Heavy tailed functional data analysis

Thomas Meinguet (thomas.meinguet@uclouvain.be)
Johan Segers (johan.segers@uclouvain.be)

Université catholique de Louvain, Institut de statistique

Louvain-la-Neuve, Belgium

June 23, 2009

- Introduction
- CM3 process
- Extremal properties
- Conclusion

Introduction

- Heavy tail = regularly varying behavior
- Two distinct cases:
  - *the observed process itself is heavy tailed*
    (finance, telecommunication, environment ...)

  - *a transformation of observed process is heavy tailed*
    (into unit-Fréchet margins, to study its extremal properties)

Functional data

- A sequence of observations of spatial phenomena is a *time series of function-valued data*.
- Goal: to deal with *space and time dependencies for extremes*.

  Space Cross-sectional *tail dependence*

  Time Temporal *clusters of extremes*

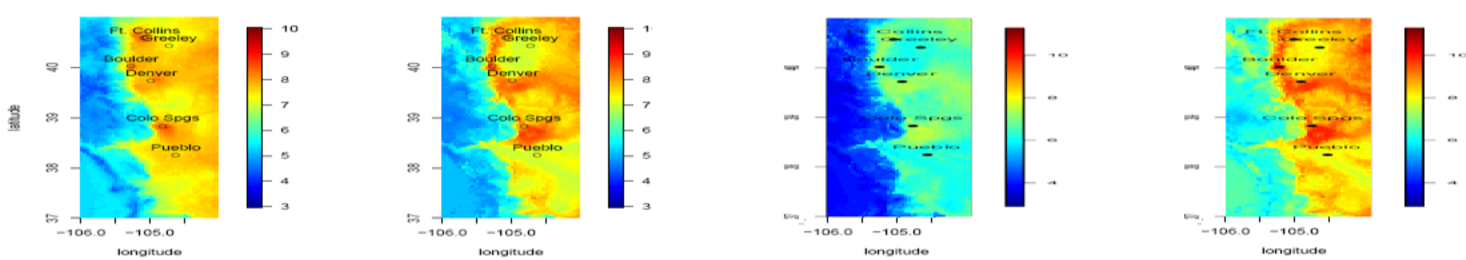


Figure 1: Rainfall in Colorado.

Functional data

- Functions are *infinite dimensional objects*.

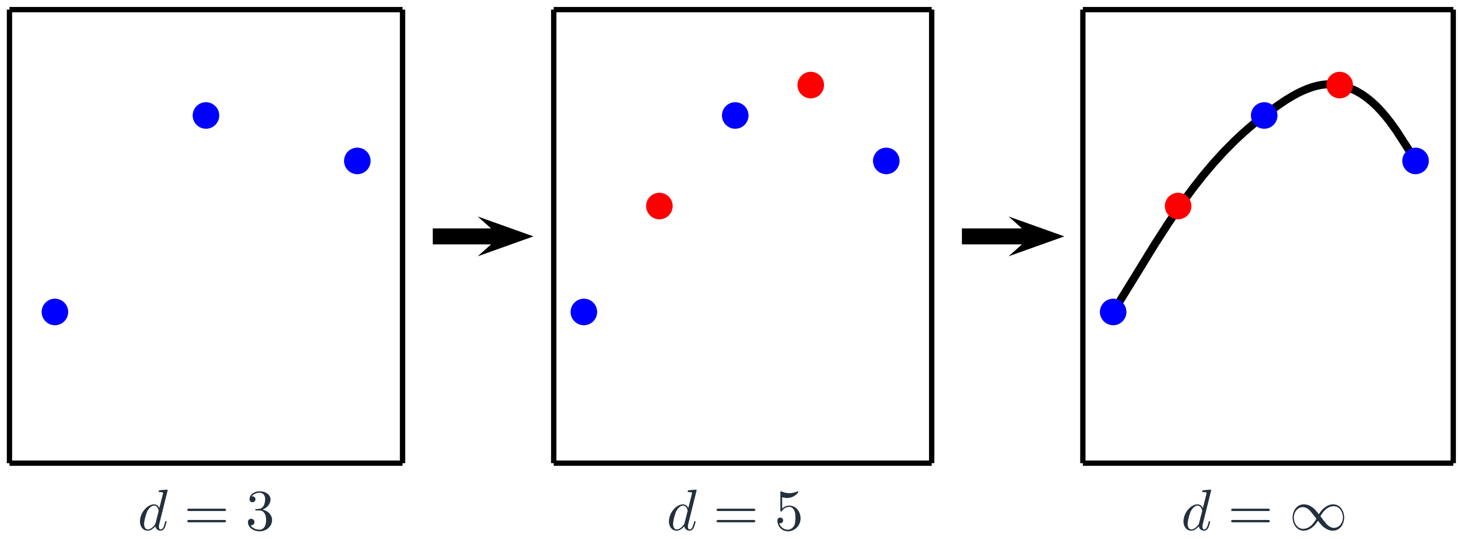


Figure 2: Functional observations $\Leftrightarrow$ "∞ measurement points".

- Sets of functions: $C([0,1]^d, \mathbb{R})$, $D([0,1]^d, \mathbb{R})$, $USC([0,1]^d)$ ...
- For this talk we focus on $C([0,1]^d, \mathbb{R})$.

- *CM3 process*: definition and motivation
- For general processes in $C([0,1]^d, \mathbb{R})$:
  - *spectral measure*
  - *spectral process*
- Illustration on the CM3 process

CM3 process

Goal

- Continuous version of Smith and Weissman's M4 process (1996).
- Designed to study extremes of spatial time series ...
- *... after standardization of the margins into unit-Fréchet*.

- *Moving Maximum (M2)*:
  - $(Z_i)_{i \geqslant 0}$ independent unit-Fréchet random variables.
  - $(a_i)_{i \geqslant 0}$ positive real numbers (weights).
  - Normalizing condition: $\sum_i a_i = 1$.
  - Process:

$$X_t = \sup_{i \geqslant 0} a_i Z_{t-i}.$$

  - X_t is still unit-Fréchet.

- *Maximum of Moving Maxima (M3)*:
  - $(Z_i^{(j)})_{i \geqslant 0, j \in \mathbb{Z}}$ independent unit-Fréchet random variables.
  - $(a_i^{(j)})_{i \geqslant 0, j \in \mathbb{Z}}$ positive real numbers (weights).
  - Normalizing condition: $\sum_i \sum_j a_i^{(j)} = 1.$
  - Process:

$$X_t = \sup_{j \in \mathbb{Z}} \sup_{i \geqslant 0} (a_i^{(j)} Z_{t-i}^{(j)}).$$

  - X_t is still unit-Fréchet.
  - Not a moving maximum.

- *Multivariate Maximum of Moving Maxima (M4)*:
  - $(Z_i^{(j)})_{i \geqslant 0, j \in \mathbb{Z}}$ independent unit-Fréchet random variables.
  - $(a_i^{(j)})_{i \geqslant 0, j \in \mathbb{Z}}$ d-dimensional vectors of positive numbers.
  - Normalizing condition: $\sum_i \sum_j a_i^{(j)}(k) = 1, 1 \leqslant k \leqslant d$.
  - Process:

$$X_t(k) = \sup_{j \in \mathbb{Z}} \sup_{i \geqslant 0} (a_i^{(j)}(k) Z_{t-i}^{(j)}), \qquad 1 \leqslant k \leqslant d.$$

  - X_t has unit-Fréchet margins for each $t \in \mathbb{Z}$.
  - Dependence structure: in case of large value of $Z_i^{(j)}$, persistence of the shock for $(X_t)_{t \geqslant i}$.

- CM3 = "*Continuous Maximum of Moving Maxima*".
  - Analogue of the M4 ...
  - ... with continuous functions instead of vectors:
  - $(Z_i^{(j)})$ array of independent unit-Fréchet RV.
  - $a_i^{(j)} : [0,1]^d \to \mathbb{R}_+$ continuous functions.
  - Normalizing condition: $\forall x, \ \sum_i \sum_j a_i^{(j)}(x) = 1$.
  - Process:

$$X_t(x) = \sup_{j \in \mathbb{Z}} \sup_{i \geqslant 0} (a_i^{(j)}(x) Z_{t-i}^{(j)}).$$

  - Properties: stationary, max-stable, unit-FM., M4-dep. struct.

Extremal properties

- Is the CM3 process *jointly* *regularly varying*?

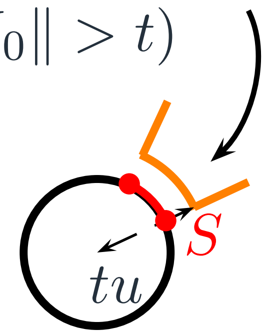
  → *Regular variation* in $C([0,1]^d, \mathbb{R})$?
  Answer: *spectral measure*.

  → *Joint regular variation* for the series ?
  Answer: *spectral process*.

Multivariate case

- Let X be a random vector in $\mathbb{R}^d$.
- Denote $\mathbb{S} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ the unit ball of $\mathbb{R}^d$.
- If, given $S \in \mathcal{B}_{\mathbb{S}}$, as $t \to \infty$,

$$\frac{P\left(\|X_0\| > tu, X_0/\|X_0\| \in S\right)}{P(\|X_0\| > t)} \xrightarrow{w} \frac{1}{u^\alpha}\lambda(S)$$



with a probability measure λ, then λ is the *spectral measure* of X_0.

- This define the *regular variation* of X_0 in $\mathbb{R}^d$:

Functional case

- The construction with spheres makes sense in $C([0,1]^d, \mathbb{R})$.
- Put $\mathbb{S} = \{x \in C([0,1]^d, \mathbb{R}) : \|x\| = 1\}$.
- **Definition**: A random continous function X is *regularly varying* if it possesses a *spectral measure* in the sense

$$\frac{P\left(\|X\| > tu, X/\|X\| \in \cdot\,\right)}{P(\|X\| > t)} \xrightarrow{w} \frac{1}{u^{\alpha}}\lambda(\,\cdot\,)$$

as $t \to \infty$, for a probability measure λ on $\mathbb{S}$.

- For the *CM3 process*

$$X_t(x) = \sup_{j\in\mathbb{Z}} \sup_{i\geqslant 0} (a_i^{(j)}(x) Z_{t-i}^{(j)}),$$

we have

$$\frac{P\left(\|X_0\| > tu, X_0/\|X_0\| \in S\right)}{P(\|X_0\| > t)} \xrightarrow{w} \frac{1}{u^\alpha}\lambda(S),$$

where

$$\lambda(S) = \sum_i \sum_j 1_{\{a_i^{(j)}/\|a_i^{(j)}\|\in S\}} \frac{\|a_i^{(j)}\|}{\sum_l \sum_k \|a_k^{(l)}\|}$$

is its spectral measure.

- After transformation, the spectral measure λ also can be given by the limit

$$\mathcal{L}\left(X_0/\|X_0\| \mid \|X_0\| \geqslant x \right) \xrightarrow{d} \lambda(\,\cdot\,).$$

- Can we replace X_0 by $(X_t)_{t\in\mathbb{Z}}$ to obtain a *limit process* $(\Theta_t)_{t\in\mathbb{Z}}$ such that

$$\mathcal{L}\left((X_t/\|X_0\|)_{t\in\mathbb{Z}} \;\Big|\; \|X_0\| \geqslant x \right) \xrightarrow{d} \mathcal{L}\left((\Theta_t)_{t\in\mathbb{Z}} \right) ?$$

- The limit distribution $(\Theta_t)_{t\in\mathbb{Z}}$ would include all features about the joint tail distribution of $(X_t)_{t\in\mathbb{Z}}$.

- With this definition $\mathcal{L}(\Theta_0)$ is equal to λ.

Special cases

- Independent series $(X_t)_{t\in\mathbb{Z}}$:

$$\mathcal{L}(\Theta_t) = 0 \qquad \text{if } t \neq 0.$$

- Fully dependent series $(X_t)_{t\in\mathbb{Z}}$:

$$\mathcal{L}(\Theta_t) = \lambda \qquad \forall t \in \mathbb{Z}.$$

- For the CM3 process,

$$\mathcal{L}\left(\left(X_t/\|X_0\|\right)_{t\in\mathbb{Z}} \;\middle|\; \|X_0\| \geqslant x\right) \xrightarrow{d} \mathcal{L}\left((\Theta_t)_{t\in\mathbb{Z}}\right)$$

with

$$P(\Theta_{-s} \in A_{-s}, \ldots, \Theta_t \in A_t) =$$

$$\sum_i \sum_j 1_{\left\{\left(\frac{a^{(j)}_{-s+i}}{\|a^{(j)}_i\|}, \ldots, \frac{a^{(j)}_{t+i}}{\|a^{(j)}_i\|}\right) \in A_{-s}\times\ldots\times A_t\right\}} \frac{\|a^{(j)}_i\|}{\sum_l \sum_k \|a^{(l)}_k\|}.$$

- Given a stationary process $(X_t)_{t\in\mathbb{Z}}$, if $(\Theta_t)_{t\in\mathbb{Z}}$ exists such that

$$\mathcal{L}\left((X_t/\|X_0\|)_{t\in\mathbb{Z}} \;\middle|\; \|X_0\| \geqslant x \right) \xrightarrow{d} \mathcal{L}\left((\Theta_t)_{t\in\mathbb{Z}} \right)$$

  we call $(\Theta_t)_{t\in\mathbb{Z}}$ the *spectral process* of $(X_t)_{t\in\mathbb{Z}}$.

- **General theorem**: A stationary process $(X_t)_{t\in\mathbb{Z}}$ in $C([0,1]^d, \mathbb{R})$ admits a *spectral process* $(\Theta_t)_{t\in\mathbb{Z}}$ if and only if $(X_t)_{t\in\mathbb{Z}}$ is *jointly regularly varying* in the sense of, for every k,

$$(X_0, \ldots, X_k)$$

  is regularly varying with the same index.

- Extremal index θ of the series $(\|X_t\|)_{t\in\mathbb{Z}}$:

$$P(\max_{1\leqslant t\leqslant n} \|X_t\| \leqslant nx) \to e^{-\theta/x}.$$

- Explicit expression thanks to the spectral process:

$$\theta = E[\sup_{i\geqslant 0} \|\Theta_i\|^\alpha - \sup_{i\geqslant 1} \|\Theta_i\|^\alpha].$$

- CM3 process:

$$\theta = \frac{\sum_j \max_i \|a_i^{(j)}\|}{\sum_j \sum_i \|a_i^{(j)}\|}.$$

Conclusion

- Theory of joint regular variation for spatial time series
- CM3 processes extend M4 processes to the spatial case
- Other implications for functional processes:
  - Point processes limit for spatial extremes
  - Tail dependance quantities / extremograms
  - Spatial linear processes
- All is actually valid in separable Banach spaces
- Other ideas: semi-continuous functions...