Predicting data with a geometrical structure : Random fields indexed by metric spaces and Kriging estimation

Nil Venet

CEA Tech Occitanie, Institut de Mathématiques de Toulouse

10 January 2017

Plan of the talk

Data with a geometrical structure

1. Data may come with a geometrical structure
   - Spatial data: we get real-valued data $(x_{p_1}, \cdots, x_{p_n})$ with the p_i in a space with a distance ($\mathbb{R}^n$, the sphere, a graph...).

Data with a geometrical structure

1. Data may come with a geometrical structure
   - Spatial data: we get real-valued data $(x_{p_1}, \cdots, x_{p_n})$ with the p_i in a space with a distance ($\mathbb{R}^n$, the sphere, a graph...).
2. ... or we can chose to put an adequate geometrical structure on it
   - Imagery: Assume our data are cancer risk scores $(S_{I_1}, \cdots, S_{I_n})$ of brain images I_i. We need a distance between two images.

Data with a geometrical structure

1. Data may come with a geometrical structure
   - Spatial data: we get real-valued data $(x_{p_1}, \cdots, x_{p_n})$ with the p_i in a space with a distance ($\mathbb{R}^n$, the sphere, a graph...).
2. ... or we can chose to put an adequate geometrical structure on it
   - Imagery: Assume our data are cancer risk scores $(S_{I_1}, \cdots, S_{I_n})$ of brain images I_i. We need a distance between two images.
   - Functional data: Data may be scores of functions/distributions. Again we need to choose a distance between functions/distributions...

Data with a geometrical structure

1. Data may come with a geometrical structure
   - Spatial data: we get real-valued data $(x_{p_1}, \cdots, x_{p_n})$ with the p_i in a space with a distance ($\mathbb{R}^n$, the sphere, a graph...).
2. ... or we can chose to put an adequate geometrical structure on it
   - Imagery: Assume our data are cancer risk scores $(S_{I_1}, \cdots, S_{I_n})$ of brain images I_i. We need a distance between two images.
   - Functional data: Data may be scores of functions/distributions. Again we need to choose a distance between functions/distributions...

- We expect data x_p and x_q to be as strongly correlated as the distance $d(p, q)$ is small.

Data with a geometrical structure

1. Data may come with a geometrical structure
   - Spatial data: we get real-valued data $(x_{p_1}, \cdots, x_{p_n})$ with the p_i in a space with a distance ($\mathbb{R}^n$, the sphere, a graph...).
2. ... or we can chose to put an adequate geometrical structure on it
   - Imagery: Assume our data are cancer risk scores $(S_{I_1}, \cdots, S_{I_n})$ of brain images I_i. We need a distance between two images.
   - Functional data: Data may be scores of functions/distributions. Again we need to choose a distance between functions/distributions...

- We expect data x_p and x_q to be as strongly correlated as the distance $d(p, q)$ is small.
- We need random models that respect that geometrical structure.

Random fields as models for our data

Typically,

We have a space E endowed with a distance d.

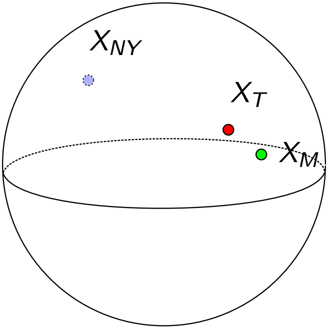

Figure: A familiar metric space : the Earth endowed with the geodesic distance

Random fields as models for our data

Typically,

We have a space E endowed with a distance d and we want a collection of random variables $(X_P)_{P \in E}$ such that for two points P and Q in E, the two random variables X_P and X_Q are as decorrelated as $d(P, Q)$ is large.

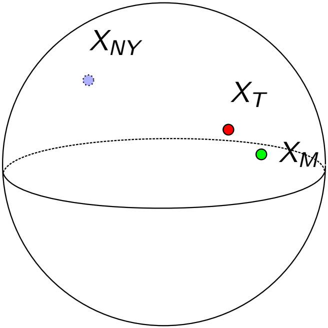

Figure: A familiar metric space : the Earth endowed with the geodesic distance

The notion of metric space is very general

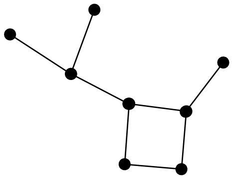

Figure: A graph

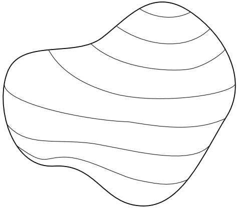

Figure: A dented sphere

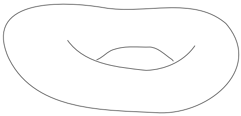

Figure: A surface with a hole

One can also think of a functional or an image space, but they are infinite dimensional and uneasy to draw.

Gaussian random fields

For practical reasons we will assume that our random fields $(X_P)_{P \in E}$ are *Gaussian*. Indeed:

A very nice property

The statistical properties of a Gaussian field $(X_P)P \in E$ depend only on its

- mean function $P \mapsto \mathbb{E}(X_P)$ and
- covariance function $(P, Q) \mapsto \mathbb{E}(X_P - \mathbb{E}(X_P))(X_Q - \mathbb{E}(X_Q))$.

Without loss of generality we will assume that $\mathbb{E}(X_P) = 0$ for every $P \in E$.

Other enjoyable properties we may ask for

Stationarity

We say that $(X_P)_{P\in E}$ is *stationary* if

$$\mathbb{E}(X_P X_Q) = f(d(P,Q)).$$

The statistical properties of (X_P) don't depend on where we are.

Stationarity, independence of the increments

- The statistical properties of the variations of the random field between $X_P - X_Q$ depend only on the distance $d(P,Q)$.
- One can also ask that two different increments be independent (Lévy processes).

The most typical example here is the Brownian motion.

Fractional Brownian motions/fields

Given a parameter H in $[0,1]$, consider the covariance

$$\mathbb{E}(X_P X_Q) = \frac{1}{2}\left(d^{2H}(O,P) + d^{2H}(O,Q) - d^{2H}(P,Q)\right).$$

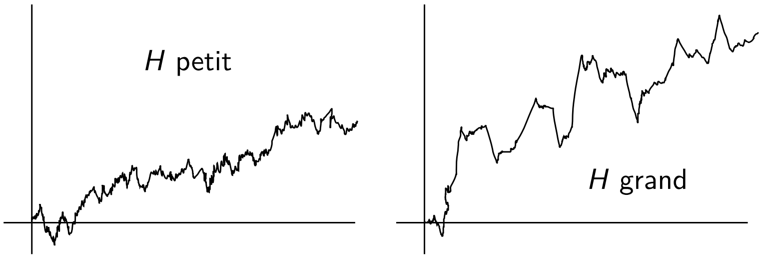

Figure: Two sample paths

The existence problem

Problem

Given a function K of two variables in E, there does not always exist a random process with covariance $\mathbb{E}(X_P X_Q) = K(P, Q)$.

Gaussian case

In order for such a Gaussian process to exist it is necessary and sufficient that K be a *positive definite kernel*, that is to say for every $P_1 \cdots, P_n \in E$ and $\lambda_1, \cdots, \lambda_n \in \mathbb{R}$,

$$\sum_{i,j=1}^{n} \lambda_i \lambda_j K(P_i, P_j) \geq 0.$$

Positive definite kernels are also crucial for *Support Vector Machines* methods.

Existence Issues for the fractional Brownian motion, 1

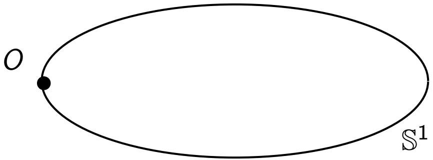

Figure: Fractional brownian field indexed by the circle

Existence Issues for the fractional Brownian motion, 1

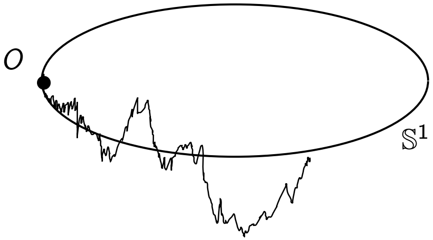

Figure: Fractional brownian field indexed by the circle

Existence Issues for the fractional Brownian motion, 1

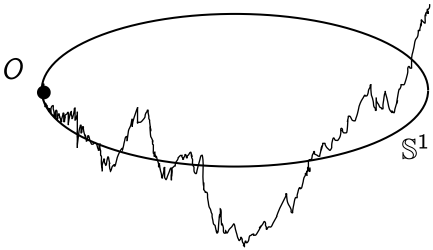

Figure: Fractional brownian field indexed by the circle

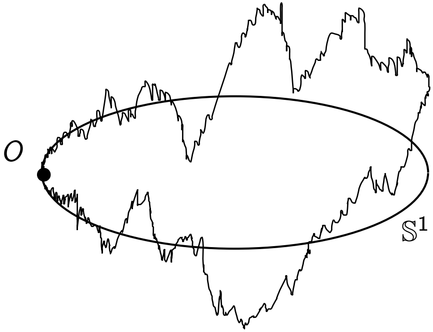

Figure: Fractional brownian field indexed by the circle

Existence Issues for the fractional Brownian motion, 1

The fractional Brownian field indexed by the circle exists if and only if

$$0 < H \leq \frac{1}{2}.$$

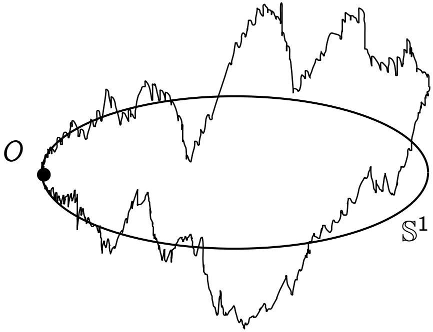

Figure: Fractional brownian field indexed by the circle

Existence Issues for the fractional Brownian motion, 1

The fractional Brownian field indexed by the circle exists if and only if

$$0 < H \leq \frac{1}{2}.$$

My PhD

The problematic of my PhD was to understand for which H the fractional Brownian field exists, for other metric spaces. It is a broad question that goes back to Paul Lévy (1960).

Existence Issues for the fractional Brownian motion, 2

I have showed that it is very unlikely to have existence of the Brownian field when there is a circle (minimal closed geodesic) in the metric space.

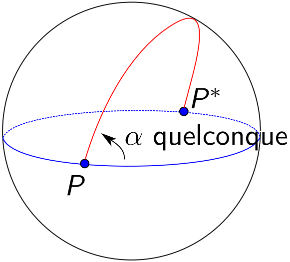

Figure: A sphere is OK

Existence Issues for the fractional Brownian motion, 2

I have showed that it is very unlikely to have existence of the Brownian field when there is a circle (minimal closed geodesic) in the metric space.

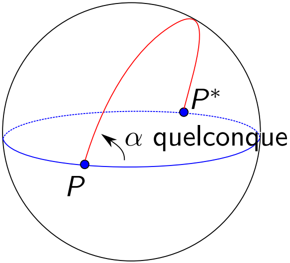

Figure: A sphere is OK

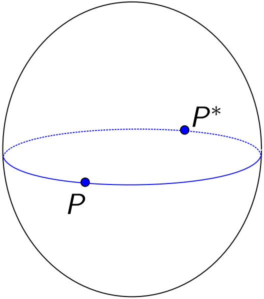

Figure: An ellipsoid is not

Existence Issues for the fractional Brownian motion, 3

I have showed that it is very unlikely to have existence of the Brownian field when there is a circle (minimal closed geodesic) in the metric space.

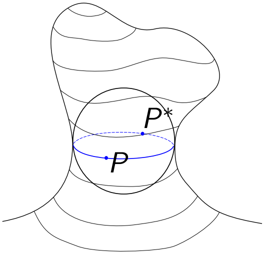

Figure: Not OK

I have showed that it is very unlikely to have existence of the Brownian field when there is a circle (minimal closed geodesic) in the metric space.

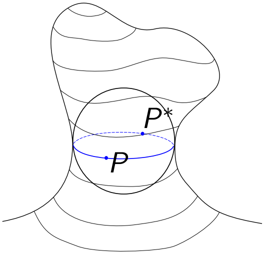

Figure: Not OK

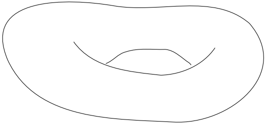

Figure: Still not OK

What are we looking for ?

We want a Gaussian random field in order to do some *Kriging estimation*

What are we looking for ?

We want a Gaussian random field in order to do some *Kriging estimation* (extra short introduction):

What are we looking for ?

We want a Gaussian random field in order to do some *Kriging estimation* (extra short introduction):

- We need a valid, nondegenerate covariance kernel.

What are we looking for ?

We want a Gaussian random field in order to do some *Kriging estimation* (extra short introduction):

- We need a valid, nondegenerate covariance kernel.
- We want noise.

What are we looking for ?

We want a Gaussian random field in order to do some *Kriging estimation* (extra short introduction):

- We need a valid, nondegenerate covariance kernel.
- We want noise.
- We want correlation.

What are we looking for ?

We want a Gaussian random field in order to do some *Kriging estimation* (extra short introduction):

- We need a valid, nondegenerate covariance kernel.
- We want noise.
- We want correlation.
- A stationarity property eases our life.

What are we looking for ?

We want a Gaussian random field in order to do some *Kriging estimation* (extra short introduction):

- We need a valid, nondegenerate covariance kernel.
- We want noise.
- We want correlation.
- A stationarity property eases our life.
- A family of covariances is even better.

Strategies for existence proofs

1. Analytic proof of the positive definiteness of a covariance kernel.

Strategies for existence proofs

1. Analytic proof of the positive definiteness of a covariance kernel.
   - for stationary random fields, harmonic analysis may be used.

Strategies for existence proofs

1. Analytic proof of the positive definiteness of a covariance kernel.
    - for stationary random fields, harmonic analysis may be used.
    - in order to do that we need a “symmetric" space.

Strategies for existence proofs

1. Analytic proof of the positive definiteness of a covariance kernel.
   - for stationary random fields, harmonic analysis may be used.
   - in order to do that we need a “symmetric" space.

2. Direct constructions through integration of Gaussian white noise.

Strategies for existence proofs

1. Analytic proof of the positive definiteness of a covariance kernel.
   - for stationary random fields, harmonic analysis may be used.
   - in order to do that we need a “symmetric" space.
2. Direct constructions through integration of Gaussian white noise.
   - again it seems that we need some kind of homogeneity of the space to obtain stationary random fields.

Plan of the talk

1. Random fields indexed by metric spaces

2. Existence questions

3. Results on data with distribution inputs

The Wasserstein space of probability distributions

Consider the space $\mathcal{W}$ of probability distributions μ on the real line $\mathbb{R}$ with a second order moment, that is to say

$$\int_{\mathbb{R}} x^2 d\mu < \infty.$$

The Wasserstein space of probability distributions

Consider the space $\mathcal{W}$ of probability distributions μ on the real line $\mathbb{R}$ with a second order moment, that is to say

$$\int_{\mathbb{R}} x^2 d\mu < \infty.$$

Definition

The *Wasserstein distance* between $\mu, \nu \in \mathcal{W}$ is

$$d(\mu,\nu) = \inf_{(X,Y)\in\Pi(X,Y)} \left(\mathbb{E}(X-Y)^2\right)^{1/2},$$

where $\Pi(X,Y)$ is the set of all random vectors (X,Y) such that $X \sim \mu$ and $Y \sim \nu$.

Results (with F. bachoc, F. Gamboa, and J.M. Loubes)

Stationary random fields

The kernels

$$F(d(\mu, \nu))$$

are valid covariances for a large class of functions, including $e^{-td^{2H}(\mu,\nu)}$ for $t > 0$ and $H \in [0, 1]$. Hence we have the existence of Gaussian stationary random fields $(X_\mu)_{\mu \in \mathcal{W}}$.

Results (with F. bachoc, F. Gamboa, and J.M. Loubes)

Stationary random fields

The kernels

$$F(d(\mu, \nu))$$

are valid covariances for a large class of functions, including $e^{-td^{2H}(\mu,\nu)}$ for $t > 0$ and $H \in [0, 1]$. Hence we have the existence of Gaussian stationary random fields $(X_\mu)_{\mu \in \mathcal{W}}$.

Fractional Brownian fields

The fractional Brownian field $(X_\mu^H)_{\mu \in \mathcal{W}}$ exists if and only if $0 \leq H \leq 1$.

Results (with F. bachoc, F. Gamboa, and J.M. Loubes)

Stationary random fields

The kernels

$$F(d(\mu, \nu))$$

are valid covariances for a large class of functions, including $e^{-td^{2H}(\mu,\nu)}$ for $t > 0$ and $H \in [0, 1]$. Hence we have the existence of Gaussian stationary random fields $(X_\mu)_{\mu \in \mathcal{W}}$.

Fractional Brownian fields

The fractional Brownian field $(X_\mu^H)_{\mu \in \mathcal{W}}$ exists if and only if $0 \leq H \leq 1$. We have nondegeneracy for these kernels.

Results (with F. bachoc, F. Gamboa, and J.M. Loubes)

Given some scored distributions $(\mu_i, s_i)_{i=1}^n$,

Results (with F. bachoc, F. Gamboa, and J.M. Loubes)

Given some scored distributions $(\mu_i, s_i)_{i=1}^n$,

- we looked at the question of the choice of the best stationary covariance $F(d(\mu,\nu))$, and showed the consistency and the normal asymptoticity of the *maximum-likelihood* estimator in a parametric model.

Results (with F. bachoc, F. Gamboa, and J.M. Loubes)

Given some scored distributions $(\mu_i, s_i)_{i=1}^n$,

- we looked at the question of the choice of the best stationary covariance $F(d(\mu,\nu))$, and showed the consistency and the normal asymptoticity of the *maximum-likelihood* estimator in a parametric model.
- we proved the consistency of the *Kriging estimator* under the estimated covariance.

Results (with F. bachoc, F. Gamboa, and J.M. Loubes)

Given some scored distributions $(\mu_i, s_i)_{i=1}^n$,

- we looked at the question of the choice of the best stationary covariance $F(d(\mu,\nu))$, and showed the consistency and the normal asymptoticity of the *maximum-likelihood* estimator in a parametric model.
- we proved the consistency of the *Kriging estimator* under the estimated covariance.
- on simulated data the method provides significant improvements compared to classical functional methods.