Regularity estimation in multivariate functional data analysis

Omar KASSI & Nicolas KLUTCHNIKOFF & Valentin PATILEA

November 14, 2023

ENSAI & Rennes

Based on joint work

Valentin Patilea (CREST Ensai)

Nicolas Klutchnikoff (Univ. Rennes)

Introduction

- New types of data
    - financial data
    - energy data
    - sensors (cars, airplanes,...)
    - medical devices (cardiograms, fMRI, blood pressure, oxygen or glucose devices,...)
    - environmental devices (daily temperature, wind speed, solar radiation, pollution levels,...)
    - sports data

- The observation unit (entity), the datum, could be one or several curves, image(s), or several such objects

- *Related fields* : Signal Processing, Longitudinal Data...

- Ideally, one would keep the observation unit as it was collected
  - model data as random realizations in a suitable space
- Data are (in)dependent realizations of some variable

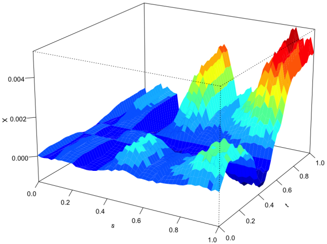
$$X : (\Omega, \mathcal{A}) \to (\mathcal{X}, \mathcal{F})$$

- When $\mathcal{X}$ is a space of curves/images/signals
  - Functional Data problem
- Functional Data Analysis (FDA) deals with the statistical description and modeling of samples of random variable taking values in spaces of functions

Multivariate functional data

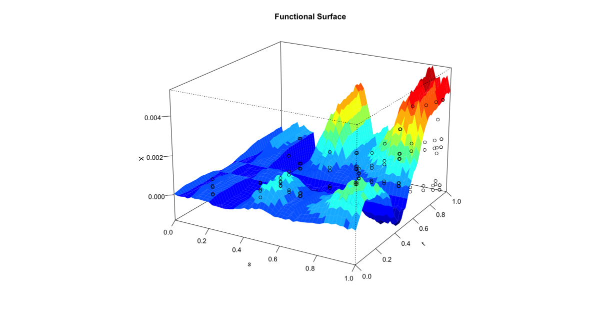
- The realizations of the stochastic process X are surfaces

  - Satellite images
  - Measurements of temperature or salinity in oceanology



Data

- Ideally, data represent the continuous time measurements of sample paths of same stochastic process
- Real data are
    - discretely observed, possibly at random points, which may be sparsely distributed
    - noisy measurements



Data

- $\mathcal{T}$: open, bounded bi-dimensional rectangle, $\overline{\mathcal{T}} \subset (0, \infty)^2$
- $X^{(1)}, \ldots, X^{(j)}, \ldots, X^{(N)}$ are independent realizations of X
- The data associated to a sample path $X^{(j)}$ consist of the pairs $(Y_m^{(j)}, \boldsymbol{t}_m^{(j)}) \in \mathbb{R} \times \mathcal{T}$, where for $1 \leq j \leq N$ and $1 \leq m \leq M_j$

$$Y_m^{(j)} = X^{(j)}(\boldsymbol{t}_m^{(j)}) + \varepsilon_m^{(j)}, \quad \text{with} \quad \varepsilon_m^{(j)} = \sigma(\boldsymbol{t}_m^{(j)}, X^{(j)}(\boldsymbol{t}_m^{(j)}))e_m^{(j)}$$

- $M_1, \ldots, M_N$ be an independent sample of an integer-valued random variable M, $\mathbb{E}[M] = \mathfrak{m}$
- The $\left(\boldsymbol{t}_m^{(j)}, \quad 1 \leq m \leq M_j\right)$ represent the observation points for the sample path $X^{(j)}$.

Methodology

First steps : univariate case (1/2)

- For B^H a fBm with Hurst index $H \in (0,1)$,

$$\mathbb{E}\left[\left\{B^H(t) - B^H(s)\right\}^2\right] = |t-s|^{2H}, \quad s,t \in \mathbb{R}_+$$

- Estimating equation for the Hurst parameter :

$$H = \frac{\log\left(\mathbb{E}\left[\left\{B^H(t) - B^H(s)\right\}^2\right]\right)}{2\log|t-s|}$$

First steps : univariate case (2/2)

- Let X be a process defined on a subset of $\mathbb{R}$, with non-differentiable sample paths
- GKP (2022) : $H(t_0) \in (0,1)$ and $L(t_0) > 0$ exist such that

$$\mathbb{E}\left[\{X(t) - X(s)\}^2\right] \approx L(t_0)^2 |t - s|^{2H(t_0)}, \quad \forall s \leq t_0 \leq t$$

  for t and s close to t_0

- Estimating equation :

$$H(t_0) \approx \frac{\log(\theta(t_1, t_2)) - \log(\theta(t_1, t_3))}{2\log(2)}, \qquad t_0 \in [t_1, t_2] \subset [t_1, t_3]$$

  where

$$\theta(t, s) = \mathbb{E}\left[\{X(t) - X(s)\}^2\right] \quad \text{and} \quad |t_1 - t_2| = 2|t_1 - t_3|.$$

Multivariate case : notation

- $H_1, H_2 : \mathcal{T} \to (0,1)$ are continuously differentiable functions. Let

$$\overline{H} = \max\{H_1, H_2\}$$

- $L_1^{(1)}, L_2^{(1)}, L_1^{(2)}, L_2^{(2)}$: Non negative Lipschitz continuous functions defined on $\mathcal{T}$ such that

$$L_k^{(1)}(\boldsymbol{t}) + L_k^{(2)}(\boldsymbol{t}) > 0, \qquad \forall \boldsymbol{t} \in \mathcal{T} \subset \mathbb{R}^2,\ k = 1, 2.$$

- For $X \in \mathcal{L}^2$, we denote for sufficiently small scalars Δ

$$\theta_{\boldsymbol{t}}^{(i)}(\Delta) = \mathbb{E}\left[\left\{X\left(\boldsymbol{t} + \frac{\Delta}{2}e_i\right) - X\left(\boldsymbol{t} - \frac{\Delta}{2}e_i\right)\right\}^2\right], \quad i = 1, 2,$$

where (e_1, e_2) is canonical basis of $\mathbb{R}^2$

A class of multivariate processes

Definition

We say $X \in \mathcal{H}^{H_1,H_2}(\boldsymbol{L}, \mathcal{T})$ if three constants $\Delta_0, C, \beta > 0$ exist such that for any $\boldsymbol{t} \in \mathcal{T}$ and $0 < \Delta \leq \Delta_0$,

$$\left|\theta_{\boldsymbol{t}}^{(i)}(\Delta) - L_1^{(i)}(\boldsymbol{t})\Delta^{2H_1(\boldsymbol{t})} - L_2^{(i)}(\boldsymbol{t})\Delta^{2H_2(\boldsymbol{t})}\right| \leq C\Delta^{2\overline{H}(\boldsymbol{t})+\beta}, \quad i = 1, 2.$$

Let

$$\mathcal{H}^{H_1,H_2} = \bigcup_{\boldsymbol{L}} \mathcal{H}^{H_1,H_2}(\boldsymbol{L}, \mathcal{T}),$$

where $\boldsymbol{L} = (L_1^{(1)}, L_2^{(1)}, L_1^{(2)}, L_2^{(2)})$.

The functions H_1, H_2 define the local regularity of the process, while $\boldsymbol{L}$ represent the local Hölder constants.

Example : Sum of two fractional Brownian motion

- Let $B_1^{H_1}$ and $B_2^{H_2}$ be two independent fBm with Hurst index H_1 and H_2.
- Let

$$X_1(\boldsymbol{t}) = B_1^{H_1}(t_1) + B_2^{H_2}(t_2), \quad \forall \boldsymbol{t} = (t_1, t_2) \in \mathbb{R}^2.$$

Then $X_1 \in \mathcal{H}^{H_1,H_2}$ where $\boldsymbol{L} = (1, 0, 0, 1)$.

- Let $\beta > 0$ and define

$$X_2(\boldsymbol{t}) = X_1\left(\begin{pmatrix} \cos\beta & \sin\beta \\ -\sin\beta & \cos\beta \end{pmatrix} \boldsymbol{t}\right), \qquad \forall \boldsymbol{t} \in \mathbb{R}^2.$$

Then $X_2 \in \mathcal{H}^{H_1,H_2}$ with

$$\boldsymbol{L} = (|\cos\beta|^{2H_1}, |\sin\beta|^{2H_2}, |\sin\beta|^{2H_1}, |\cos\beta|^{2H_2}).$$

Example : multifractional Brownian sheet

- Let $\boldsymbol{\eta} = (\eta_1, \eta_2) : [0, \infty)^2 \to (0,1)^2$ be a deterministic map
- The multifractional Brownian (MfB) sheet W with Hurst functional parameter $\boldsymbol{\eta}$ is defined as :

$$W(\boldsymbol{u}) = \left(\prod_{k=1}^{2} \frac{1}{C(\eta_k(\boldsymbol{u}))} \right) \int_{\mathbb{R}^2} \prod_{k=1}^{2} \frac{e^{iu_k\zeta_k} - 1}{|\zeta_k|^{\eta_k(\boldsymbol{u})+\frac{1}{2}}} \widehat{\boldsymbol{B}}(\mathrm{d}\boldsymbol{\zeta}), \quad \boldsymbol{u} \in (0,\infty)^2,$$

where

- $\widehat{\boldsymbol{B}}$ is the FT of the white noise in $\mathbb{R}^2$
- $$C(x) = \left[\frac{2\pi}{\Gamma(2x+1)\sin(\pi x)} \right]^{1/2}$$

- The process W is a centered Gaussian process
- The covariance function

$$\mathbb{E}[W(\boldsymbol{u})W(\boldsymbol{v})] = \prod_{i=1,2} D(\eta_i(\boldsymbol{u}), \eta_i(\boldsymbol{v})) \\ \times \left[u_i^{\eta_i(\boldsymbol{u})+\eta_i(\boldsymbol{v})} + v_i^{\eta_i(\boldsymbol{u})+\eta_i(\boldsymbol{v})} - |u_i - v_i|^{\eta_i(\boldsymbol{u})+\eta_i(\boldsymbol{v})}\right],$$

where

$$D(x, y) = C^2((x+y)/2) \cdot (2C(x)C(y))^{-1},$$

- With respect to our Definition, $W \in \mathcal{H}^{H_1,H_2}$ with $(H_1, H_2) = (\eta_1, \eta_2)$ and

$$(L_1^{(1)}(\boldsymbol{t}), L_2^{(1)}(\boldsymbol{t}), L_1^{(2)}(\boldsymbol{t}), L_2^{(2)}(\boldsymbol{t})) = (t_2^{2H_2(\boldsymbol{t})}, 0, 0, t_1^{2H_1(\boldsymbol{t})})$$

Identification issues

- Let $H_1, H_2, \tilde{H}_1$ and $\tilde{H}_2$ be some continuously differentiable functions taking values in $(0, 1)$
- Assume $X \in \mathcal{H}^{H_1,H_2}$ and $X \in \mathcal{H}^{\tilde{H}_1,\tilde{H}_2}$
- We then necessarily have

$$\min\{H_1(\boldsymbol{t}), H_2(\boldsymbol{t})\} = \min\{\tilde{H}_1(\boldsymbol{t}), \tilde{H}_2(\boldsymbol{t})\}$$

  and

$$\max\{H_1(\boldsymbol{t}), H_2(\boldsymbol{t})\} = \max\{\tilde{H}_1(\boldsymbol{t}), \tilde{H}_2(\boldsymbol{t})\}$$

- Notation :

$$\underline{H}(\boldsymbol{t}) = \min\{H_1(\boldsymbol{t}), H_2(\boldsymbol{t})\}, \quad \overline{H}(\boldsymbol{t}) = \max\{H_1(\boldsymbol{t}), H_2(\boldsymbol{t})\}.$$

Estimating equations for $\underline{H}$ and $\overline{H}$

- Recall

$$\theta_{\boldsymbol{t}}^{(i)}(\Delta) = \mathbb{E}\left[\left\{X\left(\boldsymbol{t} - \Delta e_i/2\right) - X\left(\boldsymbol{t} + \Delta e_i/2\right)\right\}^2\right], \quad i = 1, 2,$$

- Denote for any $\boldsymbol{t} \in \mathcal{T}$

$$\gamma_{\boldsymbol{t}}(\Delta) = \theta_{\boldsymbol{t}}^{(1)}(\Delta) + \theta_{\boldsymbol{t}}^{(2)}(\Delta)$$

- Then

$$\underline{H}(\boldsymbol{t}) \approx \frac{\log(\gamma_{\boldsymbol{t}}(2\Delta)) - \log(\gamma_{\boldsymbol{t}}(\Delta))}{2\log(2)}.$$

- Let

$$\alpha_{\boldsymbol{t}}(\Delta) = \left|\frac{\gamma_{\boldsymbol{t}}(2\Delta)}{(2\Delta)^{2\underline{H}(\boldsymbol{t})}} - \frac{\gamma_{\boldsymbol{t}}(\Delta)}{\Delta^{2\underline{H}(\boldsymbol{t})}}\right|.$$

- Then

$$\overline{H}(\boldsymbol{t}) - \underline{H}(\boldsymbol{t}) \approx \frac{\log(\alpha_{\boldsymbol{t}}(2\Delta)) - \log(\alpha_{\boldsymbol{t}}(\Delta))}{2\log(2)}.$$

Estimators for $\underline{H}$ and $\overline{H}$: presmoothing

- In general, the sheets $X^{(j)}$, $j \in \{1, \ldots, N\}$, are not available
- Let $\widetilde{X}^{(j)}$ be an observable approximation of $X^{(j)}$.
    - If X is observed everywhere and without noise, then

    $$\widetilde{X}^{(j)}(\boldsymbol{t}) = X^{(j)}(\boldsymbol{t}), \quad \forall \boldsymbol{t} \in \mathcal{T}$$

    - If X is observed with noise or/and on a discrete grid, then $\widetilde{X}^{(j)}$ is an estimator of $X^{(j)}$ (local polynomial, splines, interpolation...)

Estimators for $\underline{H}$ and $\overline{H}$ (1/2)

- The observable approximation allows to build estimates :

$$\widehat{\theta}_{\boldsymbol{t}}^{(i)}(\Delta) = \frac{1}{N}\sum_{j=1}^{N}\left\{\widetilde{X}^{(j)}(\boldsymbol{t}-(\Delta/2)e_i) - \widetilde{X}^{(j)}(\boldsymbol{t}+(\Delta/2)e_i)\right\}^2,$$

$$\widehat{\gamma}_{\boldsymbol{t}}(\Delta) = \widehat{\theta}_{\boldsymbol{t}}^{(1)}(\Delta) + \widehat{\theta}_{\boldsymbol{t}}^{(2)}(\Delta).$$

- The first estimator follows :

$$\widehat{\underline{H}}(\boldsymbol{t}) = \begin{cases} \frac{\log(\widehat{\gamma}_{\boldsymbol{t}}(2\Delta)) - \log(\widehat{\gamma}_{\boldsymbol{t}}(\Delta))}{2\log(2)} & \text{if} \quad \widehat{\gamma}_{\boldsymbol{t}}(2\Delta), \widehat{\gamma}_{\boldsymbol{t}}(\Delta) > 0 \\ 1 & \text{otherwise} \end{cases}.$$

Estimators for $\underline{H}$ and $\overline{H}$ (2/2)

- Moreover

$$\widehat{\alpha}_{\boldsymbol{t}}(\Delta) = \begin{cases} \left| \dfrac{\widehat{\gamma}_{\boldsymbol{t}}(2\Delta)}{(2\Delta)^{2\widehat{\underline{H}}(\boldsymbol{t})}} - \dfrac{\widehat{\gamma}_{\boldsymbol{t}}(\Delta)}{\Delta^{2\widehat{\underline{H}}(\boldsymbol{t})}} \right| & \text{if} \quad \dfrac{\widehat{\gamma}_{\boldsymbol{t}}(2\Delta)}{(2\Delta)^{2\widehat{\underline{H}}(\boldsymbol{t})}} \neq \dfrac{\widehat{\gamma}_{\boldsymbol{t}}(\Delta)}{\Delta^{2\widehat{\underline{H}}(\boldsymbol{t})}} \\ 1 & \text{otherwise} \end{cases} .$$

- Hence

$$(\widehat{\overline{H} - \underline{H}})(\boldsymbol{t}) = \frac{\log(\widehat{\alpha}_{\boldsymbol{t}}(2\Delta)) - \log(\widehat{\alpha}_{\boldsymbol{t}}(\Delta))}{2\log(2)}$$

- We then set

$$A_N(\tau) = \left\{ (\widehat{\overline{H} - \underline{H}})(\boldsymbol{t}) \geq \tau \right\},$$

and define

$$\widehat{\overline{H}}(\boldsymbol{t}) = \widehat{\underline{H}}(\boldsymbol{t}) + (\widehat{\overline{H} - \underline{H}})(\boldsymbol{t}) 1_{A_N(\tau)}.$$

Estimating equations for $L_1^{(i)}(\boldsymbol{t})$ and $L_2^{(i)}(\boldsymbol{t})$

- Recall

$$\theta_{\boldsymbol{t}}^{(i)}(\Delta) \approx L_1^{(i)}(\boldsymbol{t})\Delta^{2H_1(\boldsymbol{t})} + L_2^{(i)}(\boldsymbol{t})\Delta^{2H_2(\boldsymbol{t})}, \quad i = 1, 2$$

- Assume

$$H_1(\boldsymbol{t}) = \underline{H}(\boldsymbol{t}) < \overline{H}(\boldsymbol{t}) = H_2(\boldsymbol{t})$$

- For $i = 1, 2$,

$$L_1^{(i)}(\boldsymbol{t}) \approx \frac{\theta_{\boldsymbol{t}}^{(i)}(\Delta)}{\Delta^{2H_1(\boldsymbol{t})}}$$

and

$$L_2^{(i)}(\boldsymbol{t}) \approx \frac{1}{(4^{D(\boldsymbol{t})} - 1)\Delta^{2D(\boldsymbol{t})}} \left| \frac{\theta_{\boldsymbol{t}}^{(i)}(2\Delta)}{(2\Delta)^{2H_1(\boldsymbol{t})}} - \frac{\theta_{\boldsymbol{t}}^{(i)}(\Delta)}{\Delta^{2H_1(\boldsymbol{t})}} \right|$$

with $D(\boldsymbol{t}) = H_2(\boldsymbol{t}) - H_1(\boldsymbol{t})$

Estimators for $L_1^{(i)}(t)$ and $L_2^{(i)}(t)$

- Plug into the estimating equations for $L_j^{(i)}(\boldsymbol{t})$ the estimators of the unknown quantities, as defined above
- Special attention requires the case $\underline{H}(\boldsymbol{t}) = \overline{H}(\boldsymbol{t})$
    - A diagnostic tool is provided

Non-asymptotic results

Non-asymptotic results

- Let

$$R_p(\mathfrak{m}) = \sup_{\boldsymbol{t}\in\mathcal{T}} \mathbb{E}[|\xi^{(j)}(\boldsymbol{t})|^p], \qquad \xi^{(j)}(\boldsymbol{t}) = \widetilde{X}^{(j)}(\boldsymbol{t}) - X^{(j)}(\boldsymbol{t})$$

- Assumptions
  1. $X \in \mathcal{H}^{H_1,H_2}$, and the realizations of X are independent
  2. Constants $\mathfrak{a}$, $\mathfrak{A}$ and r exist such that, for any $\boldsymbol{t} \in \mathcal{T}$,

$$\mathbb{E}\left|X^{(j)}(\boldsymbol{t}) - X^{(j)}(\boldsymbol{s})\right|^{2p} \leq \frac{p!}{2}\mathfrak{a}\mathfrak{A}^{p-2}\|\mathrm{t}-\mathrm{s}\|^{2p\underline{H}(\mathrm{t})}, \quad \forall \boldsymbol{s} \in B(\boldsymbol{t}; r), \; \forall p \geq 1$$

  3. Constants $\mathfrak{c}$ and $\mathfrak{D}$, and a function $\rho(\mathfrak{m}) \leq 1$, exist such that

$$R_{2p}(\mathfrak{m}) \leq \frac{p!}{2}\mathfrak{c}\mathfrak{D}^{p-2}\rho(\mathfrak{m})^{2p}, \qquad \forall p \geq 1, \; \forall \mathfrak{m} > 1$$

  4. Two positive constants $\mathfrak{L}$ and ν exist such that

$$R_2(\mathfrak{m}) \leq \mathfrak{L}\mathfrak{m}^{-\nu}, \qquad \forall \mathfrak{m} > 1$$

Non-asymptotic results

- We denote

$$R(\underline{H})(\boldsymbol{t}) = \underline{H}(\boldsymbol{t}) - \frac{\log(\gamma_{\boldsymbol{t}}(2\Delta)) - \log(\gamma_{\boldsymbol{t}}(\Delta))}{2\log(2)},$$

$$R(\overline{H} - \underline{H})(\boldsymbol{t}) = \{\overline{H} - \underline{H}\}(\boldsymbol{t}) - \frac{\log(\alpha_{\boldsymbol{t}}(2\Delta)) - \log(\alpha_{\boldsymbol{t}}(\Delta))}{2\log(2)},$$

and

$$R(L_1^{(i)})(\boldsymbol{t}) = L_1^{(i)}(\boldsymbol{t}) - \frac{\theta_{\boldsymbol{t}}^{(i)}(\Delta)}{\Delta^{2H_1(\boldsymbol{t})}}, \quad i = 1, 2.$$

Non-asymptotic results

Proposition 1 : Constants $C_1, \ldots, C_5$ exist such that,

$$\forall \varepsilon, \tau \in (0,1) \quad \max\{|\log(\Delta)||R(\underline{H})(\boldsymbol{t})|,\ |R(\overline{H}-\underline{H})(\boldsymbol{t})|\} \leq \varepsilon \leq 2\tau,$$

$$\mathbb{P}\left[|\widehat{\underline{H}}(\boldsymbol{t}) - \underline{H}(\boldsymbol{t})| \geq \varepsilon\right] \leq p_1,$$

and

$$\mathbb{P}\left[\left|\widehat{\overline{H}}(\boldsymbol{t}) - \overline{H}(\boldsymbol{t})\right| \geq \varepsilon\right] \leq C_3\{p_1 + p_2 + p_3\},$$

with

- $p_1 = C_1 \exp\left(-C_2 N \times \varepsilon^2 \times \Delta^{4\underline{H}(\boldsymbol{t})} \varrho(\Delta, \mathfrak{m})\right),$
- $p_2 = \exp\left[-C_4 N \times \varepsilon^2 \times \frac{\Delta^{4\overline{H}(\boldsymbol{t})}\varrho(\Delta,\mathfrak{m})}{\log^2(\Delta)} \Delta^{4D(\boldsymbol{t})}\right] 1_{\{\underline{H}(\boldsymbol{t})<\overline{H}(\boldsymbol{t})\}},$
- $p_3 = \exp\left[-C_5 N \times \tau^2 \times \frac{\Delta^{4\overline{H}(\boldsymbol{t})}\varrho(\Delta,\mathfrak{m})}{\log^2(\Delta)} \Delta^{4D(\boldsymbol{t})}\right],$

where $\varrho(\Delta, \mathfrak{m}) = \max\{\Delta^{2\underline{H}(\boldsymbol{t})}, \rho(\mathfrak{m})^2\}^{-1}.$

Non-asymptotic results

Proposition 2 : Constants $\mathfrak{C}_1, \ldots, \mathfrak{C}_4$ exists such that, for $i = 1, 2$, and for any $\varepsilon \in (0, 1)$ such that

$$\max\left\{|R(L_1^{(i)})(\boldsymbol{t})|,\ |\log(\Delta)||R(\underline{H})(\boldsymbol{t})|,\ |R(\overline{H} - \underline{H})(\boldsymbol{t})|\right\} \leq \varepsilon,$$

$$\mathbb{P}\left(\left|\widehat{L_1^{(i)}}(\boldsymbol{t}) - L_1^{(i)}(\boldsymbol{t})\right| \geq \varepsilon\right) \leq \mathfrak{C}_1 \exp\left(-\mathfrak{C}_2 N \times \varepsilon^2 \times \frac{\Delta^{4\underline{H}(\boldsymbol{t})} \varrho(\Delta, \mathfrak{m})}{\log^2(\Delta)}\right),$$

and

$$\mathbb{P}\left(\left|\widehat{L_2^{(i)}}(\boldsymbol{t}) - L_2^{(i)}(\boldsymbol{t})\right| \geq \varepsilon\right)$$
$$\leq \mathfrak{C}_3 \exp\Bigg(-\mathfrak{C}_4 N \times \varepsilon \Delta^{4D(\boldsymbol{t})} \min\{\varepsilon, \Delta^{4D(\boldsymbol{t})}\}$$
$$\times \frac{\Delta^{4\overline{H}(\boldsymbol{t})} \varrho(\Delta, \mathfrak{m})}{\log^4(\Delta)} \times (4^{D(\boldsymbol{t})} - 1)^2\Bigg).$$

A risk bound for the anisotropy detection

Proposition 3 : Let

$$A_N(\tau) = \left\{ (\widehat{\overline{H} - \underline{H}})(\boldsymbol{t}) \geq \tau \right\}.$$

If

$$\max\{|\log(\Delta)||R(\underline{H})(\boldsymbol{t})|,\ |R(\overline{H} - \underline{H})(\boldsymbol{t})|\} \leq 2\tau,$$

and

$$2\tau \leq \left\{ \overline{H}(\boldsymbol{t}) - \underline{H}(\boldsymbol{t}) \right\} + 1_{\{\underline{H}(\boldsymbol{t}) = \overline{H}(\boldsymbol{t})\}}.$$

Then

$$\mathbb{P}\left(1_{A_N(\tau)} \neq 1_{\{\underline{H}(\boldsymbol{t}) < \overline{H}(\boldsymbol{t})\}}\right) \leq C_3 \exp\left[-C_5 N \times \tau^2 \times \frac{\Delta^{4\overline{H}(\boldsymbol{t})} \varrho(\Delta, \mathfrak{m})}{\log^2(\Delta)} \Delta^{4D(\boldsymbol{t})} \right],$$

where C_3 and C_5 are the positive constants from Proposition 1.

Applications

MfB sheet with domain deformation (1/2)

- Let W be a MfB sheet with Hurst functional parameter $\boldsymbol{\eta}$
- Let $A = (A_1, A_2)$ be a continuously differentiable deformation of a domain in the plane, satisfying some mild conditions
- Let

$$X = W \circ A$$

- Then for $\boldsymbol{t}, \boldsymbol{s} \in \mathcal{T}$, we have

$$\begin{aligned}\theta(\boldsymbol{t}, \boldsymbol{s}) &= \mathbb{E}\left[\{X(\boldsymbol{t}) - X(\boldsymbol{s})\}^2\right] \\ &\approx |A_1(\boldsymbol{t})|^{2H_1(\boldsymbol{t})} |\partial_1 A_2(\boldsymbol{t})(t_1 - s_1) + \partial_2 A_2(\boldsymbol{t})(t_2 - s_2)|^{2H_2(\boldsymbol{t})} \\ &\quad + |A_2(\boldsymbol{t})|^{2H_2(\boldsymbol{t})} |\partial_1 A_1(\boldsymbol{t})(t_1 - s_1) + \partial_2 A_1(\boldsymbol{t})(t_2 - s_2)|^{2H_1(\boldsymbol{t})},\end{aligned}$$

where

$$H_1 = \eta_1 \circ A \quad \text{and} \quad H_2 = \eta_2 \circ A.$$

MfB sheet with domain deformation (2/2)

- Assume that there exist $\rho \in (0,1)$ such that
$$0 \leq \overline{H}(\boldsymbol{t}) - \underline{H}(\boldsymbol{t}) \leq \frac{1-\rho}{2}$$

- Then $X = W \circ A \in \mathcal{H}^{H_1,H_2}(\boldsymbol{L}, \mathcal{T})$, with $\boldsymbol{L}$ given by :
$$L_1^{(1)}(\boldsymbol{t}) = |A_2(\boldsymbol{t})|^{2H_2(\boldsymbol{t})} |\partial_1 A_1(\boldsymbol{t})|^{2H_1(\boldsymbol{t})},$$
$$L_2^{(1)}(\boldsymbol{t}) = |A_1(\boldsymbol{t})|^{2H_1(\boldsymbol{t})} |\partial_1 A_2(\boldsymbol{t})|^{2H_2(\boldsymbol{t})},$$
$$L_1^{(2)}(\boldsymbol{t}) = |A_2(\boldsymbol{t})|^{2H_2(\boldsymbol{t})} |\partial_2 A_1(\boldsymbol{t})|^{2H_1(\boldsymbol{t})},$$
$$L_2^{(2)}(\boldsymbol{t}) = |A_1(\boldsymbol{t})|^{2H_1(\boldsymbol{t})} |\partial_2 A_2(\boldsymbol{t})|^{2H_2(\boldsymbol{t})}$$

- Deduce estimating equations for the components of the deformation, depending on H_1, H_2, $\boldsymbol{L}$ and the variance of X
- Estimates of A are easily obtained by plug-in

Adaptive bivariate smoothing (1/4)

- $X \in \mathcal{H}^{H_1,H_2}$ with $\boldsymbol{L} = (L_1, 0, 0, L_2)$

$$H_i(\boldsymbol{t}) = \frac{\log(\theta_{\boldsymbol{t}}^{(i)}(2\Delta)) - \log(\theta_{\boldsymbol{t}}^{(i)}(\Delta))}{2\log(2)} + O(\Delta^{\beta}), \quad i = 1, 2.$$

- New observation

$$Y_m^{new} = X^{new}(\boldsymbol{t}_m^{new}) + \varepsilon_m^{new}, \qquad 1 \leq m \leq M_0.$$

- With $\boldsymbol{B} = diag(1/h_1, 1/h_2)$

$$\widehat{X}^{new}(\boldsymbol{t}; \mathrm{B}) = \sum_{m=1}^{M_0} Y_m^{new} \frac{K\left(\mathrm{B}(\boldsymbol{t}_m^{new} - \boldsymbol{t})\right)}{\sum_{m=1}^{M_0} K\left(\mathrm{B}(\boldsymbol{t}_m^{new} - \boldsymbol{t})\right)}.$$

- We consider the risk

$$\mathcal{R}\left(\boldsymbol{t}; \mathbf{B}, M_0\right) = \mathbb{E}\left[\left\{\widehat{X}^{new}(\boldsymbol{t}; \mathbf{B}) - X^{new}(\boldsymbol{t})\right\}^2 \middle| M_0\right].$$

Adaptive bivariate smoothing (2/4) : Assumptions

- Two constants exist $\kappa, r > 0$

$$\kappa^{-1} 1_{B(0,r)}(\boldsymbol{t}) \leq K(\boldsymbol{t}) \leq \kappa 1_{B(0,1)}(\boldsymbol{t}), \quad \forall \boldsymbol{t} \in \mathcal{T},$$

  and $h_1, h_2 \in \mathcal{H}$ with $\sqrt{\mathfrak{m}} \inf \mathcal{H} \to \infty$ and $\sup \mathcal{H} \to 0$.
- $\exists c > 0$, $f_{\mathsf{T}}(\boldsymbol{t}) \geq c$, $\forall \boldsymbol{t} \in \mathcal{T}$.
- The ε_m^{new} are iid and $\mathbb{E}[\varepsilon_m^{new}] = 0, \mathbb{E}[(\varepsilon_m^{new})^2] = \sigma^2$.
- The M_0, X^{new}, $\boldsymbol{t}_m^{new}$, and ε_m^{new}, $1 \leq m \leq M_0$, are mutually independent.

- A constant $\tilde{c} > 0$ exists such that $\tilde{c}^{-1} \leq M_0/\mathfrak{m} \leq \tilde{c}$, a.s.
- $\widehat{H}_i(\boldsymbol{t})$ and $\widehat{L}_i(\boldsymbol{t})$ are independent of M_0, X^{new}, $\boldsymbol{t}_m^{new}$, ε_m^{new}. Moreover, two constants exists $a > 0$ and $\mathfrak{k}_1$ depending on a

$$\mathbb{P}\left(|\widehat{H}_i(\boldsymbol{t}) - H_i(\boldsymbol{t})| > \log^{-a}(\mathfrak{m})\right) \leq \mathfrak{k}_1 \exp\left(-\mathfrak{m}\right), \quad i = 1, 2.$$

$$\mathbb{P}\left(|\widehat{L}_i(\boldsymbol{t}) - L_i(\boldsymbol{t})| > \log^{-a}(\mathfrak{m})\right) \leq \mathfrak{k}_1 \exp\left(-\mathfrak{m}\right), \quad i = 1, 2.$$

Adaptive bivariate smoothing (3/4)

Proposition 4

$$\mathcal{R}(\boldsymbol{t}; \mathsf{B}, M_0) \leq \frac{\kappa^2}{c\pi} \frac{\sigma^2}{M_0 h_1 h_2} + 2L_1(\boldsymbol{t}) h_1^{2H_1(\boldsymbol{t})} + 2L_2(\boldsymbol{t}) h_2^{2H_2(\boldsymbol{t})} + \text{ negligible terms.}$$

Set, for $i = 1, 2$

$$\alpha_i(\boldsymbol{t}) = \frac{\omega(\boldsymbol{t})}{2\omega(\boldsymbol{t}) + 1} \times \frac{1}{H_i(\boldsymbol{t})}, \quad \Lambda_i(\boldsymbol{t}) = \kappa^2 \sigma^2 / \{4c\pi H_i(\boldsymbol{t}) L_i(\boldsymbol{t})\},$$

and

$$\omega(\boldsymbol{t}) = \frac{H_1(\boldsymbol{t}) H_2(\boldsymbol{t})}{H_1(\boldsymbol{t}) + H_2(\boldsymbol{t})}, \quad \mathcal{H}(\boldsymbol{t}) = 2H_1(\boldsymbol{t}) H_2(\boldsymbol{t}) + H_1(\boldsymbol{t}) + H_2(\boldsymbol{t}).$$

Adaptive bivariate smoothing (4/4)

Proposition 5

- With the choice

$$h_1^* = M_0^{-\alpha_1(\boldsymbol{t})} \left[\frac{\Lambda_1(\boldsymbol{t})^{2H_2(\boldsymbol{t})+1}}{\Lambda_2(\boldsymbol{t})}\right]^{\frac{1}{2\mathcal{H}(\boldsymbol{t})}}, h_2^* = M_0^{-\alpha_2(\boldsymbol{t})} \left[\frac{\Lambda_2(\boldsymbol{t})^{2H_1(\boldsymbol{t})+1}}{\Lambda_1(\boldsymbol{t})}\right]^{\frac{1}{2\mathcal{H}(\boldsymbol{t})}},$$

- We obtain :

$$\mathcal{R}(\boldsymbol{t}; \mathbf{B}^*, M_0) \leq M_0^{-\frac{2\omega(\boldsymbol{t})}{2\omega(\boldsymbol{t})+1}} \Gamma_1(\boldsymbol{t}).$$

- Estimates of h_1^* and h_2^* are easily obtained by plug-in
- We have

$$\mathcal{R}(\boldsymbol{t}; \widehat{\mathbf{B}}^*, M_0) \leq \Gamma_2(\boldsymbol{t}) M_0^{-\frac{2\omega(\boldsymbol{t})}{2\omega(\boldsymbol{t})+1} + 2\log^{-a}(\mathfrak{m})} \times \{1 + o(\log^{-a}(\mathfrak{m}))\},$$

with $\mathfrak{m} = \mathbb{E}[M_0]$.

Take away

- Functional data are noisy, discretely observed realizations of a stochastic process
- A general class of stochastic processes defined on the plane is introduced ; the MfB sheet is an example
- The characteristics of the process are estimated nonparametrically, exploiting the replication feature of the functional data
- Non-asymptotic exponential bounds of the estimators are derived
- Two applications are proposed
    - Multifractional Brownian sheet with domain deformation
    - Optimal smoothing for reconstructing the sheets

THANK YOU

QR code to the paper on arxiv