(x, x_n))^\top$
- $m = (m(x_1), ..., m(x_n))^\top$

Conditional mean

The conditional mean is $m_n(x) := \mathbb{E}(\xi(x)|\xi(x_1), ..., \xi(x_n)) = m(x) + r(x)^\top R^{-1}(y - m)$.

Conditional variance

The conditional variance is
$k_n(x, x) = var(\xi(x)|\xi(x_1), ..., \xi(x_n)) = \mathbb{E}\left[(\xi(x) - m_n(x))^2\right] = k(x, x) - r(x)^\top R^{-1} r(x)$.

Conditional distribution

Conditionally to $\xi(x_1), ..., \xi(x_n)$, ξ is a Gaussian process with (conditional) mean function m_n and (conditional) covariance function $(x, y) \to k_n(x, y) = k(x, y) - r(x)^\top R^{-1} r(y)$

Illustration of conditional mean and variance

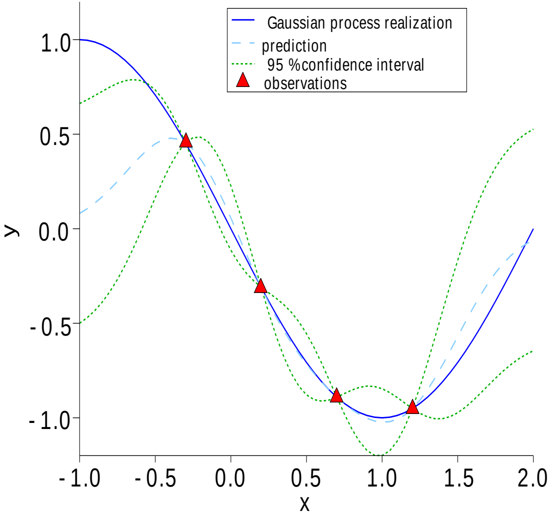
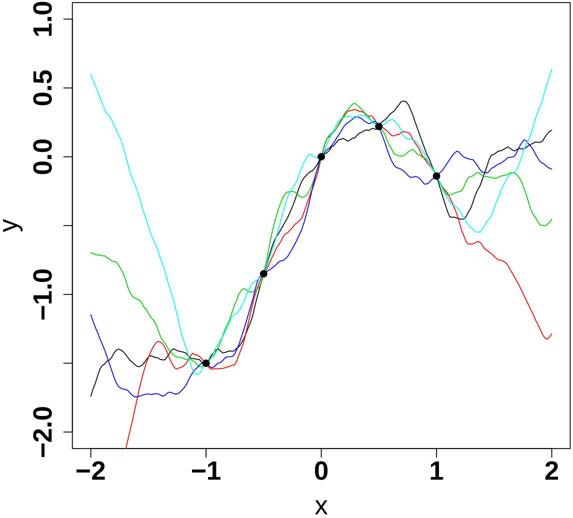


Illustration of the conditional distribution



Gaussian process model for computer experiments

Basic idea : representing the code function $\mathbb{X} \subset \mathbb{R}^d \to \mathbb{R}$ by a realization of a Gaussian process

- Bayesian framework on a fixed function

What we obtain

- Metamodel of the code : the Gaussian process conditional mean function approximates the code function, and its evaluation cost is negligible
- Error indicator with the conditional variance
- Full conditional Gaussian process $\Rightarrow$ possible goal-oriented iterative strategies for optimization, failure domain estimation, probability estimation, code calibration...

Covariance function estimation

- Assume in the rest of the talk that the mean function of ξ is zero
- One needs to select (estimate) a covariance function in order to apply the prediction formulas
- Classically, it is assumed that the covariance function k belongs to a parametric set

Parameterization

Covariance function model $\{k_\theta, \theta \in \Theta\}$ for the Gaussian process ξ

- θ is the multidimensional covariance parameter. k_θ is a covariance function

Observations

ξ is observed at $x_1, ..., x_n \in \mathbb{X}$, yielding the Gaussian vector $y = (\xi(x_1), ..., \xi(x_n))^\top$

Estimation

Objective : build estimator $\hat{\theta}(y)$

Maximum likelihood (ML) for estimation

Explicit Gaussian likelihood function for the observation vector y

Maximum likelihood

Define R_θ as the covariance matrix of $y = (\xi(x_1), ..., \xi(x_n))^\top$ with covariance function k_θ :
$R_\theta = [k_\theta(x_i, x_j)]_{i,j=1,\ldots,n}$.
The maximum likelihood estimator of θ is

$$\hat{\theta}_{ML} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_n(\theta)$$

with

$$\mathcal{L}_n(\theta) = \log(p_\theta(y)) = \log\left(\frac{1}{(2\pi)^{n/2}|R_\theta|} e^{-\frac{1}{2} y^\top R_\theta^{-1} y}\right)$$

$\Rightarrow$ Numerical optimization with $O(n^3)$ criterion
$\Rightarrow$ Most standard estimation method
$\Rightarrow$ Other estimation methods exits : empirical variogram (Book, Cressie), Cross validation (Zhang and Wang 10, Bachoc 13)

A side remark : other applications of Gaussian processes

- In this talk : the Gaussian process ξ is directly observed
  - Main example of application in this talk : computer models

    B. J. Williams, T. J. Santner, and W. I. Notz, The design and analysis of computer experiments, *Springer (2003)*
  - Other applications : machine learning

    C. E. rasmussen and C. K. I. Williams, Gaussian processes for machine learning, *The MIT press (2006)*
- In other strands of research, the Gaussian process ξ is a Bayesian prior over an indirectly observed latent function
  - Gaussian process classification,

$$P(y = 1|x) = e^{\xi(x)}/(1 + e^{\xi(x)}),$$

    Book, Rasmussen & Williams 2006
  - Deep Gaussian processes, e.g. $\xi_2(\xi_1(x))$

    A. Damianou and N. Lawrence, Deep gaussian processes, *AISTATS, Artificial Intelligence and Statistics (2013)*
  - Point processes, e^{ξ} is the spatial intensity function (e.g. epidemiology)

    J. Møller, A. R. Syversveen and R. P. Waagepeterse, Log gaussian cox processes, *Scandinavian journal of statistics, 25(3) 451-482 (1998)*

Inequality constraints

We consider a Gaussian process ξ on $\mathbb{X} = [0,1]^d$ for which we assume that additional information is available :

- $\xi(x)$ belongs to $[\ell, u]$ for $x \in [0,1]^d$ (**boundedness constraints**)
- $\partial/\partial x_i \xi(x) \geq 0$ for $x \in [0,1]^d$ and $i = 1, \ldots, d$ (**monotonicity constraints**)
- ξ is convex on $[0,1]^d$ (**convexity constraints**)
- Modifications and/or combinations of the above constraints

Application cases :

- Computer model output belongs to $\mathbb{R}^+$ (energy) or $[0,1]$ (concentration, energetic efficiency)
- Inputs are known to have positive effects (more input power $\rightarrow$ more output energy)

Impact of the constraints

Generic form of the constraints :

$$\xi \in \mathcal{E}$$

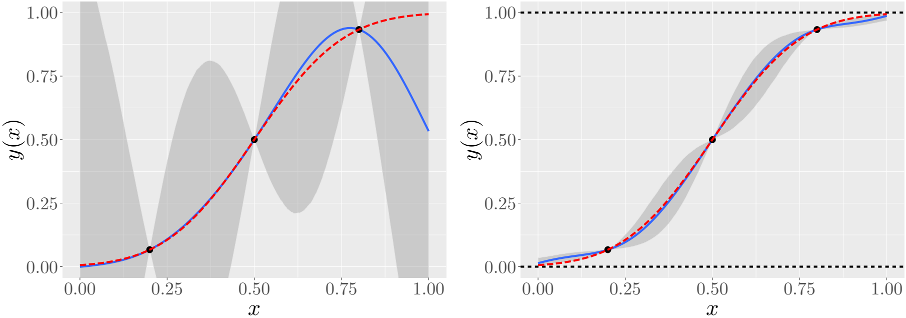
where $\mathcal{E}$ is a set of functions from $[0, 1]^d \to \mathbb{R}$ so that $P(\xi \in \mathcal{E}) > 0$

Impact :

- **New stochastic model :** The law of the realization function is $P(\xi \in .|\xi \in \mathcal{E})$
- **New conditional distribution :** Conditional distribution of ξ given $\xi \in \mathcal{E}$ and $\xi(x_1) = y_1, \dots, \xi(x_n) = y_n$
- **New estimation** of the covariance parameters θ in the covariance model $\{k_\theta; \theta \in \Theta\}$

Illustration of constraint benefits

Target function : bounded and monotonic.



Unconstrained Gaussian process.

Constrained Gaussian process.

- true function
- predictive mean
- training points
- confidence intervals

Existing work

- For boundedness constraints, it is possible to consider models of the form $y_i = T(\xi(x_i))$ with T bijective from $\mathbb{R}$ to $[\ell, u]$ and ξ a Gaussian process
- For monotonicity and convexity constraints, the approach $P(\xi \in .|\xi \in \mathcal{E})$ has become standard
- $\Longrightarrow$ but the constraint $\xi \in \mathcal{E}$ needs to be approximated
- $\xi \in \mathcal{E}$ is replaced by a finite number of constraints on inducing points in

S. Da Veiga and A. Marrel, Gaussian process modeling with inequality constraints, *Annales de la faculté des sciences de Toulouse Mathématiques 21 (2012) 529-555*.

S. Golchi, D. Bingham, H. Chipman and D.A. Campbell, Monotone emulation of computer experiments, *SIAM/ASA Journal on Uncertainty Quantification 3 (2015) 370-392*.

- ξ is replaced by a finite-dimensional approximation ξ_m in

H. Maatouk and X. Bay, Gaussian process emulators for computer experiments with inequality constraints, *Mathematical Geosciences 49(5) (2017) 557-582*.

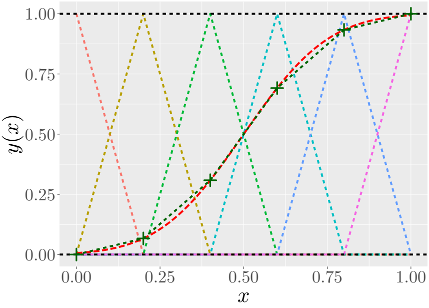
(we follow this latter approach)

Maatouk and Bay 2017 suggest to consider, in dimension $d = 1$,

$$\xi_m(t) = \sum_{j=1}^{m} \epsilon_j \phi_j(t),$$

where

- $\epsilon_j = \xi(t_j)$
- $t_1 = 0, t_2 = 1/(m-1), \dots, t_m = 1$
- the ϕ_j are hat functions, $\phi_j(t) = (1 - (m-1)|t - t_j|)^+$ for $j = 1, \dots, m$



The finite dimensional approximation

Computational benefit (Maatouk and Bay 2017) :

- $\ell \leq \xi_m \leq u \Longleftrightarrow \ell \leq \epsilon_1, \dots, \epsilon_m \leq u$
- ξ_m is a non-decreasing function $\Longleftrightarrow \epsilon_1 \leq \dots \leq \epsilon_m$
- ξ_m is a convex function $\Longleftrightarrow \epsilon_2 - \epsilon_1 \leq \dots \leq \epsilon_m - \epsilon_{m-1}$

$\Longrightarrow$ Only a finite number of inequalities $\Longrightarrow$ guarantee to satisfy the constraints everywhere on $[0,1]$

Extension to dimension 2

$$\xi_m(t_1, t_2) = \sum_{j_1, j_2 = 1}^{m} \epsilon_{j_1} \epsilon_{j_2} \phi_{j_1}(t_1) \phi_{j_2}(t_2)$$

- Becomes problematic in higher dimension
- We are developing other approaches (cf later)

With the finite-dimensional approximation

$$\xi_m(t) = \sum_{j=1}^{m} \epsilon_j \phi_j(t),$$

we study linear constraints of the form

$$\ell \leq \Lambda\epsilon \leq u$$

where

- $\epsilon = (\epsilon_1, \ldots, \epsilon_m)^\top$
- Λ is a $q \times m$ matrix
- ℓ and u are $q \times 1$ vectors
- boundedness, monotonicity, convexity constraints can be enforced, as well as combinations

$\Longrightarrow$ After observed values, the conditional distribution is

$$\mathcal{L}\left(\Lambda\epsilon \middle| \, \Phi\epsilon = y, \ell \leq \Lambda\epsilon \leq u \right),$$

where $\Phi = [\phi_j(x_i)]_{i=1,\ldots,n, j=1,\ldots,m}$ is $n \times m$

Sampling problem

Let M be the covariance matrix of $\epsilon = (\epsilon_1, \dots, \epsilon_m)^\top = (\xi(t_1), \dots, \xi(t_m))^\top$

We have

$$\mathcal{L}\left(\Lambda\epsilon | \Phi\epsilon = y\right) = \mathcal{N}\left(\Lambda M \Phi^\top (\Phi M \Phi^\top)^{-1} y, \Lambda M \Lambda^\top - \Lambda M \Phi^\top (\Phi M \Phi^\top)^{-1} \Phi M \Lambda^\top\right)$$
$$:= \mathcal{N}(\Lambda\mu, \Lambda\Sigma\Lambda^\top)$$

Hence the sampling problem is to sample

$$v \sim \mathcal{N}(\Lambda\mu, \Lambda\Sigma\Lambda^\top),$$

conditionally to $\ell \leq v \leq u$

- We take Λ injective so that $v \implies \epsilon \implies \xi_m$
- Computing $\mathrm{argmax}_{\tilde{v}}\, p_v(\tilde{v} | \ell \leq v \leq u)$ provides the mode
- Computing $\mathbb{E}(v | \ell \leq v \leq u)$ provides the conditional mean
- Sampling v given $\ell \leq v \leq u$ provides conditional samples

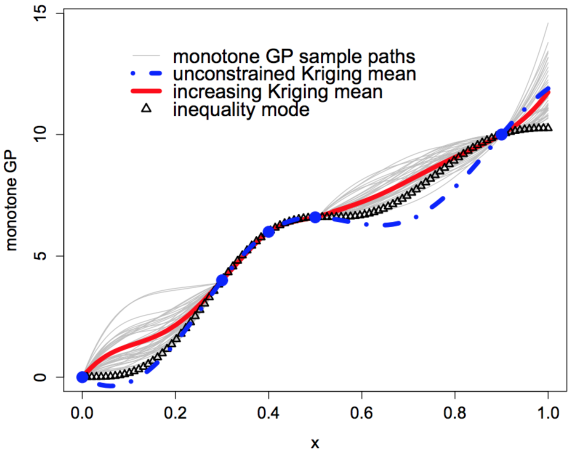


FIGURE: (from Maatouk and Bay 2017) Illustration of conditional samples with constraints (monotone GP sample paths), conditional mean without constraints (unconstrained Kriging mean), conditional mean with monotonicity constraints (increasing Kriging mean) and mode with monotonicity constraints (inequality mode)

Algorithms

The mode is obtained by solving

$$\hat{v} \in \underset{\substack{v \in \mathbb{R}^q \\ \ell \leq v \leq u}}{\operatorname{argmin}} (v - \Lambda\mu)^\top (\Lambda\Sigma\Lambda^\top)^{-1} (v - \Lambda\mu)$$

- quadratic function optimization subject to linear inequality constraints
- quite fast algorithms
- corresponds to the (unconstrained) conditional mean $\Lambda\mu$ if it satisfies the inequality constraints

Sampling $v \sim \mathcal{N}(\Lambda\mu, \Lambda\Sigma\Lambda^\top)$ subject to $\ell \leq v \leq u$:

- rejection sampling from the mode Maatouk and Bay 2017 (low acceptance rate for q large)

We investigate

- Hastings metropolis
- Gibbs sampling (never rejects) Taylor and Benjamini 2017
- Minimax tilting Botev 2017 JRSSB
- Hamiltonian Monte Carlo Pakman and Paninski 2014 JCGS

and conclude that Hamiltonian Monte Carlo is an efficient sampler in our framework

An application to nuclear engineering

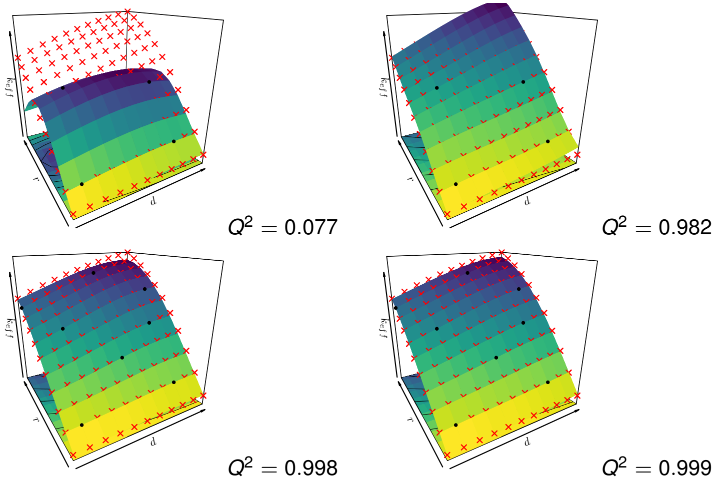


FIGURE: Two dimensional nuclear engineering example. **radius** and **density** of uranium sphere $\Longrightarrow$ **criticality coefficient**. Mononicity constraints. Left : unconstrained Gaussian process models. Right : constrained Gaussian process models. The Q^2 measures the prediction quality and should be close to 1.

The paper

A. F. López-Lopera, F. Bachoc, N. Durrande and O. Roustant, Finite-dimensional Gaussian approximation with linear inequality constraints, *SIAM/ASA Journal on Uncertainty Quantification, forthcoming*.

Adaptation to higher dimension

- In dimension $d \geq 5$, say, we can not use the full grid approach
- We aim for a representation

$$\xi_m = \text{function}(\epsilon_1, \ldots, \epsilon_m)$$

so that we keep

$$\xi_m \in \mathcal{E} \Longleftrightarrow (\epsilon_1, \ldots, \epsilon_m) \in \mathcal{C}$$

- Approach 1 : additive Gaussian processes

$$\xi_m(x_1, \ldots, x_d) = \sum_{i=1}^{d} \xi_{m,i}(x_i) + \sum_{\substack{i,j=1,\ldots,d \\ i \neq j}} \xi_{m,i,j}(x_i, x_j)$$

with grids in dimensions 1 and 2.

- Approach 2 : Tensorized grid with less grid points for less important variables

Setting

- For simplicity, let us forget about the finite-dimensional approximation ξ_m (but see the papers)
- We observe the Gaussian process ξ at $x_1, \dots, x_n \in [0,1]^d$ and let $y = (\xi(x_1), \dots, \xi(x_n))^\top$
- We assume that ξ has covariance function k
- We consider the model of covariance functions $\{k_\theta ; \theta \in \Theta\}$
- The inequality constraints are $\xi \in \mathcal{E}$

The maximum likelihood estimator of θ is

$$\hat{\theta}_{ML} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_n(\theta)$$

with

$$\mathcal{L}_n(\theta) = \log(p_\theta(y)) = \log \left(\frac{1}{(2\pi)^{n/2} |R_\theta|} e^{-\frac{1}{2} y^\top R_\theta^{-1} y} \right)$$

- (it ignores the information $\xi \in \mathcal{E}$)
- explicit expression of $\mathcal{L}_n$ with $O(n^3)$ cost

Constrained maximum likelihood estimator

The constrained maximum likelihood estimator of θ is

$$\hat{\theta}_{cML} \in \underset{\theta \in \Theta}{\operatorname{argmax}} \, \mathcal{L}_{\mathcal{C},n}(\theta)$$

with

$$\mathcal{L}_{\mathcal{C},n}(\theta) = \log(p_\theta(y)) - \log(p_\theta(\xi \in \mathcal{E})) + \log(p_\theta(\xi \in \mathcal{E} | y))$$

- The additional terms $\log(p_\theta(\xi \in \mathcal{E}))$ and $\log(p_\theta(\xi \in \mathcal{E} | y))$ have no explicit expressions
- They need to be approximated by numerical integration or Monte Carlo : Genz 1992 JCGS, Botev 2017 JRSSB

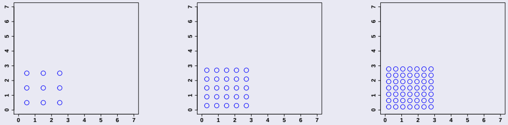
$\Longrightarrow$ We aim at comparing $\hat{\theta}_{ML}$ and $\hat{\theta}_{cML}$ asymptotically

Two asymptotic frameworks for covariance parameter estimation

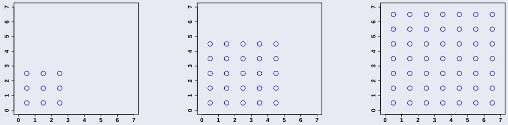
- Asymptotics (number of observations $n \to +\infty$) is an active area of research
- (case without constraints so far)
- There are several asymptotic frameworks because there are several possible location patterns for the observation points

Two main asymptotic frameworks

- fixed-domain asymptotics : The observation points are dense in a bounded domain



- increasing-domain asymptotics : number of observation points is proportional to domain volume $\longrightarrow$ unbounded observation domain.



Existing increasing-domain asymptotic results

- Consistent estimation is possible for all covariance parameters (that are identifiable in finite-sample). [asymptotic independence between observations]
- Asymptotic normality proved for maximum likelihood

Mardia K, Marshall R, Maximum likelihood estimation of models for residual covariance in spatial regression, *Biometrika 71 (1984) 135-146*.

N. Cressie and S.N Lahiri, The asymptotic distribution of REML estimators, *Journal of Multivariate Analysis 45 (1993) 217-233*.

N. Cressie and S.N Lahiri, Asymptotics for REML estimation of spatial covariance parameters, *Journal of Statistical Planning and Inference 50 (1996) 327-341*.

F. Bachoc, Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes, *Journal of Multivariate Analysis 125 (2014) 1-35*.

- Consistent estimation is impossible for some covariance parameters (identifiable in finite-sample), see e.g.

  Zhang, H., Inconsistent Estimation and Asymptotically Equivalent Interpolations in Model-Based Geostatistics, *Journal of the American Statistical Association (99), 250-261, 2004*.

  Stein M, Interpolation of Spatial Data : Some Theory for Kriging, *Springer, New York, 1999*.

  - covariance parameters that can not be estimated consistently are called non-microergodic
  - covariance parameters that can be estimated consistently are called microergodic

- For instance, consider the set of covariance functions $\{k_\theta, \theta \in (0,\infty)^2\}$ on $[0,1]$ given by $\theta = (\sigma^2, \alpha)$ and $k_\theta(t_1, t_2) = \sigma^2 e^{-\alpha|t_1 - t_2|}$
  - σ^2 is non-microergodic
  - α is non-microergodic
  - $\sigma^2\alpha$ is microergodic

$\Longrightarrow$ We address fixed-domain asymptotics here

Preservation of consistency

Setting :

- ξ is a Gaussian process on $[0,1]^d$, $d \in \mathbb{N}$, with mean zero and covariance function k
- $\theta = (\sigma^2, \alpha_1, \ldots, \alpha_d)$
- k_θ is the covariance function of the Gaussian process $(x_1, \ldots, x_d) \to \sigma^2 \xi(\alpha_1 x_1, \ldots, \alpha_d x_d)$

$\Longrightarrow$ $k = k_{\theta_0}$ with $\theta_0 = (1, \ldots, 1)$

- The constraints are given by the set $\mathcal{E}$ and are **boundedness**, **monotonicity** or **convexity**
- $(x_i)_{i \in \mathbb{N}}$ is dense in $[0,1]^d$

Proposition : preservation of consistency for ML (López-Lopera, Bachoc, Durrande, Roustant 2018)

Assume that the covariance function k satisfy technical conditions (see papers). Assume $\forall \varepsilon > 0$,

$$P(\|\widehat{\theta}_{ML} - \theta_0\| \geq \varepsilon) \xrightarrow[n \to \infty]{} 0 \quad \text{(unconditional consistency of ML)}$$

Then, we have $P(\xi \in \mathcal{E}) > 0$, and thus

$$P(\|\widehat{\theta} - \theta_0\| \geq \varepsilon \mid \xi \in \mathcal{E}) \xrightarrow[n \to \infty]{} 0 \quad \text{(conditional consistency of ML)}$$

Proposition : preservation of consistency for cML (López-Lopera, Bachoc, Durrande, Roustant 2018)

Assume that the covariance function k satisfy technical conditions (see papers). Assume that $\forall \varepsilon > 0$ and $\forall M < \infty$, (sufficient condition for unconditional consistency of ML)

$$P\left(\sup_{\|\theta - \theta_0\| \geq \varepsilon} (\mathcal{L}_n(\theta) - \mathcal{L}_n(\theta_0)) \geq -M \right) \xrightarrow[n \to \infty]{} 0$$

Then, (sufficient condition for conditional consistency of cML)

$$P\left(\sup_{\|\theta - \theta_0\| \geq \varepsilon} (\mathcal{L}_{\mathcal{C},n}(\theta) - \mathcal{L}_{\mathcal{C},n}(\theta_0)) \geq -M \,\middle|\, \xi \in \mathcal{E} \right) \xrightarrow[n \to \infty]{} 0$$

Consequently (conditional consistency of ML and cML)

$$\hat{\theta}_{ML} \xrightarrow[n \to \infty]{P|\xi \in \mathcal{E}} \theta_0 \quad \text{and} \quad \hat{\theta}_{cML} \xrightarrow[n \to \infty]{P|\xi \in \mathcal{E}} \theta_0$$

Asymptotic normality result 1 : variance estimation

Setting :

- Gaussian process ξ on $[0,1]^d$, $d \in \mathbb{N}$, with zero mean function and covariance function k
- Monotonicity, boundedness or convexity constraints (as before)
- $(x_i)_{i \in \mathbb{N}}$ is dense in $[0,1]^d$
- $\theta = \sigma^2$ and $k_\theta(u_1, u_2) = \sigma^2 k(u_1, u_2)$

Known results

- It is well-known that in this case

$$\sqrt{n}\left(\hat{\sigma}_{ML}^2 - \sigma_0^2\right) \to_{n\to\infty}^{\mathcal{L}} N(0, 2\sigma_0^4)$$

Asymptotic normality result 1 : variance estimation

Notation : we write $X_n \to_{n\to\infty}^{\mathcal{L}|\xi\in\mathcal{E}} L$ when for all bounded measurable function f :

$$\mathbb{E}(f(X_n)|\xi \in \mathcal{E}) \to_{n\to\infty} \int f(x)dL(x)$$

Theorem (Bachoc, Lagnoux, López-Lopera 2018)

Under technical conditions on k and the sequence $(x_i)_{i\in\mathbb{N}}$ (see papers), we have

$$\sqrt{n}\left(\hat{\sigma}^2_{ML} - \sigma_0^2\right) \to_{n\to\infty}^{\mathcal{L}|\xi\in\mathcal{E}} N(0, 2\sigma_0^4)$$

and

$$\sqrt{n}\left(\hat{\sigma}^2_{cML} - \sigma_0^2\right) \to_{n\to\infty}^{\mathcal{L}|\xi\in\mathcal{E}} N(0, 2\sigma_0^4)$$

- Same asymptotic distribution as the (unconstrained) maximum likelihood estimator, in the unconstrained case
- No asymptotic impact of the constraints

Asymptotic normality result 2 : Matérn model

Setting :

- Gaussian process ξ on $[0,1]^d$, $d = 1, 2, 3$, with zero mean function and covariance function k
- Monotonicity, boundedness or convexity constraints (as before)
- $(x_i)_{i \in \mathbb{N}}$ is dense in $[0,1]^d$
- $\theta = (\sigma^2, \rho) \in (0, \infty)^2$ and

$$k_{\theta,\nu}(x, x') = \sigma^2 K_\nu \left(\frac{||x - x'||}{\rho} \right) = \frac{\sigma^2}{\Gamma(\nu) 2^{\nu-1}} \left(\frac{||x - x'||}{\rho} \right)^\nu \kappa_\nu \left(\frac{||x - x'||}{\rho} \right).$$

  - Γ is the Gamma function
  - κ_ν is the modified Bessel function of the second kind
  - $\nu > 0$ (assumed known) is the smoothness parameter : $\nu > r \iff$ corresponding Gaussian process if r times differentiable

In this case :

- σ^2 is non-microergodic
- ρ is non-microergodic
- $\sigma^2 / \rho^{2\nu}$ is microergodic and

$$\sqrt{n} \left(\frac{\widehat{\sigma}^2_{ML}}{\widehat{\rho}^{2\nu}_{ML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}} \right) \xrightarrow[n \to +\infty]{\mathcal{L}} \mathcal{N} \left(0, 2 \left(\frac{\sigma_0^2}{\rho_0^{2\nu}} \right)^2 \right).$$

C. G. Kaufman and B. A. Shaby, The Role of the Range Parameter for Estimation and Prediction in Geostatistics, *Biometrika 100 (2013) 473–484*.

Asymptotic normality result 2 : Matérn model

We show

Theorem (Bachoc, Lagnoux, López-Lopera 2018)

Under technical conditions on ν and the sequence $(x_i)_{i \in \mathbb{N}}$ (see papers), we have

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2_{ML}}{\widehat{\rho}^{2\nu}_{ML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}} \right) \xrightarrow[n \to +\infty]{\mathcal{L} | \xi \in \mathcal{E}} \mathcal{N}\left(0, 2 \left(\frac{\sigma_0^2}{\rho_0^{2\nu}} \right)^2 \right)$$

and

$$\sqrt{n}\left(\frac{\widehat{\sigma}^2_{cML}}{\widehat{\rho}^{2\nu}_{cML}} - \frac{\sigma_0^2}{\rho_0^{2\nu}} \right) \xrightarrow[n \to +\infty]{\mathcal{L} | \xi \in \mathcal{E}} \mathcal{N}\left(0, 2 \left(\frac{\sigma_0^2}{\rho_0^{2\nu}} \right)^2 \right)$$

- Same conclusions as for the estimation of a variance parameter

An illustration

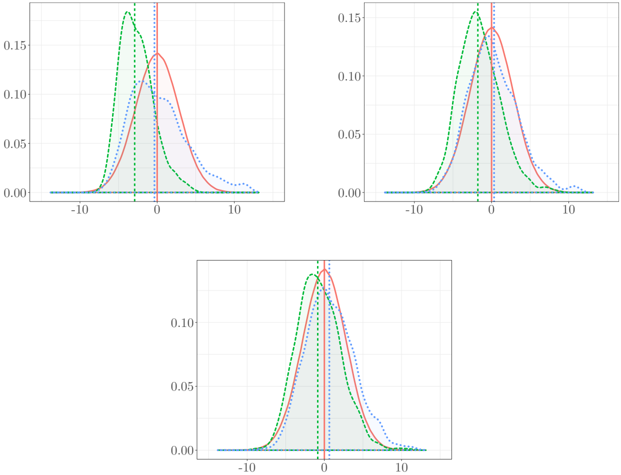


FIGURE: An example with the estimation of σ^2 with boundedness constraints. Distribution of $n^{1/2}(\hat{\sigma}^2 - \sigma_0^2)$. $n = 20$ (top left), $n = 50$ (top right) and $n = 80$ (bottom). Green : ML. Blue : cML. Red : Gaussian limit

The papers

For consistency :

A. F. López-Lopera, F. Bachoc, N. Durrande and O. Roustant, Finite-dimensional Gaussian approximation with linear inequality constraints, *SIAM/ASA Journal on Uncertainty Quantification, forthcoming*.

For asymptotic normality :

F. Bachoc, Agnès Lagnoux and A. F. López-Lopera, Maximum likelihood estimation for Gaussian processes under inequality constraints, `arxiv.org/abs/1804.03378`.

Conclusion

Summary

- Gaussian processes provide a Bayesian framework on deterministic functions (e.g. computer models)
- Inequality constraints correspond to additional information (e. g. physical knowledge)
- Taking them into account can significantly improve the predictions
- with a computational cost (explicit $\Longrightarrow$ Monte Carlo)
- The constrained maximum likelihood estimator (cML) has similar consistency guarantees as maximum likelihood (ML)
- Asymptotically, we do not see an impact of the constraints and ML $\approx$ cML
- For small sample size, cML appears to be beneficial

Ongoing work

- The finite-dimensional approach in higher dimension

Thank you for your attention !