

Estimación

1

Propiedades de $\widehat{\Gamma}$

Ejemplos

A 2x6 grid of 12 small circles, arranged in two rows of six.

Mediana

OC

Profundidad

○○○○○○○○○○
○○○○○○

Boxplot Funcional

10

Functional Data Analysis: Estimación y Boxplot

Graciela Boente

Estimación

- $X_i : \Omega \rightarrow \mathcal{H}, 1 \leq i \leq n$, observaciones **i.i.d.**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^2 < \infty$.
- Sea $\mu = \mathbb{E}X$ y $\Gamma = \mathbb{E}\{(X - \mu) \otimes (X - \mu)\}$.
- μ puede estimarse usando la media muestral

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$$

- Γ puede estimarse usando el operador de covarianza muestral

$$\hat{\Gamma} = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{\mu}) \otimes (X_i - \hat{\mu})$$

Estimación

- Observemos que $\widehat{\Gamma}$ tiene rango finito, es decir, tiene sólo un número finito de autovalores no nulos.
- Si $\mathcal{H} = L^2(\mathcal{I})$, entonces

$$\widehat{\gamma}(s, t) = \frac{1}{n} \sum_{i=1}^n (X_i(t) - \widehat{\mu}(t))(X_i(s) - \widehat{\mu}(s))$$

Autofunciones

- Podemos estimar las autofunciones de Γ usando las de $\widehat{\Gamma}$ asociadas a los autovalores no nulos.
- Si $\mathcal{H} = L^2(\mathcal{I})$, y tenemos un proceso de Wiener $W(t)$ con $\gamma(s, t) = \min(s, t)$,

$$\lambda_j = \pi^{-2}(j - 0.5)^{-2}$$

$$\phi_j(t) = \sqrt{2} \sin \left(\left(j - \frac{1}{2} \right) \pi t \right)$$

- Si $\mathcal{H} = L^2(\mathcal{I})$, y tenemos un puente Browniano $X(t)$ con $\gamma(s, t) = \min(s, t) - ts$ entonces

$$\lambda_j = \pi^{-2}j^{-2}$$

$$\phi_j(t) = \sqrt{2} \sin(j \pi t)$$

Propiedades

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad \mu = \mathbb{E}(X)$$

- Si $\{X_i\}_{i \geq 1}$ son i.i.d $X_i \sim X$, con $\mathbb{E}\|X\| < \infty$, entonces
 - $\mathbb{E}\hat{\mu} = \mu$
 - Por la ley de los grandes números

$$\hat{\mu} \xrightarrow{a.s.} \mu$$

- Si $\mathbb{E}\|X\|^2 < \infty$

$$n \mathbb{E} (\|\hat{\mu} - \mu\|^2) = O(1)$$

Propiedades: $n\mathbb{E}(\|\hat{\mu} - \mu\|^2) = O(1)$

Sea $Z_i = X_i - \mu$ entonces

$$\begin{aligned}\mathbb{E}(\|\hat{\mu} - \mu\|^2) &= \mathbb{E}\left(\left\|\frac{1}{n} \sum_{i=1}^n Z_i - \mu\right\|^2\right) = \mathbb{E}\left(\left\|\frac{1}{n} \sum_{i=1}^n (X_i - \mu)\right\|^2\right) \\ &= \frac{1}{n^2} \mathbb{E}\left\langle \sum_{i=1}^n Z_i, \sum_{j=1}^n Z_j \right\rangle = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \mathbb{E}\langle Z_i, Z_j \rangle\end{aligned}$$

Ahora bien $\mathbb{E}Z_i = 0$ y si $i \neq j$, Z_i es independiente de Z_j . Por lo tanto, si $i \neq j$

$$\mathbb{E}\{\langle Z_i, Z_j \rangle | Z_j = z\} = \mathbb{E}\{\langle Z_i, z \rangle | Z_j = z\} = \mathbb{E}\{\langle Z_i, z \rangle\} = 0$$

pues

$$\langle \mathbb{E}Z, u \rangle = \mathbb{E}\langle Z, u \rangle \quad \forall u \in \mathcal{H}$$

Luego, $\mathbb{E}\{\langle Z_i, Z_j \rangle\} = 0$ si $i \neq j$, de donde

$$\mathbb{E}(\|\hat{\mu} - \mu\|^2) = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}\langle Z_i, Z_i \rangle = \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}\|Z_i\|^2 = \frac{1}{n} \mathbb{E}\|X - \mu\|^2$$

Propiedades

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i \quad \mu = \mathbb{E}(X)$$

Si $\mathbb{E}\|X\|^2 < \infty$, por el TCL

$$\sqrt{n}(\hat{\mu} - \mu) \xrightarrow{D} N(0, \Gamma)$$

Qué podemos decir de $\widehat{\Gamma}$

- $X_i : \Omega \rightarrow \mathcal{H}$, $1 \leq i \leq n$, observaciones **i.i.d.**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^2 < \infty$.
 - Sea $\mu = \mathbb{E}X$ y $\Gamma = \mathbb{E}\{(X - \mu) \otimes (X - \mu)\}$.

$$\widehat{\Gamma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{X}_i - \widehat{\boldsymbol{\mu}}) \otimes (\mathbf{X}_i - \widehat{\boldsymbol{\mu}})$$

Espacios de operadores

- \mathcal{H} espacio de Hilbert separable
 - Let $\Upsilon : \mathcal{H} \rightarrow \mathcal{H}$ un operador lineal acotado ($\Upsilon \in \mathcal{L}$)

$$\|\Upsilon\|_{\mathcal{L}} = \sup_{\|u\| < 1} \|\Upsilon(u)\|$$

- El operador Υ se dice un operador de traza si, para toda base ortonormal $\{u_\ell : \ell \geq 1\}$

$$\sum_{\ell=1}^{\infty} \langle \Upsilon u_\ell, u_\ell \rangle < \infty$$

- El operador Υ se dice **Hilbert–Schmidt** si, para toda base ortonormal $\{u_\ell : \ell \geq 1\}$

$$\sum_{\ell=1}^{\infty} \langle \Upsilon u_\ell, \Upsilon u_\ell \rangle = \sum_{\ell=1}^{\infty} \| \Upsilon u_\ell \|^2 < \infty$$

Espacios de operadores

Llamaremos \mathcal{F} al espacio de Hilbert de los operadores Hilbert–Schmidt sobre \mathcal{H}

- Dados Υ_1 y Υ_2 operadores Hilbert–Schmidt, el producto interno en \mathcal{F} se define como

$$\langle \Upsilon_1, \Upsilon_2 \rangle_{\mathcal{F}} = \text{traza}(\Upsilon_1^* \Upsilon_2) = \sum_{\ell=1}^{\infty} \langle \Upsilon_1 u_{\ell}, \Upsilon_2 u_{\ell} \rangle$$

- Dado $\Upsilon \in \mathcal{F}$ la norma es igual a

$$\|\Upsilon\|_{\mathcal{F}} = \langle \Upsilon^*, \Upsilon \rangle_{\mathcal{F}}^{1/2} = \left\{ \sum_{\ell=1}^{\infty} \|\Upsilon u_{\ell}\|^2 \right\}^{1/2}$$

donde $\{u_{\ell} : \ell \geq 1\}$ es cualquier base ortonormal de \mathcal{H} .

Las definiciones anteriores son independientes de la elección de la base.

Espacios de operadores

\mathcal{F} : espacio de Hilbert de los operadores Hilbert–Schmidt

- Los operadores Hilbert–Schmidt tienen numerables autovalores. Si el operador de auto–adjunto (o simétrico), todos los autovalores son reales.
- Si $\Upsilon \in \mathcal{F}$ es definido-nonegativo, autoadjunto, entonces como base la base ortonormal de las autofunciones de Υ ,

$$\|\Upsilon\|_{\mathcal{F}}^2 = \sum_{\ell=1}^{\infty} \lambda_{\ell}^2$$

donde los autovalores no–nulos $\{\lambda_{\ell} : \ell \geq 1\}$ se ordenan en forma decreciente $\lambda_{\ell} \geq \lambda_{\ell+1}$.

Espacios de operadores

- Si $\mathcal{H} = L^2(\mathcal{I})$ y Γ es el operador de covarianza con núcleo de covarianza

$$\gamma(s, t) = \text{Cov}(X(s), X(t))$$

entonces

$$\|\Gamma\|_{\mathcal{F}}^2 = \sum_{\ell=1}^{\infty} \lambda_{\ell}^2 = \|\gamma\|^2 = \int_{\mathcal{I}} \int_{\mathcal{I}} \gamma^2(t, s) dt ds$$

- Si $\Upsilon = u \otimes u$, para $u \in \mathcal{H}$, entonces

$$\|\Upsilon\|_{\mathcal{F}}^2 = \|u\|^2$$

$$\|\Upsilon\|_{\mathcal{F}}^2 = \|u\|^2$$

Sea $\{u_\ell : \ell \geq 1\}$ una base ortonormal de \mathcal{H} , usando que

$$(u \otimes u)w = \langle u, w \rangle u$$

obtenemos

$$\begin{aligned} \|\Upsilon\|_{\mathcal{F}}^2 &= \sum_{\ell=1}^{\infty} \|\Upsilon u_\ell\|^2 = \sum_{\ell=1}^{\infty} \|(u \otimes u)u_\ell\|^2 \\ &= \sum_{\ell=1}^{\infty} \|u \langle u, u_\ell \rangle\|^2 = \sum_{\ell=1}^{\infty} \|u\|^2 \langle u, u_\ell \rangle^2 \\ &= \|u\|^2 \sum_{\ell=1}^{\infty} \langle u, u_\ell \rangle^2 = \|u\|^4 \end{aligned}$$

donde la última igualdad vale por la igualdad de Parseval.

Propiedades de $\widehat{\Gamma}$

- $X_i : \Omega \rightarrow \mathcal{H}$, $1 \leq i \leq n$, observaciones **i.i.d.**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^2 < \infty$.

1

$$\mathbb{E}(\widehat{\Gamma}) = \frac{n-1}{n} \Gamma$$

- Por la ley de los grandes números en \mathcal{F}

$$\widehat{\Gamma} = \frac{1}{n} \sum_{i=1}^n (X_i - \widehat{\mu}) \otimes (X_i - \widehat{\mu}) \xrightarrow{a.s.} \Gamma$$

Estimación

1

Propiedades de $\hat{\Gamma}$

10

Ejemplos

100

Mediana

OC

Profundidad

A 3x10 grid of circles, arranged in three rows and ten columns.

Boxplot Funcional

A 2x8 grid of 16 empty circles arranged in two rows of eight.

Propiedades de $\widehat{\Gamma}$: $\mathbb{E}(\widehat{\Gamma}) = \frac{n-1}{n} \Gamma$

Observemos que

$$\begin{aligned}
\widehat{\Gamma} &= \frac{1}{n} \sum_{i=1}^n (X_i - \widehat{\mu}) \otimes (X_i - \widehat{\mu}) = \frac{1}{n} \sum_{i=1}^n \{(X_i - \mu) + (\mu - \widehat{\mu})\} \otimes (X_i - \widehat{\mu}) \\
&= \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \otimes (X_i - \widehat{\mu}) + (\mu - \widehat{\mu}) \otimes \left\{ \frac{1}{n} \sum_{i=1}^n (X_i - \widehat{\mu}) \right\} \\
&= \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \otimes (X_i - \widehat{\mu}) = \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \otimes \{(X_i - \mu) + (\mu - \widehat{\mu})\} \\
&= \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \otimes (X_i - \mu) + \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \otimes (\mu - \widehat{\mu}) \\
&= \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \otimes (X_i - \mu) + (\widehat{\mu} - \mu) \otimes (\mu - \widehat{\mu}) \\
&= \frac{1}{n} \sum_{i=1}^n (X_i - \mu) \otimes (X_i - \mu) - (\widehat{\mu} - \mu) \otimes (\widehat{\mu} - \mu)
\end{aligned}$$

Estimación

9

三

Propiedades de $\hat{\Gamma}$

1

Ejemplos

999999

100

Mediana

Q

Profundidad

○○○○○○○○○○

10

○○○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○

○○○○○○○○○○

Propiedades de $\widehat{\Gamma}$: $\mathbb{E}(\widehat{\Gamma}) = \frac{n-1}{n} \Gamma$

Luego,

$$\begin{aligned}\mathbb{E}(\widehat{\Gamma}) &= \mathbb{E} \left\{ \frac{1}{n} \sum_{i=1}^n (\boldsymbol{X}_i - \boldsymbol{\mu}) \otimes (\boldsymbol{X}_i - \boldsymbol{\mu}) \right\} - \mathbb{E} \{ (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \otimes (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \} \\ &= \Gamma - \mathbb{E} \{ (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \otimes (\widehat{\boldsymbol{\mu}} - \boldsymbol{\mu}) \}\end{aligned}$$

Como $\hat{\mu} - \mu = (1/n) \sum_{i=1}^n Z_i$ con $Z_i = X_i - \mu$, tenemos

$$\begin{aligned}\text{Cov}(\langle u, \hat{\mu} - \mu \rangle, \langle v, \hat{\mu} - \mu \rangle) &= \frac{1}{n^2} \text{Cov} \left(\langle u, \sum_{i=1}^n Z_i \rangle, \langle v, \sum_{s=1}^n Z_s \rangle \right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{s=1}^n \text{Cov}(\langle u, Z_i \rangle, \langle v, Z_s \rangle)\end{aligned}$$

Estimación

3

三

Propiedades de $\hat{\Gamma}$

•

○○○●○○○○○

Ejemplos

5

10

Mediana

Q

Profundidad

○○○○○○○○○○

100

Boxplot Funcional

○○○○○○○○○

Page 10

Propiedades de $\widehat{\Gamma}$: $\mathbb{E}(\widehat{\Gamma}) = \frac{n-1}{n} \Gamma$

$$\begin{aligned}\text{Cov}(\langle u, \hat{\mu} - \mu \rangle, \langle v, \hat{\mu} - \mu \rangle) &= \frac{1}{n^2} \sum_{i=1}^n \text{Cov}(\langle u, Z_i \rangle, \langle v, Z_i \rangle) \\ &= \frac{1}{n} \text{Cov}(\langle u, X - \mu \rangle, \langle v, X - \mu \rangle) \\ &= \frac{1}{n} \langle u, \Gamma v \rangle \\ &= \langle u, \frac{1}{n} \Gamma v \rangle\end{aligned}$$

Por lo tanto, el operador de covarianza de $\hat{\mu} - \mu$ es $(1/n)\Gamma$, donde

$$\mathbb{E}(\widehat{\Gamma}) = \Gamma - \frac{1}{n}\Gamma = \frac{n-1}{n}\Gamma$$

10

○○○○○
○○○○●○○○○

The diagram consists of three horizontal rows of small circles. The top row contains 6 circles. The middle row contains 5 circles. The bottom row contains 8 circles.

○○

The diagram consists of three horizontal rows of small circles. The top row contains 8 circles. The middle row contains 5 circles. The bottom row contains 10 circles.

Propiedades de $\widehat{\Gamma}$

- $X_i : \Omega \rightarrow \mathcal{H}$, $1 \leq i \leq n$, observaciones **i.i.d.**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^4 < \infty$.
 - $\tilde{\otimes}$ es el producto tensorial en \mathcal{F}
 - $\{\phi_\ell : \ell \geq 1\}$ son las autofunciones de Γ con autovalores asociados $\{\lambda_\ell : \ell \geq 1\}$ tales que $\lambda_\ell \geq \lambda_{\ell+1}$.
 - ξ_m las coordenadas estandarizadas de $X - \mu$ en la base $\{\phi_\ell : \ell \geq 1\}$, o sea,

$$\xi_m = \lambda_m^{-1/2} \langle X - \mu, \phi_m \rangle$$

- Por lo tanto,

$$\mathbb{E}(\xi_m) = 0$$

$$\mathbb{E}(\xi_m^2) = 1$$

$$\mathbb{E}(\xi_m$$

Propiedades de $\widehat{\Gamma}$

- $X_i : \Omega \rightarrow \mathcal{H}$, $1 \leq i \leq n$, observaciones **i.i.d.**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^4 < \infty$.
 - Dauxois et al. (1982) mostraron que

$$\sqrt{n} \left(\widehat{\Gamma} - \Gamma \right) \xrightarrow{D} U \quad \text{en } \mathcal{F}$$

con U un elemento Gaussiano de \mathcal{F} con operador de covarianza $\boldsymbol{\Upsilon}$ dado por

$$\begin{aligned}\Upsilon &= \sum_{m,r,o,p} \lambda_m^{1/2} \lambda_r^{1/2} \lambda_o^{1/2} \lambda_p^{1/2} \mathbb{E}(\xi_m \xi_r \xi_o \xi_p) \phi_m \otimes \phi_r \tilde{\otimes} \phi_o \otimes \phi_p \\ &\quad - \sum_{m,r} \lambda_m \lambda_r \phi_m \otimes \phi_m \tilde{\otimes} \phi_r \otimes \phi_r\end{aligned}$$

Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12, 136-154

Estimación

10

Propiedades de Γ

100

Ejemplos

A 3x5 grid of 15 empty circles, arranged in three rows and five columns.

Mediana

OC

Profundidad

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Boxplot Funcional

○○○○○○○○

Propiedades de $\hat{\phi}_j$

- $X_i : \Omega \rightarrow \mathcal{H}$, $1 \leq i \leq n$, observaciones **i.i.d.**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^4 < \infty$.
 - Sean $\widehat{\phi}_j$ las autofunciones de $\widehat{\Gamma}$ con autovalores asociados $\widehat{\lambda}_j$.
 - Supongamos $\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}$
 - $\widehat{\lambda}_j \xrightarrow{a.s.} \lambda_j$, $j = 1, \dots, p$
 - signo $\left(\langle \widehat{\phi}_j, \phi_j \rangle \right) \widehat{\phi}_j \xrightarrow{a.s.} \phi_j$, $j = 1, \dots, p$

Estimación
○○○
○○○

Propiedades de $\widehat{\Gamma}$
○○○○○
○○○○○●○

Ejemplos
○○○○○○
○○○○○
○○○○○○○○○○

Mediana
○○

Profundidad
○○○○○○○○○○
○○○○○
○○○○○○○○○○○○

Boxplot Funcional
○○○○○○○○○○
○○○○○○○○○○○○

Propiedades de $\widehat{\phi}_j$

- $X_i : \Omega \rightarrow \mathcal{H}$, $1 \leq i \leq n$, observaciones **i.i.d.**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^4 < \infty$.
- Sean $\widehat{\phi}_j$ las autofunciones de $\widehat{\Gamma}$ con autovalores asociados $\widehat{\lambda}_j$.

$$\limsup n \mathbb{E} \left\{ \left(\widehat{\lambda}_j - \lambda_j \right)^2 \right\} < \infty$$

$$\limsup n \mathbb{E} \left\{ \|\text{signo} \left(\langle \widehat{\phi}_j, \phi_j \rangle \right) \widehat{\phi}_j - \phi_j\|^2 \right\} < \infty$$

- Dauxois *et al.* (1982) obtuvieron la distribución asintótica de $\sqrt{n} (\widehat{\phi}_j - \phi_j)$

Propiedades de $\hat{\phi}_j$

- En particular, si $\mathcal{H} = L^2([0, 1])$ y todos los autovalores no nulos son distintos (tienen multiplicidad 1), Dauxois *et al.* (1982) muestran que si X_i es Gaussiano

a) $\sqrt{n} \left(\hat{\lambda}_j - \lambda_j \right) \xrightarrow{D} N(0, 2 \lambda_j^2)$

- b) Sea Υ_j el operador definido por $(\Upsilon_j v)(t) = \int_0^1 \rho_j(t, s)v(s)ds$, donde

$$\rho_j(t, s) = \lambda_j \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \frac{\lambda_k}{(\lambda_j - \lambda_k)^2} \phi_k(t) \phi_j(s).$$

Entonces, se tiene que $\sqrt{n} \left(\hat{\phi}_j - \phi_j \right) \xrightarrow{D} N(0, \Upsilon_j)$.

Estimación

10

Propiedades de $\widehat{\Gamma}$

A diagram consisting of two rows of small circles. The top row contains four circles arranged horizontally. Below it, the bottom row contains seven circles arranged horizontally.

Ejemplos

A diagram consisting of three rows of small circles. The top row contains 5 circles. The middle row contains 4 circles. The bottom row contains 8 circles.

Mediana

○○

Profundidad

○○○○○○○○○○

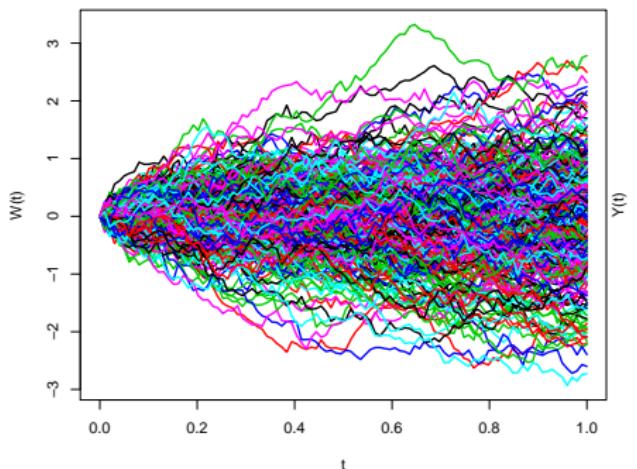
○○○○○○○○

○○○○○○○○○○

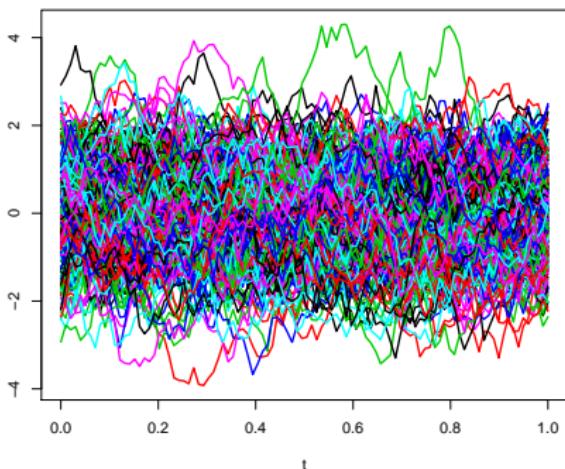
Boxplot Funcional

Ejemplos

Wiener



$$Y(t)$$



Estimación

○○○

三〇〇

Propiedades de $\widehat{\Gamma}$

• 00000

○○○○○○○○

Ejemplos

5

— 1 —

Mediana

○

Profundidad

○○○○○○○○○○

○○○○○○○

Boxplot Funcional

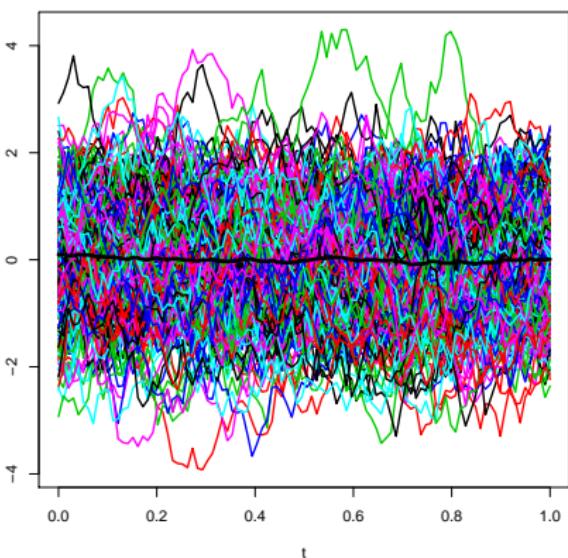
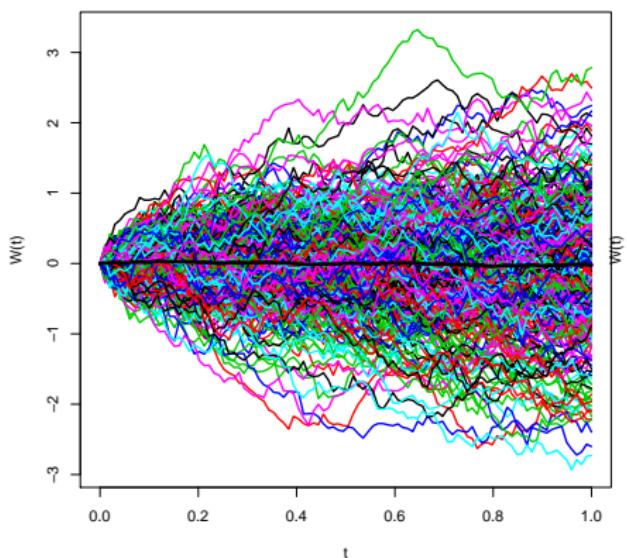
○○○○○○○○○

○○○○○○○○○○

Ejemplos

Wiener

$$Y(t)$$



Estimación

6

Propiedades de $\widehat{\Gamma}$

A 2x5 grid of ten empty circles, arranged in two rows of five.

Ejemplos

A diagram consisting of three rows of open circles. The top row contains 5 circles, the middle row contains 4 circles, and the bottom row contains 7 circles, representing the numbers 5, 4, and 7 respectively.

Mediana

○○

Profundidad

A 3x6 grid of 18 small circles, arranged in three rows and six columns.

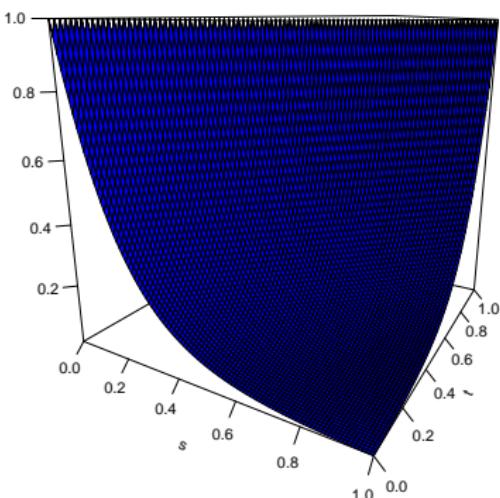
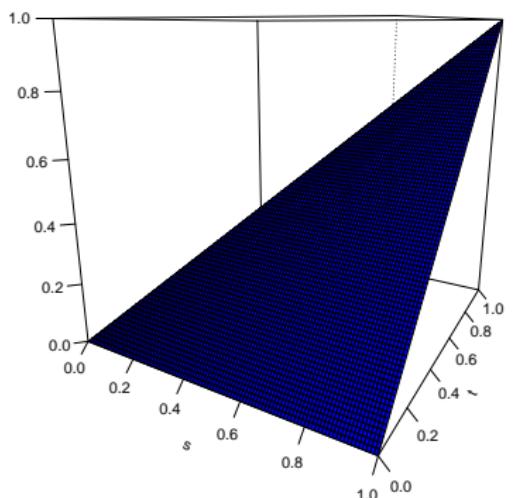
Boxplot Funcional

A diagram consisting of two rows of circles. The top row contains 8 circles arranged horizontally. Below it, another row of 8 circles is also arranged horizontally, creating a total of 16 circles in a 2x8 grid pattern.

Ejemplos: Γ

Wiener

$$Y(t)$$



Estimación

10

Propiedades de $\widehat{\Gamma}$

A horizontal row of ten small, light-gray circles arranged in a single line.

Ejemplos

A diagram consisting of three rows of five circles each. The circles in the first two rows are white, while the circle in the bottom right corner is shaded.

Mediana

○○

Profundidad

○○○○○○○○○○

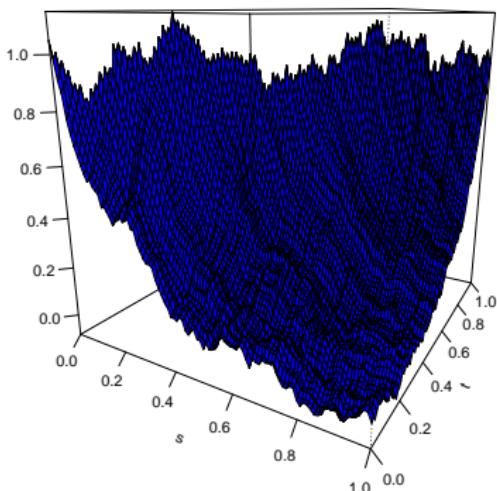
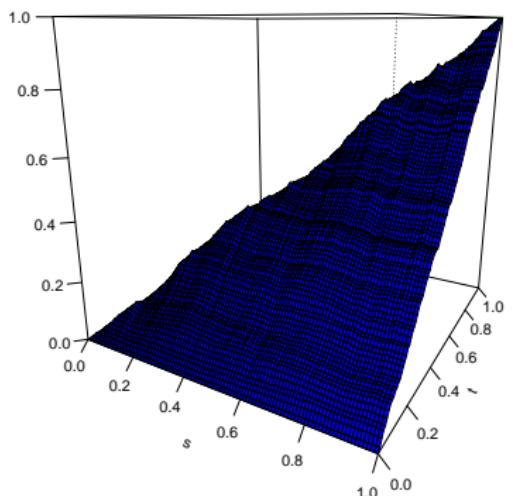
Boxplot Funcional

The diagram consists of two rows of seven circles each. The top row has circles at positions (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), and (1,7). The bottom row has circles at positions (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), and (2,7).

Ejemplos: $\widehat{\Gamma}$

Wiener

$$Y(t)$$



Estimación

1

Propiedades de $\widehat{\Gamma}$

A 2x7 grid of 14 small circles, arranged in two rows of seven.

Ejemplos

A 3x8 grid of circles, arranged in three rows and eight columns.

Mediana

oc

Profundidad

○○○○○○○○○○

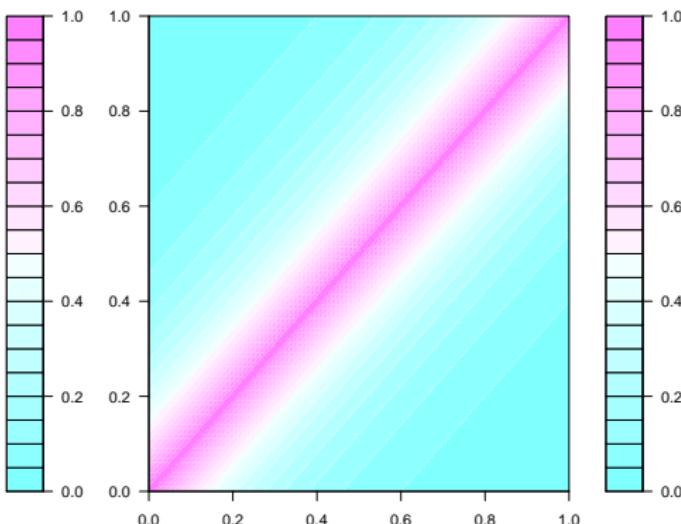
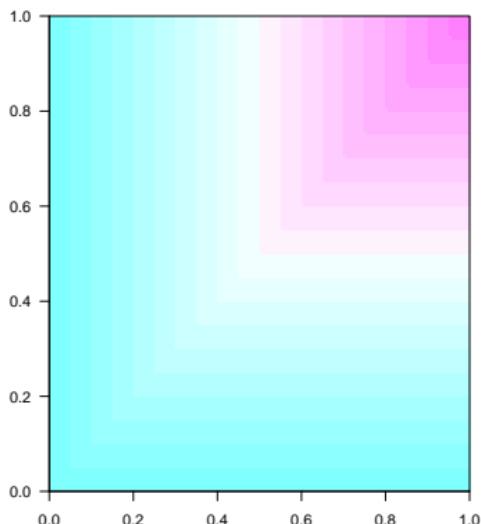
Boxplot Funcional

○○○○○○○○○○

Ejemplos: Γ

Wiener

$$Y(t)$$



Estimación

○○○
○○○

Propiedades de $\hat{\Gamma}$

○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○

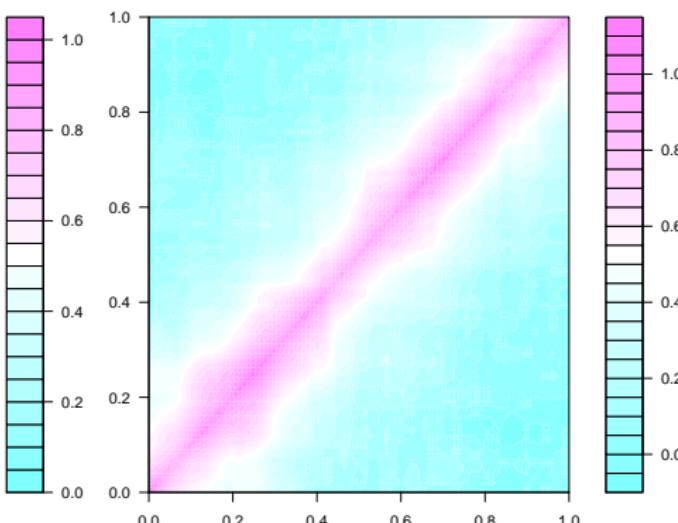
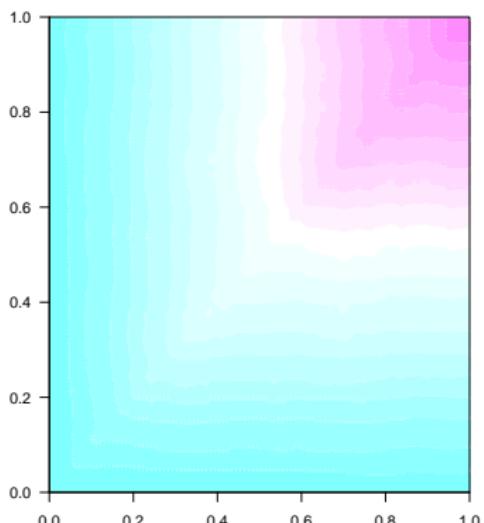
Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Ejemplos: $\hat{\Gamma}$

Wiener

$Y(t)$



Estimación

10

Propiedades de $\widehat{\Gamma}$

A diagram consisting of two rows of small circles. The top row contains four circles arranged horizontally. The bottom row contains seven circles arranged horizontally, positioned directly below the top row.

Ejemplos

Mediana

○○

Profundidad

A horizontal row of ten small, hollow circles arranged in a single line.

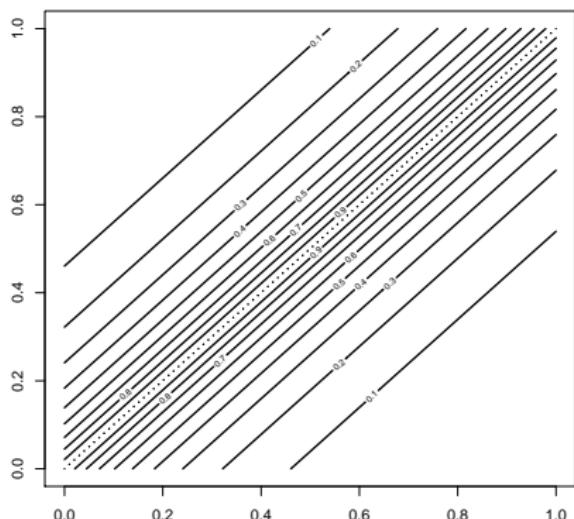
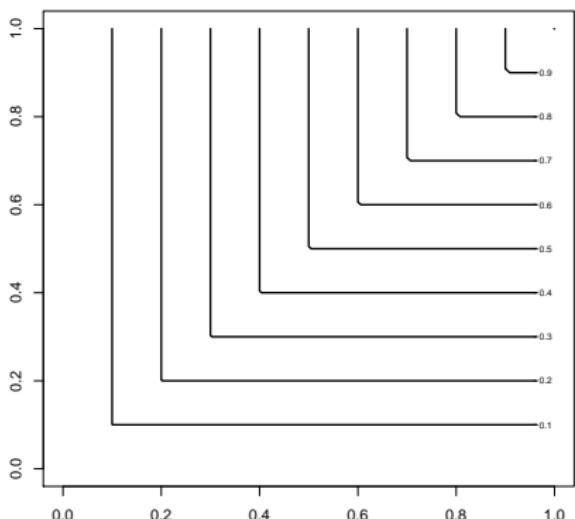
Boxplot Funcional

A 2x10 grid of 20 small circles, arranged in two rows of ten.

Ejemplos: Γ

Wiener

$$Y(t)$$



Estimación

Propiedades de $\widehat{\Gamma}$

Ejemplos

Mediana

Profundidad

Boxplot Funcional

6

The diagram consists of two rows of small circles. The top row contains five circles arranged horizontally. Below it, the bottom row contains seven circles arranged horizontally.

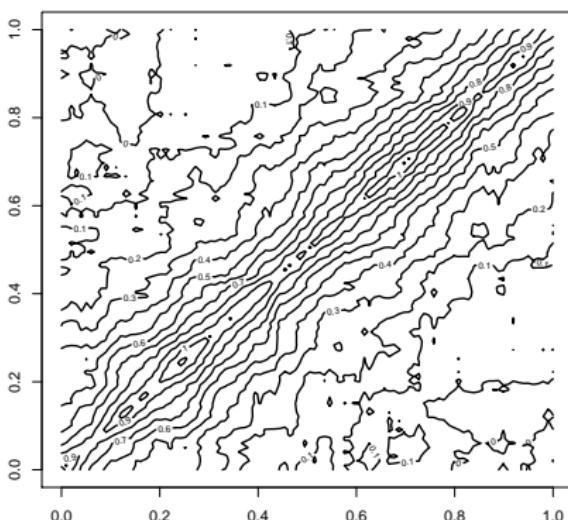
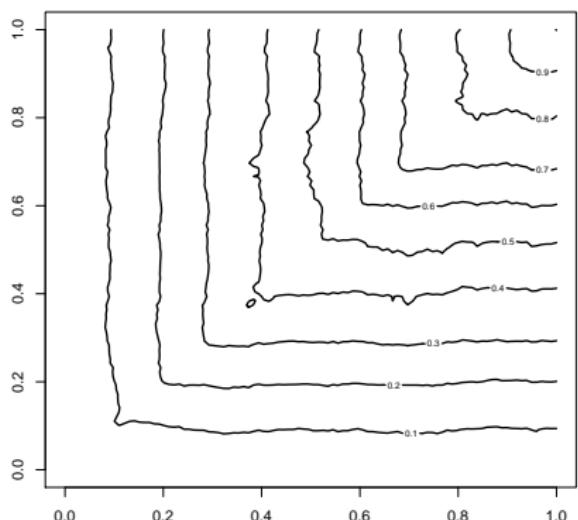
○○

A diagram consisting of three rows of small circles. The top row contains 8 circles. The middle row contains 5 circles. The bottom row contains 10 circles.

Ejemplos: $\hat{\Gamma}$

Wiener

$$Y(t)$$



Estimación

- Podemos estimar las autofunciones de Γ usando las de $\widehat{\Gamma}$ asociadas a los autovalores no nulos.
 - Si $\mathcal{H} = L^2(\mathcal{I})$, y tenemos un proceso de Wiener (o movimiento Browniano) $W(t)$ con

$$\gamma(s, t) = \min(s, t)$$

entonces

$$\phi_j(t) = \sqrt{2} \sin \left(\left(j - \frac{1}{2} \right) \pi t \right)$$

Estimación

○○○
○○○Propiedades de $\hat{\Gamma}$ ○○○○○
○○○○○○○○○○

Ejemplos

○●○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

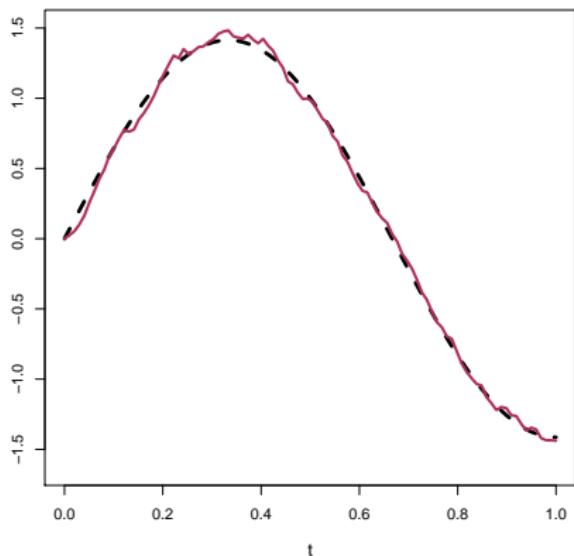
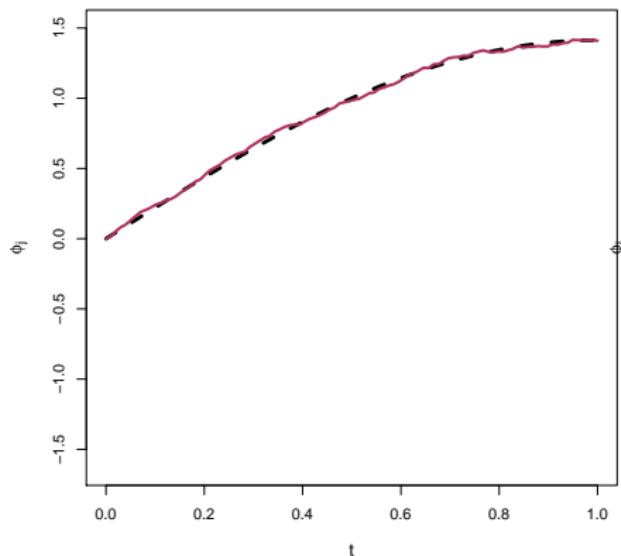
Profundidad

○○○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○○○
○○○○○○○○○○○○○○○○○○

Ejemplos: $\hat{\phi}$ para $W(t)$

 $j = 1$ $j = 2$ 

Estimación

○○○
○○○Propiedades de $\hat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○●○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

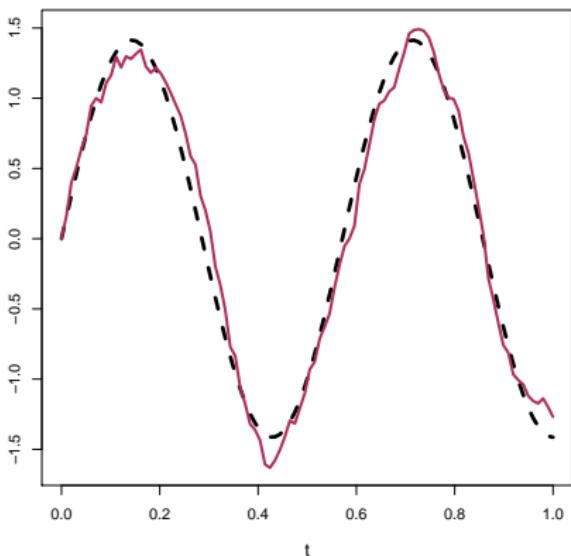
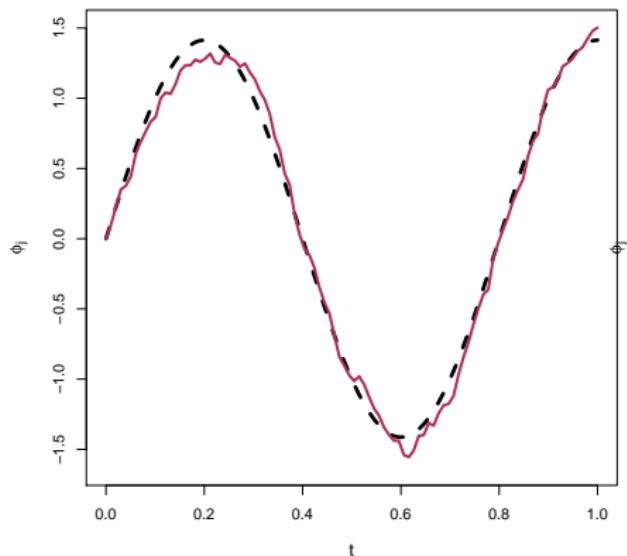
Profundidad

○○○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Ejemplos: $\hat{\phi}$ para $W(t)$

 $j = 3$ $j = 4$ 

Estimación

- Podemos estimar las autofunciones de Γ usando las de $\widehat{\Gamma}$ asociadas a los autovalores no nulos.
 - Si $\mathcal{H} = L^2(\mathcal{I})$, y tenemos un puente Browniano $X(t)$ con

$$\gamma(s, t) = \min(s, t) - t s$$

entonces

$$\lambda_j = \pi^{-2} j^{-2}$$

$$\phi_j(t) = \sqrt{2} \sin(j\pi t)$$

Estimación

10

Propiedades de $\widehat{\Gamma}$

Ejemplos

Mediana

OC

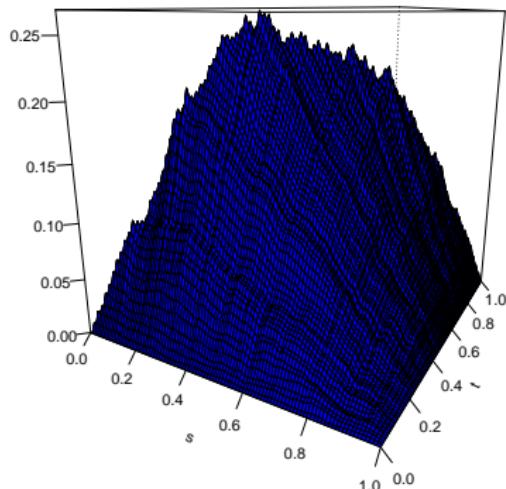
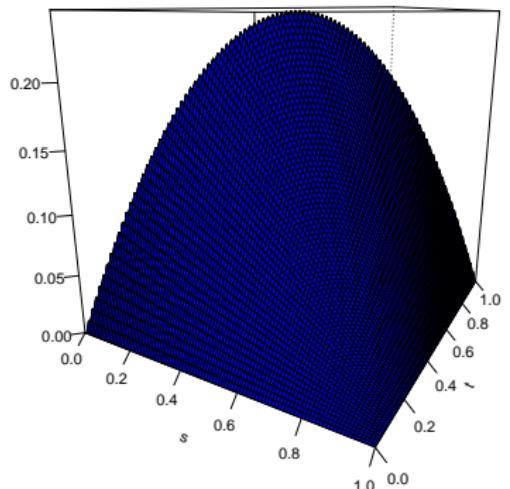
Profundidad

○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○

Ejemplos: $\widehat{\Gamma}$ para $X(t)$



Estimación

○○○
○○○Propiedades de $\hat{\Gamma}$ ○○○○○
○○○○○○○○○

Ejemplos

○○○○○●○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

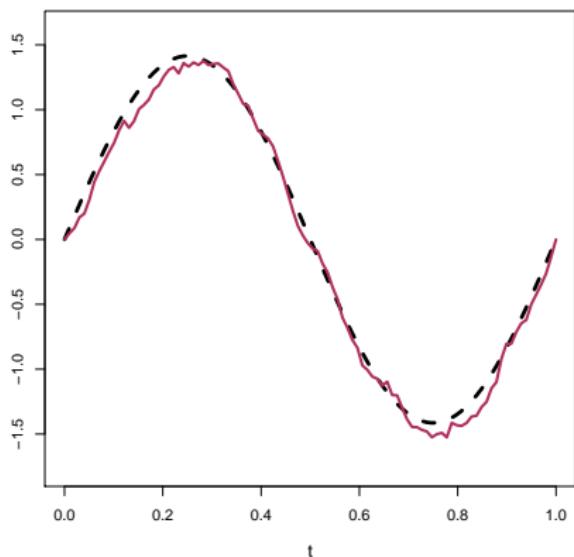
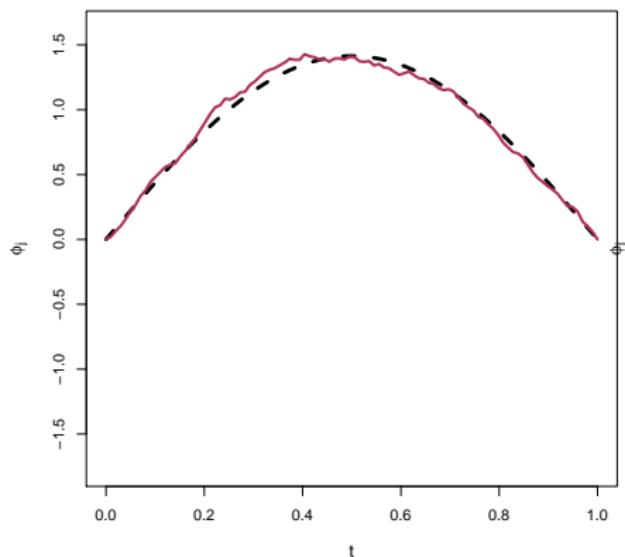
Profundidad

○○○○○○○○○○○○
○○○○○○○○
○○○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○○○
○○○○○○○○○○○○○○○○○○

Ejemplos: $\hat{\phi}$ para $X(t)$

 $j = 1$ $j = 2$ 

Estimación
○○○
○○○

Propiedades de $\hat{\Gamma}$
○○○○○
○○○○○○○○

Ejemplos
○○○○○○
●○○○○
○○○○○○○○○○

Mediana
○○

Profundidad
○○○○○○○○○○
○○○○○
○○○○○○○○○○○○

Boxplot Funcional
○○○○○○○○○○
○○○○○○○○○○○○

Estimadores suaves

Para obtener estimadores suaves de las autofunciones podemos usar **Bases**:

- $X_i(t) \approx \hat{X}_i(t) = \sum_{j=1}^p a_{ij}\phi_j(t)$
- Si la trayectoria $X_i(t)$ se observa en los puntos $t_1 < \dots < t_m$, los coeficientes a_{ij} se pueden elegir de modo a minimizar

$$\sum_{\ell=1}^m \left(X_i(t_\ell) - \sum_{j=1}^p a_{ij}\phi_j(t_\ell) \right)^2$$

ver detalles en Ramsay y Silverman (2005), Sección 3.2

- Se calcula luego el operador de covarianza muestral de $\hat{X}_i(t)$ y sus autofunciones que dan estimadores de las autofunciones de Γ .

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○●○○○○
○○○○○○○○○○○○

Mediana

○○

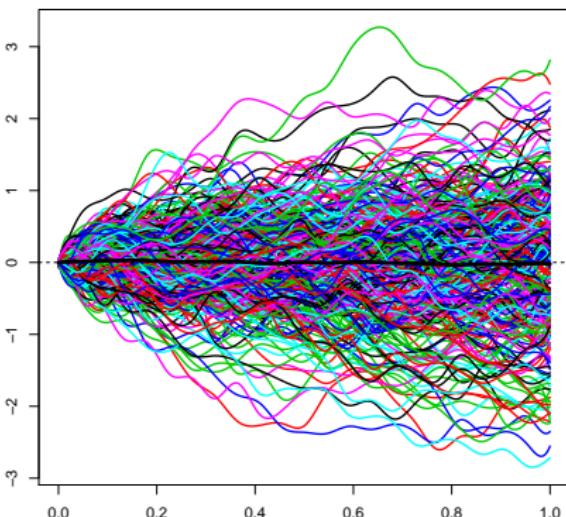
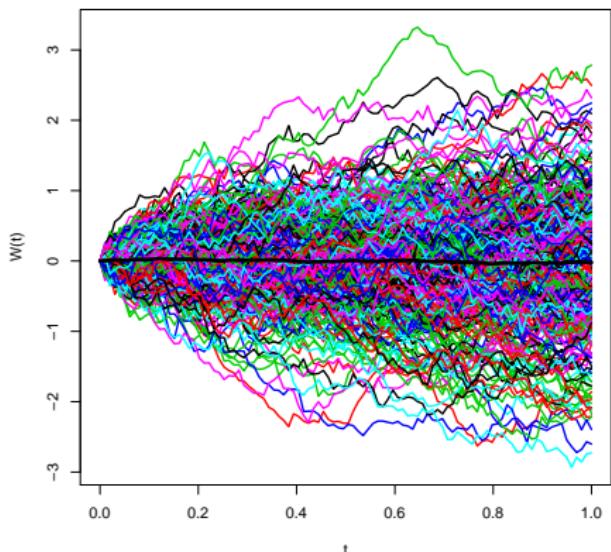
Profundidad

○○○○○○○○○○○○
○○○○○○○○
○○○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○○○

Ejemplo: Wiener

Wiener $W_i(t)$ $\widehat{W}_i(t)$ 

Estimación

6

Propiedades de $\widehat{\Gamma}$

A 2x5 grid of 10 circles, arranged in two rows of five.

Ejemplos

Mediana

○○

Profundidad

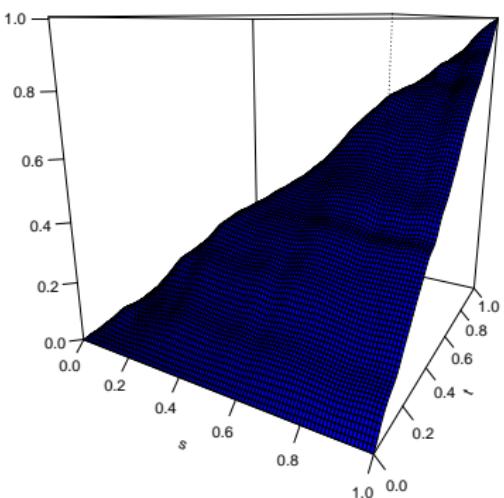
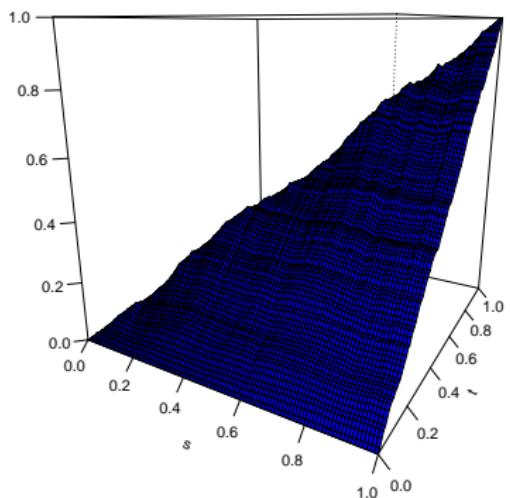
○○○○○○○○○○

Boxplot Funcional

Ejemplos: $\hat{\Gamma}$

Γ̂_W

۱۰۷



Estimación

800

100

Propiedades de $\widehat{\Gamma}$

•

100

Ejemplos

8888888

10

Mediana

○

Profundidad

0000000000

10 of 10

○○○○○○○○○○○○○○○○

Boxplot Funcional

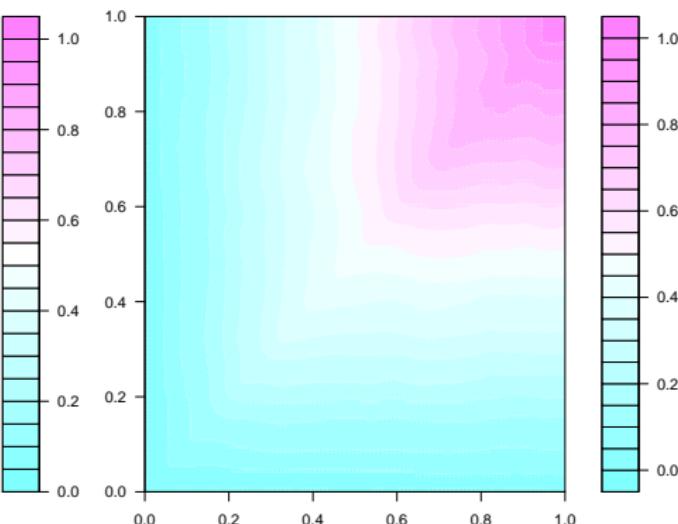
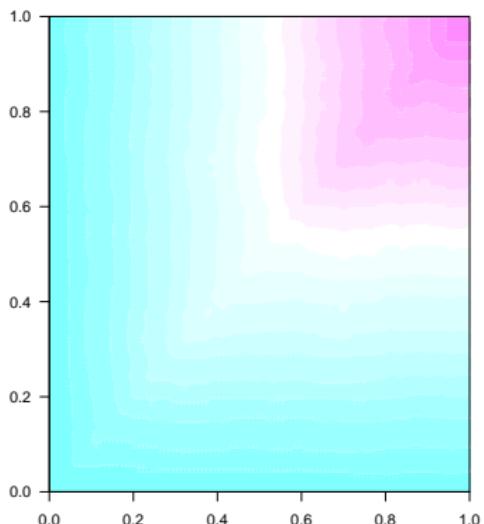
88888888

10 of 10

Ejemplos: $\hat{\Gamma}$

Γ̂_w

۱۷۸



Estimación

800

三

Propiedades de $\widehat{\Gamma}$

10

100

Ejemplos

8888888

○○○○○●

Mediana

Q

Profundidad

0000000000

○○○○○○

○○○○○○○○○○○○

Boxplot Funcional

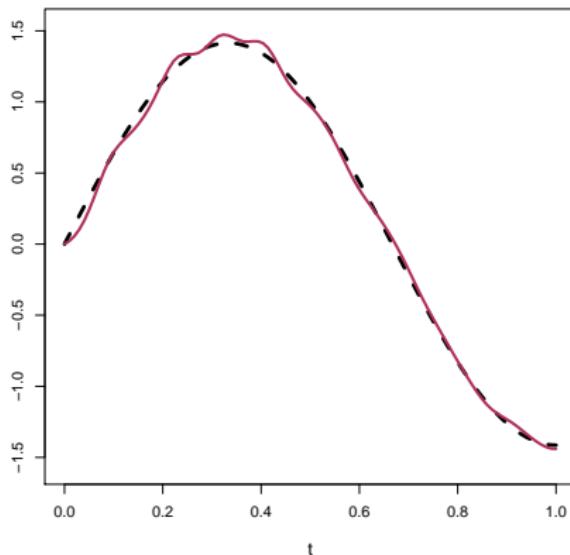
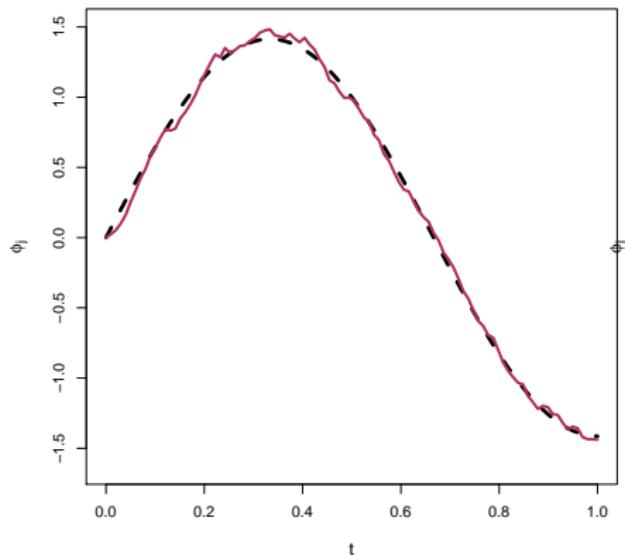
000000000

○○○○○○○○○○

$\widehat{\phi}_j$ para $W(t)$ y $\widehat{W}(t)$: $j = 2$

$$\hat{\phi}_{2,W}$$

$$\widehat{\phi}_{2,\widehat{W}}$$



Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
●○○○○○○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Estimadores suaves

Para obtener estimadores suaves de las autofunciones podemos usar **Suavizador basado en núcleos**

- Definamos

$$X_{i,h}(t) = \int K\left(\frac{t-s}{h}\right) X_i(s) ds$$

- Si la trayectoria $X_i(t)$ se observa en los puntos $t_1 < \dots < t_m$,

$$X_{i,h}(t) = \frac{\sum_{\ell=1}^m K\left(\frac{t-t_\ell}{h}\right) X_i(t_\ell)}{\sum_{\ell=1}^m K\left(\frac{t-t_\ell}{h}\right)}$$

donde $K : \mathbb{R} \rightarrow \mathbb{R}$, $K \geq 0$, $\int K(u) du = 1$, $\int K^2(u) du < \infty$

Estimadores suaves

Para obtener estimadores suaves de las autofunciones podemos usar **Suavizador basado en núcleos**

- Definamos

$$X_{i,h}(t) = \int K\left(\frac{t-s}{h}\right) X_i(s) ds$$

- Si la trayectoria $X_i(t)$ se observa en los puntos $t_1 < \dots < t_m$,

$$X_{i,h}(t) = \frac{1}{mh} \sum_{\ell=1}^m K\left(\frac{t - t_\ell}{h}\right) X_i(t_\ell)$$

donde $K : \mathbb{R} \rightarrow \mathbb{R}$, $K \geq 0$, $\int K(u)du = 1$, $\int K^2(u)du < \infty$

1

The diagram consists of two rows of small circles. The top row contains four circles arranged horizontally. Below it, the bottom row contains seven circles arranged horizontally.

○○

The diagram consists of three horizontal rows of small circles. The top row contains 10 circles arranged in a single line. The middle row contains 5 circles arranged in a single line. The bottom row contains 10 circles arranged in a single line.

Estimadores suaves

- El estimador del operador de covarianza se define como

$$\left(\widehat{\Gamma}_h u\right)(t) = \int_0^1 \widehat{\gamma}_h(t, s) u(s) ds \quad \text{for } u \in L^2(0, 1)$$

dónde

$$\hat{\gamma}_h(t, s) = \frac{1}{n} \sum_{i=1}^n (X_{i,h}(t) - \bar{X}_h(t)) (X_{i,h}(s) - \bar{X}_h(s)) .$$

- Observemos que $\hat{\gamma}_h(t, s)$ es una versión suavizada del núcleo de covarianza empírico $\hat{\gamma}(t, s)$

$$\widehat{\gamma}_h(t, s) = \int \int \overline{K}_h(t - u, s - v) \widehat{\gamma}(u, v) du dv$$

donde $\overline{K}_h(u, v) = K_h(u)K_h(v)$.

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○●○○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○○○

Propiedades

Boente, G. & Fraiman, R. (2000). Kernel-based functional principal components. *Statistics and Probability Letters*, **48**, 335-345.

Sean $\widehat{\phi}_{j,h}$ la autofunción de $\widehat{\Gamma}_h$ asociada al autovalor no nulo $\widehat{\lambda}_{jh}$.

- Si K es continuo, $\widehat{\phi}_{j,h}$ puede elegirse de modo que sea continua.
- Si K es Lipschitz (o sea, $|K(u) - K(v)| \leq A|u - v|$), $\widehat{\phi}_{j,h}$ puede elegirse de modo que sea Lipschitz.
- Supongamos que, para algún $A > 0$ y para todo $m \geq 2$,

$$\mathbb{E} (\|X_1\|^{2m}) \leq \frac{m!}{2} A^{m-2} \mathbb{E} (\|X_1\|^2)$$

entonces, si $h \rightarrow 0$,

$$\|\widehat{\Gamma}_h - \Gamma\|_{\mathcal{F}} \xrightarrow{a.s.} 0$$

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○●○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Propiedades: Boente, G. & Fraiman, R. (2000)

- $\widehat{\phi}_{j,h}$ la autofunción de $\widehat{\Gamma}_h$ asociada al autovalor no nulo $\widehat{\lambda}_{jh}$.
- Sea

$$\mathcal{M}_j = \{k \geq 1 : \lambda_k = \lambda_j\}$$

- Para cada $j \geq 1$, si λ_j tiene multiplicidad $m_j = \#\mathcal{M}_j$
 - a) existen m_j sucesiones $\{\widehat{\lambda}_{k,h}, k \in \mathcal{M}_j\}$ que convergen a λ_j , en casi todo punto.
 - b) En particular, si $m_j = 1$,

$$\widehat{\lambda}_{j,h} \xrightarrow{a.s.} \lambda_j$$

y

$$\text{signo} \left(\langle \widehat{\phi}_{j,h}, \phi_j \rangle \right) \widehat{\phi}_{j,h} \xrightarrow{a.s.} \phi_j \quad \text{en } L^2[0,1]$$

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○●○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Propiedades: Boente, G. & Fraiman, R. (2000)

- c) Más aún, si $m_j = 1$, y X_i es Gaussiano

$$\sqrt{n} \left(\widehat{\lambda}_{j,h} - \lambda_j \right) \xrightarrow{D} N(0, 2\lambda_j^2)$$

y

$$\sqrt{n} \left(\widehat{\phi}_{j,h} - \phi_j \right) \xrightarrow{D} N(0, \Upsilon_j)$$

donde Υ_j es el operador definido por

$$(\Upsilon_j v)(t) = \int_0^1 \rho_j(t, s)v(s)ds, \text{ con}$$

$$\rho_j(t, s) = \lambda_j \sum_{\substack{k=1 \\ k \neq j}}^{\infty} \frac{\lambda_k}{(\lambda_j - \lambda_k)^2} \phi_k(t) \phi_j(s).$$

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○●○○○○

Mediana

○○

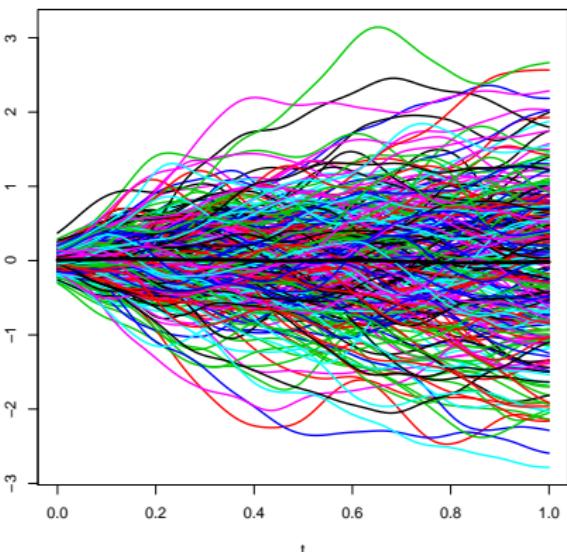
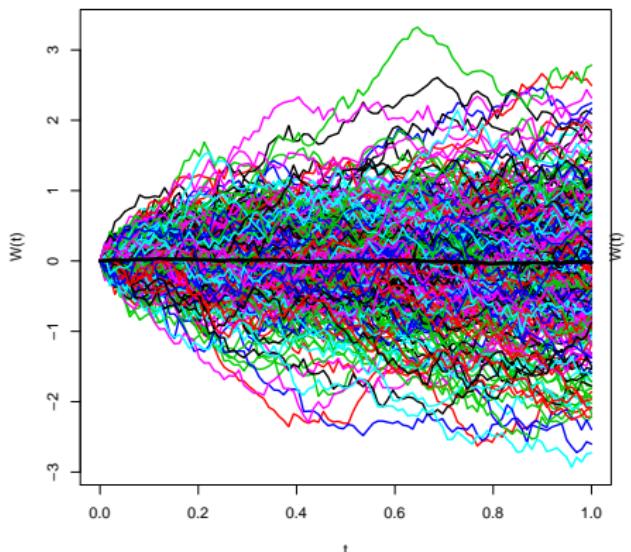
Profundidad

○○○○○○○○○○○○
○○○○○○○○
○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Ejemplo: Wiener

Wiener $W_i(t)$ $\widehat{W}_{i,h}(t)$ 

Estimación

10

Propiedades de $\widehat{\Gamma}$

Ejemplos

Mediana

○○

Profundidad

A horizontal row of ten small, empty circles arranged in a single line.

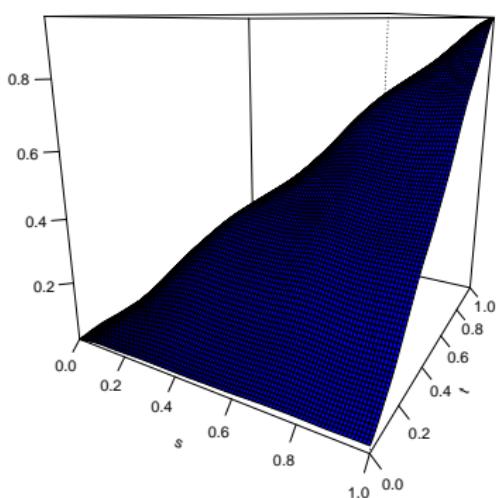
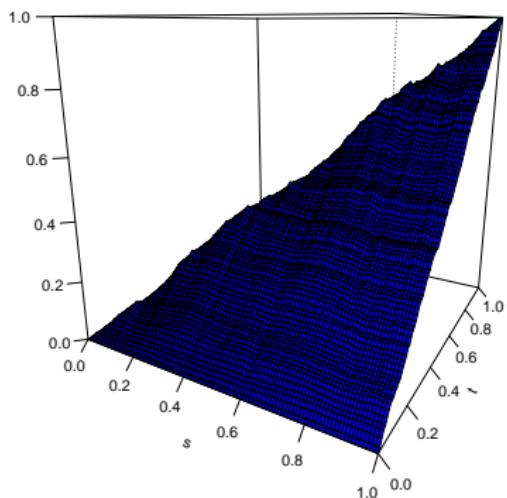
Boxplot Funcional

A 2x10 grid of 20 small circles, arranged in two rows of ten.

Wiener: $\hat{\Gamma}$ y $\hat{\Gamma}_h$

1

1



Estimación

10

Propiedades de $\widehat{\Gamma}$

A horizontal row of eight small, light-gray circles, arranged in a single line.

Ejemplos

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Mediana

OC

Profundidad

○○○○○○○○○○

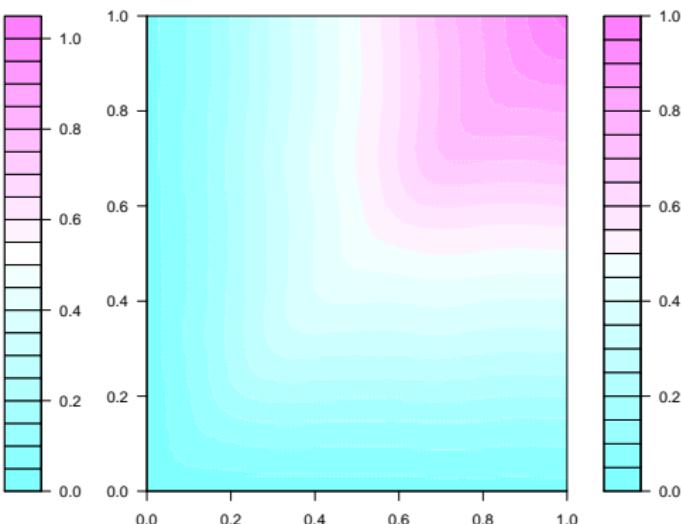
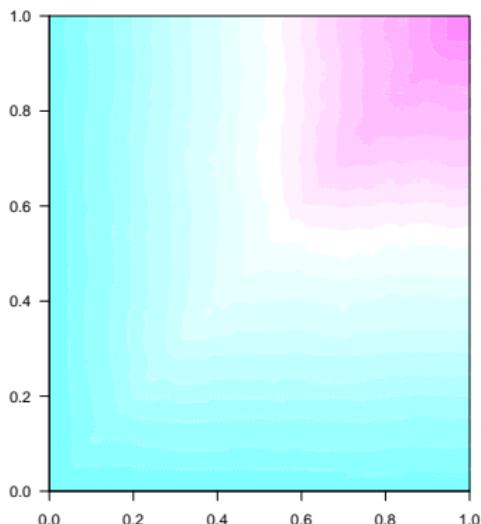
Boxplot Funcional

A 2x10 grid of 20 small circles, arranged in two rows of ten.

$$\hat{\Gamma} \text{ y } \hat{\Gamma}_h$$

六

$$\hat{\Gamma}_h$$



Estimación

10

Propiedades de $\widehat{\Gamma}$

A horizontal row of eight small, light-gray circles, evenly spaced.

Ejemplos

Mediana

○○

Profundidad

○○○○○○○○○○

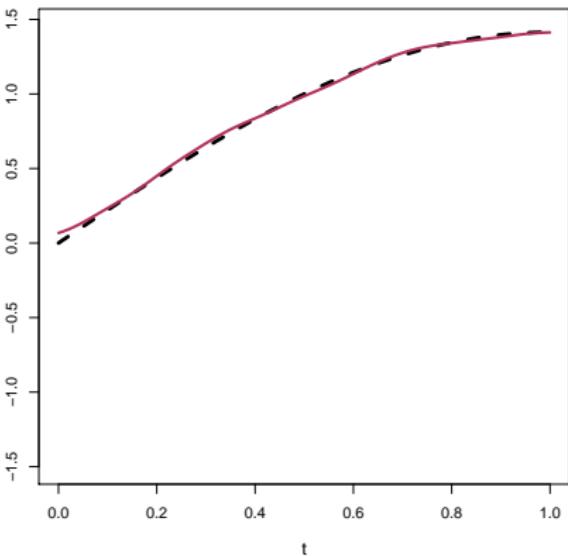
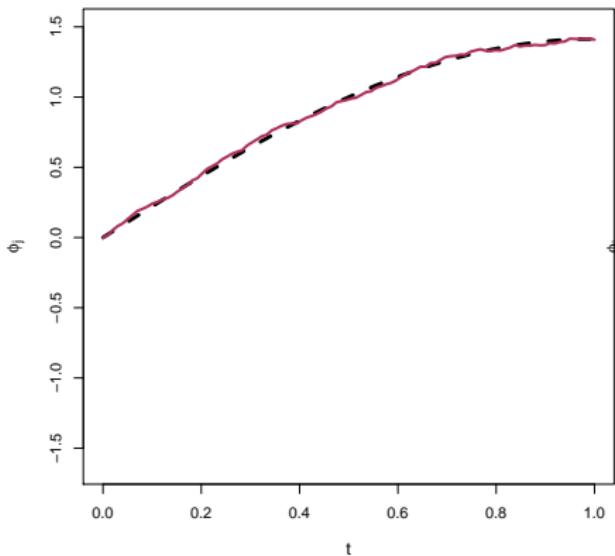
Boxplot Funcional

A 2x10 grid of 20 small circles, arranged in two rows of ten.

$$\hat{\phi}_j \text{ y } \hat{\phi}_{j,h}: j = 1$$

1

$$\hat{\phi}_{1,h}$$



Estimación

10

Propiedades de $\widehat{\Gamma}$

A horizontal row of seven small, light-gray circles arranged in a single line.

Ejemplos

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Mediana

○○

Profundidad

○○○○○○○○○○

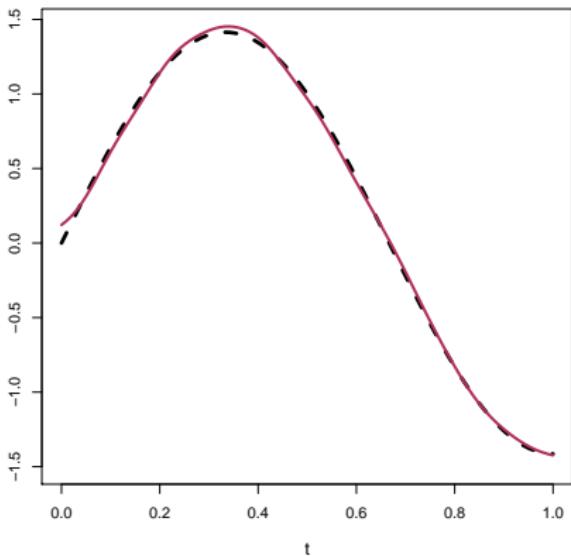
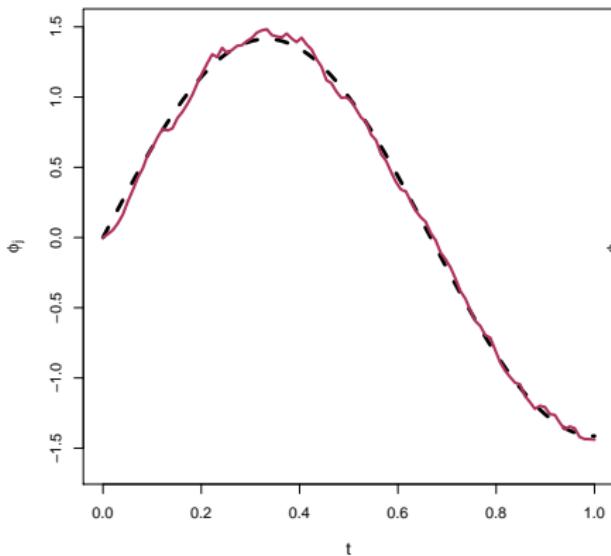
Boxplot Funcional

○○○○○○○○○○

$\hat{\phi}_j$ y $\hat{\phi}_{j,h}$: $j = 2$

$\hat{\phi}_2$

$$\hat{\phi}_{2,h}$$





Mediana

- La *mediana espacial* o *mediana geométrica funcional* $\mu_{GM} = \mu_{GM}(P)$ del elemento aleatorio $X \sim P$, se define como

$$\mu_{\text{GM}} = \operatorname*{argmin}_{u \in \mathcal{H}} \mathbb{E} (\|X - u\| - \|X\|) ,$$

- El estimador asociado es el valor $\hat{\mu}_{GM}$ tal que

$$\hat{\mu}_{\text{GM}} = \operatorname*{argmin}_{u \in \mathcal{H}} \sum_{i=1}^n (\|X_i - u\| - \|X_i\|) ,$$



Mediana

- Gervini (2008) probó que si μ_{GM} es única, entonces para todo $u \in \mathcal{H}$,

$$\langle \hat{\mu}_{\text{GM}}, u \rangle \xrightarrow{a.s.} \langle \mu_{\text{GM}}, u \rangle$$

- Cardot *et al.* (2013) (Teorema 3.1) muestran que

$$\|\hat{\mu}_{\text{GM}} - \mu_{\text{GM}}\| \xrightarrow{a.s.} 0$$

y dan un resultado de normalidad asintótica.

- Cardot, H., Cénac, P. & Zitt, P. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm, *Bernoulli*, **19**, 18-431.
 - Gervini, D. (2008). Robust functional estimation using the median and spherical principal components, *Biometrika*, **95**, 587-600.

Estimación

○○○
○○○Propiedades de $\hat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

Profundidad

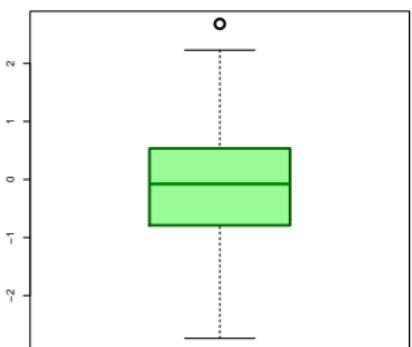
●○○○○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○○○

Boxplot Funcional

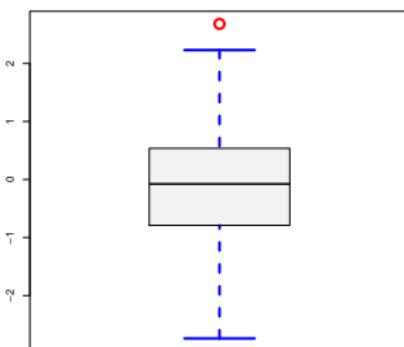
○○○○○○○○○○
○○○○○○○○○○○○○○○○

Boxplot

Representa la distribución de una muestra **univariada** y **unimodal**.



(a) Mediana y caja.



(b) Bigotes y outliers

Exhibe características de la distribución a partir de estimadores robustos. Utiliza la noción de profundidad que en \mathbb{R} se define como

$$\min(F_n(x), 1 - F_n(x))$$

Estimación

1

Propiedades de $\widehat{\Gamma}$

○○○○
○○○○○○○

Ejemplos

A diagram consisting of three rows of circles. The top row contains 5 circles, the middle row contains 4 circles, and the bottom row contains 7 circles, totaling 16 circles.

Mediana

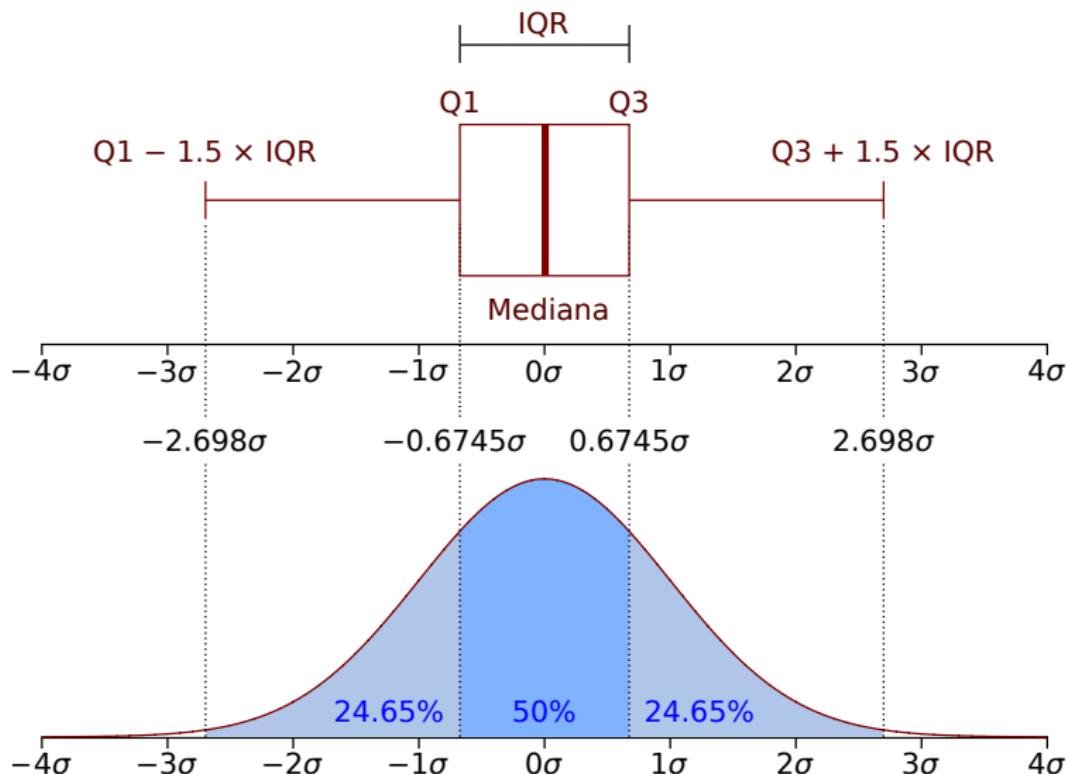
00

Profundidad

Boxplot Funcional

A 2x8 grid of circles, arranged in two rows of eight circles each.

El factor 1.5: Relación entre las partes del boxplot y la densidad $\mathcal{N}(0, 1)$



Estimación

○○○
○○○Propiedades de $\hat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

Profundidad

○○●○○○○○○○○
○○○○○○
○○○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○○○
○○○○○○○○○○○○○○○○○○

Profundidad

La medida de profundidad mide cuan profundo está un dato en la muestra

El boxplot utiliza la profundidad de semi-espacios de la muestra
 $X_1, \dots, X_n \in \mathbb{R}$

$$hsprof(x) = \min(F_n(x), 1 - F_n(x))$$

- Si x es la mediana, entonces

$$F_n(x) = \frac{1}{2} \quad hsprof(x) = \frac{1}{2}$$

que es la mayor profundidad posible.

- Si $x = X^{(n)} = \max_i X_i$, entonces

$$F_n(x) = 1 \quad hsprof(x) = 0$$

que es la menor profundidad posible.

Estimación
○○○
○○○

Propiedades de $\widehat{\Gamma}$
○○○○○
○○○○○○○○

Ejemplos
○○○○○○
○○○○○○
○○○○○○○○○○

Mediana
○○

Profundidad
○○○●○○○○○○
○○○○○○
○○○○○○○○○○○○○○

Boxplot Funcional
○○○○○○○○○○
○○○○○○○○○○○○○○

Profundidad

Otras medidas posibles de profundidad son la profundidad simplicial

$$SD_n(x) = F_n(x) \{1 - F_n(x^-)\}$$

y la profundidad

$$D_n(x) = 1 - \left| \frac{1}{2} - F_n(x) \right|$$

- Si x es la mediana, entonces $F_n(x) = \frac{1}{2}$ y $D_n(x) = 1$, que es la mayor profundidad posible.
- Si $x = X^{(n)} = \max_i X_i$, entonces $F_n(x) = 1$ y $D_n(x) = \frac{1}{2}$, que es la menor profundidad posible.
- $D_n(x) = 3/2 - hsprof(x)$.

Profundidad en \mathbb{R}^p : Tukey (1975)

- La profundidad de semi-espacios $\text{iprof}(\mathbf{x}, \mathcal{X})$ de un punto \mathbf{x} respecto de una muestra bivariada $\mathcal{X} = \{\mathbf{x}_i = (x_{i,1}, x_{i,2})^T\}_{i=1}^n$ es igual a

$$\text{hsprof}(\mathbf{x}, \mathcal{X}) = \min_{\mathcal{S}_{\mathbf{x}}} \{\#\{\mathbf{x}_i \in \mathcal{S}_{\mathbf{x}}\}\}$$

S_x es un semi-espacio cerrado cuya frontera pasa por x .

- La región de profundidad por lo menos k , \mathcal{D}_k , es

$$\mathcal{D}_k = \{\mathbf{x} : \text{hsprof}(\mathbf{x}, \mathcal{X}) \geq k\}$$

\mathcal{D}_k son polígonos convexos $\mathcal{D}_{k+1} \subset \mathcal{D}_k$

Estimación

10

Propiedades de Γ

○○○○
○○○○○○○○

Ejemplos

A diagram consisting of three rows of small circles. The top row contains 5 circles, the middle row contains 4 circles, and the bottom row contains 6 circles, totaling 15 circles.

Mediana

OC

Profundidad

Boxplot Funcional

○○○○○○○○

Mediana: Donoho y Gasko (1992)

En el caso multivariado, la noción de mediana puede generalizarse

- La mediana de Tukey θ^* se define como el valor con máxima profundidad si es único

En caso contrario, θ^* se define como el centro de gravedad de la región con mayor profundidad.

Estimación

10

Propiedades de Γ

Ejemplos

A diagram consisting of three rows of small circles. The top row contains 5 circles, the middle row contains 4 circles, and the bottom row contains 6 circles, totaling 15 circles.

Mediana

OC

Profundidad

○○○○●○○○○

Boxplot Funcional

○○○○○○○○

Bagplot: Rousseeuw, Ruts and Tukey (1999)

El bagplot es un gráfico bidimensional.

- Su componente principal es la **bolsa (bag)** que contiene el 50% de las observaciones.
 - Dentro de la **bolsa** está la mediana de Tukey
 - El gráfico contiene una **cerca (fence)**, que separa los outliers de las demás observaciones.
 - El gráfico contiene también un **bucle (loop)**, donde yacen las observaciones que no están en la bolsa pero que se encuentran dentro de la cerca.

Estimación

10

Propiedades de $\widehat{\Gamma}$

Ejemplos

A diagram consisting of three rows of open circles. The top row contains 6 circles, the middle row contains 5 circles, and the bottom row contains 8 circles, totaling 19 circles.

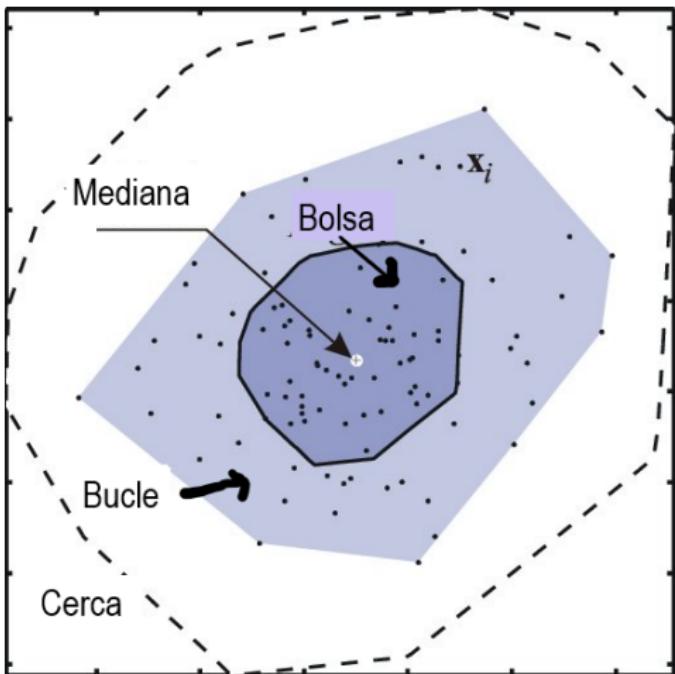
Mediana

○○

Profundidad

Boxplot Funcional

The diagram consists of two rows of small circles. The top row contains 8 circles arranged horizontally. The bottom row contains 9 circles, also arranged horizontally, positioned directly below the top row.



Estimación

Propiedades de $\hat{\Gamma}$

Ejemplos

Mediana

Profundidad

Boxplot Funcional

1

○○○○○

○○○○○○○○

○○○○○○○
○○○○○○○
○○○○○○○

OC

○○○○○○○○○○

```
library(mrfDepth)
```

```
cardata =structure(as.integer( c(2560,2345,1845,2260,2440, 2285, 2275, 2350, 2295,
1900, 2390, 2075, 2330, 3320, 2885, 3310, 2695, 2170, 2710, 2775, 2840, 2485, 2670,
2640, 2655, 3065, 2750, 2920, 2780, 2745, 3110, 2920, 2645, 2575, 2935, 2920, 2985,
3265, 2880, 2975, 3450, 3145, 3190, 3610, 2885, 3480, 3200, 2765, 3220, 3480, 3325,
3855, 3850, 3195, 3735, 3665, 3735, 3415, 3185, 3690, 97, 114, 81, 91, 113, 97, 97,
98, 109, 73, 97, 89, 109, 305, 153, 302, 133, 97, 125, 146, 107, 109, 121, 151, 133,
181, 141, 132, 133, 122, 181, 146, 151, 116, 135, 122, 141, 163, 151, 153, 202, 180,
182, 232, 143, 180, 180, 151, 189, 180, 231, 305, 302, 151, 202, 182, 181, 143, 146,
146)), .Dim = as.integer(c(60, 2)), .Dimnames = list(NULL, c("Weight", "Disp."))))
```

```
bcar=compBagplot(cardata, type = "hdepth", sizesubset = 500)
```

```
bagplot(bcarr, colorbag =NULL, colorloop = NULL, colorchull = NULL, databag =  
TRUE, dataloop = TRUE, plot.fence = FALSE)
```

```
bagplot(bcar, colorbag =NULL, colorloop = NULL, colorchull = NULL, databag = TRUE, dataloop = TRUE, plot.fence = TRUE)
```

Estimación

○○○
○○○

Propiedades de $\hat{\Gamma}$

○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

Profundidad

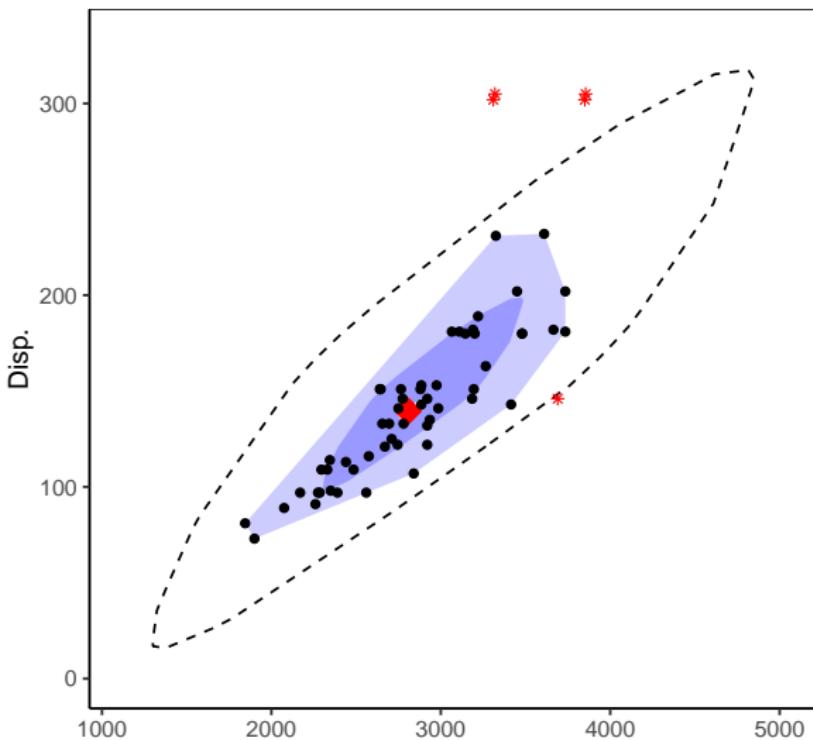
○○○○○○○○○●○
○○○○○○
○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Bagplot del Car Data

Bagplot based on halfspace depth



Estimación

10

Propiedades de $\widehat{\Gamma}$

Ejemplos

Mediana

○○

Profundidad

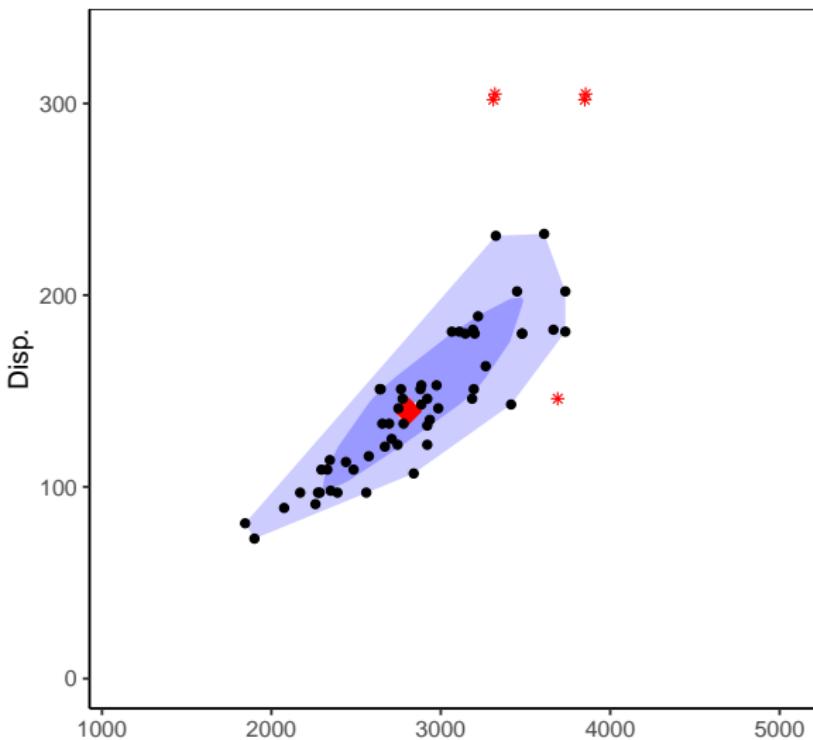
10

Boxplot Funcional

A 2x10 grid of 20 empty circles, arranged in two rows of ten.

Bagplot del Car Data

Bagplot based on halfspace depth





Profundidad de datos funcionales

Como en el caso multivariado, hay muchas maneras de definir profundidades.

Un review de varias medidas puede verse en

- Cuevas, A., Febrero-Bande, M. & Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Computational Statistics*, **22**, 481-496.
 - Febrero, M., Galeano, P. & González-Manteiga, W. (2008). Outlier detection in functional data by depth measures with application to identify abnormal NO_x levels. *Environmetrics*, **19**, 331-345.
 - Nieto-Reyes, A. & Battey, H. (2016). A topological valid definition of depth for functional data. *Statistical Science*, **31**, 61-79.

Nieto-Reyes & Battey (2016) formalizan la noción de profundidad funcional y estudian las propiedades de las ya definidas.

Estimación

○○○

3

Propiedades de $\hat{\Gamma}$

-

○○○○○○○○○○

Ejemplos

5 - 1

— 6 —

○○○○○○○○○○

Mediana

○

Profundidad

○○○○○○○○○○

10

○○○○○○○○○○○○○○

Boxplot Funcional

100

○○○○○○○○○○

Profundidad de datos funcionales

Las medidas de profundidad son útiles para

- definir estimadores robustos de posición y del operador de dispersión.
 - dar procedimientos de clasificación
 - detección de outliers.

Estimación

Propiedades de Γ

Ejemplos

Mediana

Profundidad

Boxplot Funcional

10

A 2x5 grid of circles, representing the second step in a process of forming a larger rectangle.

A 3x8 grid of 24 small circles, arranged in three rows and eight columns.

OC

A horizontal row of twelve empty white circles with black outlines. The fourth circle from the left is filled black.

○○○○○○○○○○

Profundidad Integrada de Fraiman & Muñiz (2001)

$$X_i : \Omega \rightarrow L^2([0, 1]), \quad 1 \leq i \leq n$$

Definamos la distribución empírica marginal como

$$F_{n,t}(a) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}(X_i(t) \leq a) \quad a \in \mathbb{R}$$

Sea D una profundidad univariada definida sobre \mathbb{R} y D_n la versión empírica.

La profundidad de $x(t)$ respecto de $X_1(t), \dots, X_n(t)$, está dada por

$$D_n(x(t))$$

Estimación

Propiedades de Γ

Ejemplos

Mediana

Profundidad

Boxplot Funcional

10

The diagram consists of two rows of small circles. The top row contains 5 circles arranged horizontally. The bottom row contains 7 circles arranged horizontally, positioned directly below the top row.

OC

○○○○○○○○○○

Profundidad Integrada de Fraiman & Muñiz (2001)

Esto permite definir la profundidad funcional integrada de $x \in L^2([0, 1])$ como

$$I_n(x) = \int_0^1 D_n(x(t)) dt$$

Una posible elección es

$$D_n(x(t)) = 1 - \left| \frac{1}{2} - F_{n,t}(x(t)) \right|$$

La función `depth.FM` de la librería `fda.usc` permite su cálculo.

Estimación
○○○
○○○

Propiedades de $\widehat{\Gamma}$
○○○○○
○○○○○○○○

Ejemplos
○○○○○○
○○○○○○
○○○○○○○○○○

Mediana
○○

Profundidad
○○○○○○○○○○○○
○○○○●○
○○○○○○○○○○○○

Boxplot Funcional
○○○○○○○○○○
○○○○○○○○○○○○

Profundidad Modal

Dado un dato funcional x_i , cuán densamente está por las otras observaciones de la muestra de datos funcionales?

- La h -profundidad (Cuevas, Febrero & Fraiman, 2007) en $x \in \mathcal{H}$ con respecto a la medida de probabilidad $X \sim P$ es

$$D_h(x, P) = \mathbb{E} K_h(\|x - X\|)$$

donde

$$K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$$

donde $K(t)$ es un núcleo.

- La moda funcional es el dato funcional que está más densamente rodeado por otros, o sea,

$$\mu_{\text{MODA}} = \underset{x}{\operatorname{argmax}} D_h(x, P)$$

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○●
○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○○○○○

Profundidad Modal

Dado un dato funcional x_i , cuán densamente está por las otras observaciones de la muestra de datos funcionales?

- La versión muestral de $D_h(x, P)$ es

$$D_h(x, P_n) = \frac{1}{n} \sum_{i=1}^n K_h(\|x - X_i\|)$$

- Elección de h

En la función `depth.mode` de la librería `fda.usc`, h se elige por defecto como el cuartil 15% de

$$\|X_j - X_k\|$$

$$j, k = 1, \dots, n, j \neq k.$$

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○○○○○○○○○
●○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○○○
○○○○○○○○○○○○

Profundidad de bandas: López Pintado & Romo (2009)

Se basa en la representación gráfica de los datos y de las bandas que determinan en el plano.

$$X_i : \Omega \rightarrow \mathcal{C}(\mathcal{I}), 1 \leq i \leq n, \mathcal{I} = [0, 1].$$

- Dada $x \in \mathcal{C}([0, 1])$

$$G(x) = \{(t, x(t)) : t \in \mathcal{I}\}$$

- La banda en \mathbb{R}^2 determinada por k de curvas, x_1, \dots, x_k , es

$$\begin{aligned} V(x_1, x_2, \dots, x_k) &= \{(t, y) : t \in \mathcal{I}, \min_{r=1, \dots, k} x_r(t) \leq y \leq \max_{r=1, \dots, k} x_r(t)\} \\ &= \{(t, y) : t \in \mathcal{I}, y = \alpha_t \min_{r=1, \dots, k} x_r(t) + (1 - \alpha_t) \max_{r=1, \dots, k} x_r(t), \\ &\quad \alpha_t \in [0, 1]\} \end{aligned}$$

Estimación

○○○
○○○

Propiedades de $\hat{\Gamma}$

○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

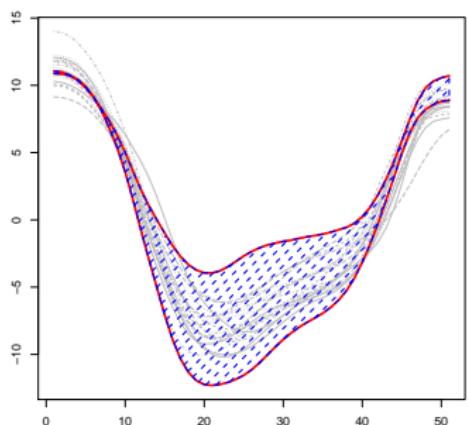
Profundidad

○○○○○○○○○○○○
○○○○○○
○●○○○○○○○○○○○○

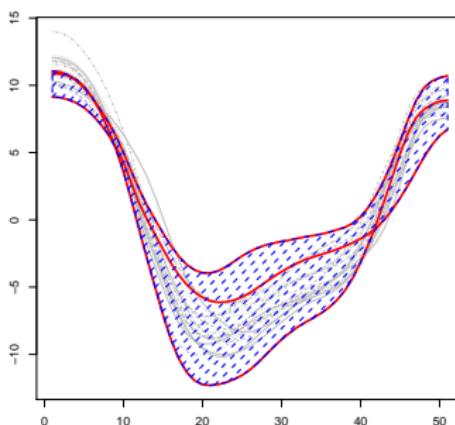
Boxplot Funcional

○○○○○○○○○○
○○○○○○○○○○○○

Profundidad de bandas: López Pintado & Romo (2009)



(a) Banda formada por dos curvas



(b) Banda formada por tres curvas

Las funciones x_i se muestran en gris

Estimación

○○○
○○○Propiedades de $\widehat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○○○○○
○○●○○○○○○○○○○○○

Boxplot Funcional

○○○○○○○○○○○○
○○○○○○○○○○○○○○○○○○

Profundidad de bandas: López Pintado & Romo (2009)

La profundidad de una curva en la muestra se medirá en términos de cuántas bandas la contienen

- Para cualquier función $x \in \{X_1, \dots, X_n\}$ y $j \geq 2$

$$BD_n^{(j)}(x) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_j \leq n} \mathbb{I}\{G(x) \subset V(X_{i_1}, X_{i_2}, \dots, X_{i_j})\}$$

- $BD_n^{(j)}(x)$ es la proporción de bandas formadas por j curvas que contienen el gráfico de la curva x .
- Se define la **profundidad de banda** de una función x como

$$BD_{n,J}(x) = \sum_{j=2}^J BD_n^{(j)}(x), \quad J \geq 2.$$

Estimación

○○○
○○○

Propiedades de $\widehat{\Gamma}$

The diagram consists of two rows of small circles. The top row contains 5 circles arranged horizontally. The bottom row contains 7 circles arranged horizontally, positioned below the top row.

Ejemplos

A diagram consisting of three rows of small circles. The top row contains 5 circles, the middle row contains 4 circles, and the bottom row contains 6 circles, totaling 15 circles.

Mediana

00

Profundidad

A 3x8 grid of circles. The bottom-middle circle is filled black, while all other circles are unfilled.

Boxplot Funcional

100

○○○○○○○○○○

Profundidad de bandas: López Pintado & Romo (2009)

- Se puede ver que

$$BD_n^{(j)}(x) = \binom{n}{k}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_j \leq n} \mathbb{I}\{x \in \mathcal{B}(X_{i_1}, X_{i_2}, \dots, X_{i_j})\}$$

donde

$$\mathcal{B}(X_{i_1}, X_{i_2}, \dots, X_{i_j}) = \{y \in \mathcal{C}(\mathcal{I}) : y(t) = \sum_{\ell=1}^j \alpha_\ell X_{i_\ell}(t),$$

$$\sum_{\ell=1}^j \alpha_\ell = 1 \quad \alpha_\ell \geq 0 \}$$

Estimación

○○○
○○○

Propiedades de $\hat{\Gamma}$

The diagram consists of two rows of five empty circles each. The top row has its circles centered, and the bottom row has its circles shifted to the right, creating a staggered effect.

Ejemplos

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Mediana

00

Profundidad

100

Boxplot Funcional

A horizontal row of ten small, light-colored circles arranged in a single line.

Profundidad de bandas: Versión poblacional

X_1, X_2, \dots, X_n i.i.d., $X_i \sim X$

$$BD^{(j)}(x) = \mathbb{P} \left\{ G(x) \subset V(X_{i_1}, X_{i_2}, \dots, X_{i_j}) \right\}$$

$$BD_J(x) = \sum_{j=2}^J BD^{(j)}(x) = \sum_{j=2}^J \mathbb{P}\left\{G(x) \subset V(X_{i_1}, X_{i_2}, \dots, X_{i_j})\right\}.$$

Las definiciones anteriores dependen de J que es la cantidad máxima de curvas que generan una banda.

López-Pintado y Romo (2009) recomiendan trabajar con BD_3 , o sea, $J = 3$.

10

The diagram consists of two rows of small circles. The top row contains 5 circles arranged horizontally. Below it, the bottom row contains 7 circles arranged horizontally.

A diagram consisting of three rows of small circles. The top row contains 6 circles arranged horizontally. The middle row contains 5 circles arranged horizontally. The bottom row contains 8 circles arranged horizontally.

○○

○○○○○○○○○○

Mediana

- Definimos la mediana muestral como la observación $\hat{m}_{n,J}$ con mayor profundidad

$$\hat{m}_{n,J} = \operatorname*{argmax}_{x \in \{X_1, \dots, X_n\}} BD_{n,J}(x) .$$

- La mediana poblacional es la función $m_J \in \mathcal{C}(\mathcal{I})$ que maximiza $BD_J(\cdot)$

Estimación

○○○
○○○Propiedades de $\hat{\Gamma}$ ○○○○○
○○○○○○○○

Ejemplos

○○○○○○○
○○○○○○
○○○○○○○○○○○○

Mediana

○○

Profundidad

○○○○○○○○○○○○
○○○○○○○○○○○○
○○○○○●○○○○○○

Boxplot Funcional

○○○○○○○○○○○○
○○○○○○○○○○○○

Profundidad de bandas modificada (MBD): López Pintado & Romo (2009)

- Cuando las curvas son muy irregulares, pocas bandas contendrán por completo una curva.
- Varias curvas de la muestra tendrán el mismo valor de profundidad, lo cual resulta en un ordenamiento pobre de la muestra con muchos empates.
- Por ejemplo, la banda definida por dos curvas ($J = 2$) que se corten en un punto, con probabilidad 1 no contendrá ninguna otra curva y no contribuirá al valor de la profundidad.

La definición anterior es restrictiva.

- **Causa:** Usar la función indicadora.
- **Solución:** Medir el conjunto donde la función queda contenida en la correspondiente banda.

Profundidad de bandas modificada (MBD)

- Para cualquier función x y $j \geq 2$

$$A_{i_1, \dots, i_j}(x) = \left\{ t \in I : \min_{r=i_1, \dots, i_j} X_r(t) \leq x(t) \leq \max_{r=i_1, \dots, i_j} X_r(t) \right\},$$

el conjunto de puntos del intervalo \mathcal{I} donde la función x está dentro de la banda determinada por las observaciones X_{i_1}, \dots, X_{i_j}

- Sea λ la medida usual (de Lebesgue) en \mathcal{I} ,

$$\lambda_r(A_{i_1, \dots, i_j}(x)) = \frac{\lambda(A_{i_1, \dots, i_j}(x))}{\lambda(\mathcal{I})}$$

es la “proporción de tiempo” que x está en la banda.

Profundidad de bandas modificada (MBD)

- Si x está siempre dentro de la banda, $\lambda_r(A_{i_1, \dots, i_j}(x)) = 1$
- Definamos para $j \geq 2$

$$MBD_n^{(j)}(x) = \binom{n}{j}^{-1} \sum_{1 \leq i_1 \leq i_2 \leq \dots \leq i_j \leq n} \lambda_r(A_{i_1, \dots, i_j}(x))$$

- $MBD_n^{(j)}(x)$ es una generalización de $BD_n^{(j)}(x)$
- Dado $J \geq 2$, la profundidad de banda modificada de una curva x es

$$MBD_{n,J}(x) = \sum_{j=2}^J MBD_n^{(j)}(x).$$

Estimación

10

Propiedades de $\hat{\Gamma}$

○○○○
○○○○○○○

Ejemplos

A 3x8 grid of 24 small circles, arranged in three rows and eight columns.

Mediana

○○

Profundidad

Boxplot Funcional

○○○○○○○○

Profundidad de banda modificada: Versión poblacional

X_1, X_2, \dots, X_n i.i.d., $X_i \sim X$

$$MBD^{(j)}(x) = \mathbb{E} \left(\lambda_r \left(A(x; X_{i_1}, X_{i_2}, \dots, X_{i_j}) \right) \right), \quad \text{si } j \geq 2$$

$$MBD_J(x) = \sum_{j=2}^J MBD^{(j)}(x), \quad \text{si } J \geq 2$$

$$A(x; X_{i_1}, X_{i_2}, \dots, X_{i_j}) = \left\{ t \in I : \min_{r=i_1, \dots, i_j} X_r(t) \leq x(t) \leq \max_{r=i_1, \dots, i_j} X_r(t) \right\},$$

Profundidad de banda modificada

- Las simulaciones dadas en López-Pintado y Romo (2009), muestran que el orden inducido por esta definición es estable cuando aumenta J .

Para evitar problemas computacionales, recomiendan $J = 2$.

- Cuando $J = 2$, $MBD_J(x)$ coincide con la profundidad integrada si se usa la profundidad univariada simplicial, cuando $X \sim P$ y P tiene marginales absolutamente continuas. (ver Claeskens *et al.* (2014))

Claeskens, G., Hubert, M., Slaets, L. & Vakili, K.

(2014). Multivariate functional halfspace depth. *Journal American Statistical Association*, **109**, 411-423.

Estimación

10

Propiedades de Γ

Ejemplos

A diagram consisting of three rows of circles. The top row contains 5 circles, the middle row contains 4 circles, and the bottom row contains 7 circles, totaling 16 circles.

Mediana

OC

Profundidad

A diagram consisting of three horizontal rows of small circles. The top row contains 8 circles. The middle row contains 5 circles. The bottom row contains 8 circles.

Boxplot Funcional

○○○○○○○○

○○○○○○○○

BD y MBD

- Como la profundidad de banda modificada toma en cuenta la proporción de tiempo que una curva permanece dentro de una banda es más conveniente para obtener la curva más representativa en términos de magnitud.
 - La profundidad de banda depende más de la forma de las curvas dando más empates.

Estimación

10

Propiedades de Γ

A horizontal row of five small, hollow circles arranged in a single line.

Ejemplos

A diagram consisting of three rows of small circles. The top row contains 6 circles. The middle row contains 5 circles. The bottom row contains 8 circles.

Mediana

OC

Profundidad

A diagram consisting of three horizontal rows of small circles. The top row contains 8 circles, the middle row contains 5 circles, and the bottom row contains 8 circles, representing the numbers 8, 5, and 8 respectively.

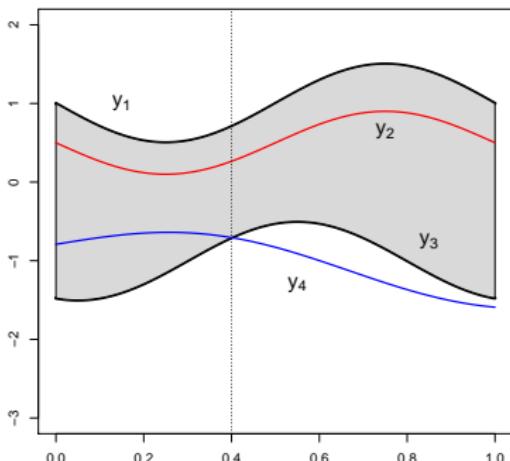
Boxplot Funcional

○○○○○○○○

BD y MBD

- La atipicidad de un dato, en el caso funcional, puede ser de dos tipos: de magnitud o forma. Los *outliers de magnitud* están lejos de la media y los (*outliers de forma*) tienen un patrón diferente de las otras curvas.
 - Una diferencia entre *BD* y *MBD* es su comportamiento ante curvas que abandonan el centro de la muestra en un intervalo corto, o sea permanecen en el interior de la muestra casi todo el tiempo pero tomando valores extremos en subintervalos pequeños: el valor de *MBD* sigue siendo grande mientras que *BD* es sensible a este tipo de datos y pasa a tomar valores pequeños.

Un ejemplo del cálculo de BD y MBD



- $BD(y_2) = 5/6 = 0.83$ ya que solo la banda delimitada por $y_3(t)$ e $y_4(t)$ no contiene completamente a la curva $y_2(t)$
- $BD(y_4) = 3/6 = 0.5$
- $BD(y_1) = BD(y_3) = 0.5$
- $y_4(t)$ solo pertenece a la banda pintada el 40% del tiempo, entonces

$$MBD(y_4) = (3 + 0.4 + 0.4)/6 = 0.63$$

- Para las otras dos curvas, $MBD(y_1) = 0.5$ y $MBD(y_3) = 0.7$.

Boxplot funcional

- El boxplot funcional **no se obtiene** de los boxplots puntuales en cada punto t .
 - Las nociones de profundidad presentadas permiten ordenar curvas definiendo estadísticos de orden y extender el boxplot a conjuntos de datos funcionales.
 - $X_i : \Omega \rightarrow L^2(\mathcal{I}) \cup \mathcal{C}(\mathcal{I})$, $1 \leq i \leq n$, $\mathcal{I} = [0, 1]$.
 - Sea $X_{[i]}(t)$ la observación asociada al i -ésimo mayor valor de profundidad.
 - $X_{[1]}(t), \dots, X_{[n]}(t)$ son los estadísticos de orden con $X_{[1]}(t)$ la curva más profunda (más central) y $X_{[n]}(t)$ la más exterior.

Boxplot funcional

- Buscamos el 50% de las observaciones más profundas:

$$C_{0,5} = \{(t, x(t)) : \min_{r=1, \dots, \lceil \frac{n}{2} \rceil} X_{[r]}(t) \leq x(t) \leq \max_{r=1, \dots, \lceil \frac{n}{2} \rceil} X_{[r]}\}$$

donde $\lceil \frac{n}{2} \rceil$ es el menor entero mayor a $\frac{n}{2}$.

- Esta región central es análoga a la determinada por la distancia intercuartil y da una medida robusta de la dispersión del 50% de las curvas más centrales.
 - La curva $X_{[1]}(t)$ es la mediana.

Estimación
ooo
ooo

Propiedades de $\hat{\Gamma}$

Ejemplos

○○○○○○
○○○○○○
○○○○○○○○○○○

Mediana

Profundidad
oooooooooooo
oooooo
oooooooooooo

Boxplot Funcional
○○●○○○○○○○○

Boxplot funcional

- Los bigotes del boxplot univariado son las líneas verticales del gráfico que se extienden desde la caja hasta la última observación que no es atípica.
 - Definimos la *región exterior* inflando la región central 1.5 veces su tamaño.
 - Cualquier curva fuera de estos límites se clasifica como un potencial outlier.

Si las funciones fueran constantes el boxplot funcional se reduce al boxplot univariado.

Estimación

10

Propiedades de $\widehat{\Gamma}$

A diagram consisting of two rows of small circles. The top row contains four circles arranged horizontally. Below it, the bottom row contains seven circles arranged horizontally.

Ejemplos

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Mediana

OC

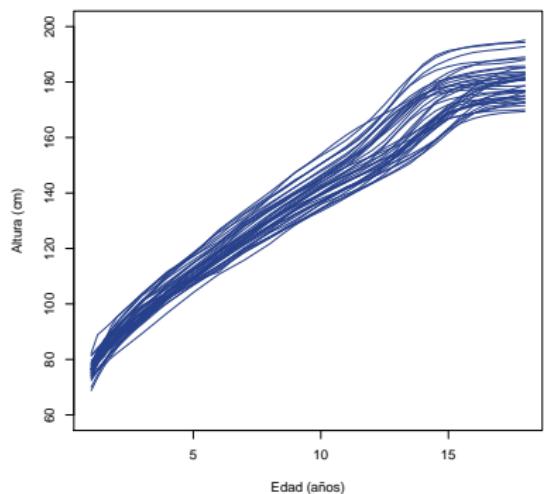
Profundidad

A diagram consisting of three rows of small circles. The top row contains 8 circles, the middle row contains 5 circles, and the bottom row contains 8 circles, representing the numbers 8, 5, and 8 respectively.

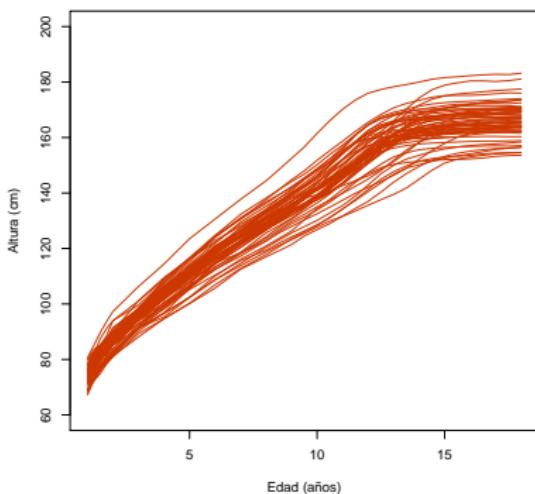
Boxplot Funcional

Altura de 39 niños y 54 niñas

Curvas para 39 varones



Curvas para 54 mujeres



Estimación

Propiedades de $\widehat{\Gamma}$

Ejemplos

Mediana

Profundidad

Boxplot Funcional

10

The diagram consists of two rows of small circles. The top row contains five circles arranged horizontally. Below it, the bottom row contains seven circles arranged horizontally. This visual representation serves as a concrete example for the mathematical concept of 'more than'.

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

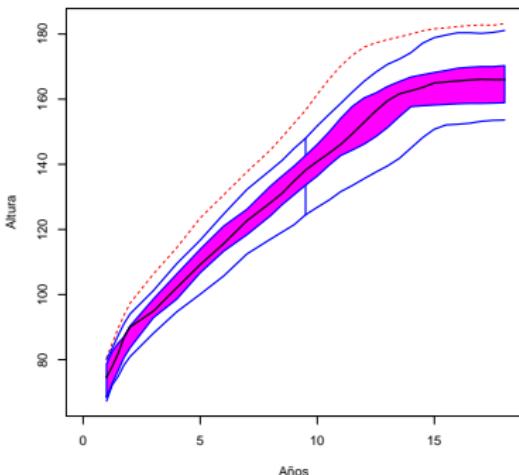
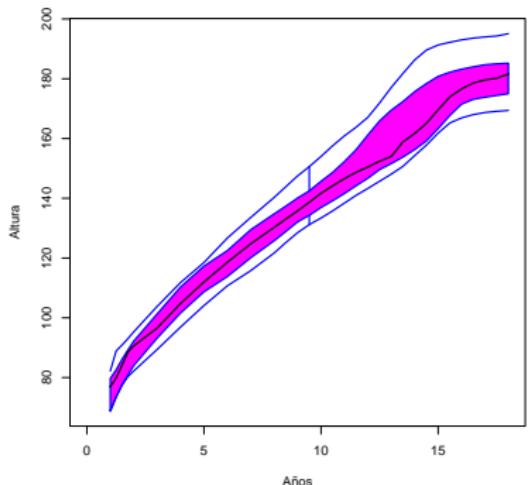
2

The diagram consists of three horizontal rows of small circles. The top row contains 8 circles. The middle row contains 5 circles. The bottom row contains 10 circles.

Boxplot funcional

Varones

Mujeres (con un outlier en rojo)



En negro se dibuja la mediana, la zona pintada es la región central. El dato atípico se representa en rojo.

```
fbplot(growth$hgtm, x = growth$age, ylab="Altura", xlab="Años", xlim=c(0,18),  
ylim=c( min(growth$hgtm), max(growth$hgtm)))  
fbplot(growth$hgtf, x = growth$age, ylab="Altura", xlab="Años", xlim=c(0,18),  
ylim=c( min(growth$hgtf), max(growth$hgtf)))
```

Estimación

Propiedades de $\widehat{\Gamma}$

10

Ejemplos

A 2x5 grid of 10 small circles, arranged in two rows of five.

○○○○○○○
○○○○○○○
○○○○○○○○○○

Mediana

○○

Profundidad

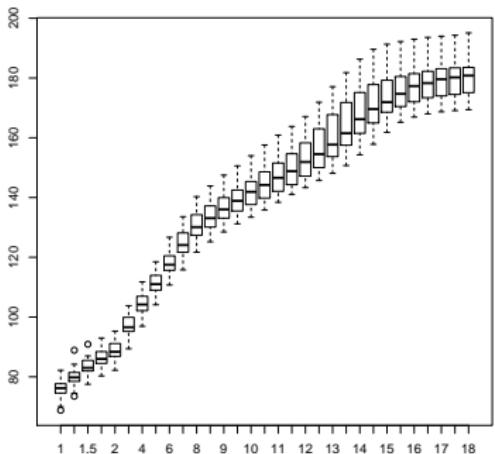
A horizontal row of ten small, light-colored circles arranged in two rows of five.

Boxplot Funcional

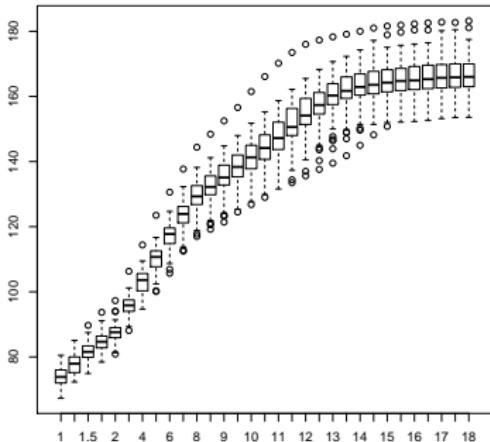
ANSWER

El boxplot funcional NO se obtiene de los boxplots puntuales en cada punto t .

Varones



Mujeres



Estimación

1

Propiedades de $\widehat{\Gamma}$

The diagram consists of two rows of small circles. The top row contains 5 circles arranged horizontally. Below it, the bottom row contains 7 circles also arranged horizontally.

Ejemplos

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Mediana

OC

Profundidad

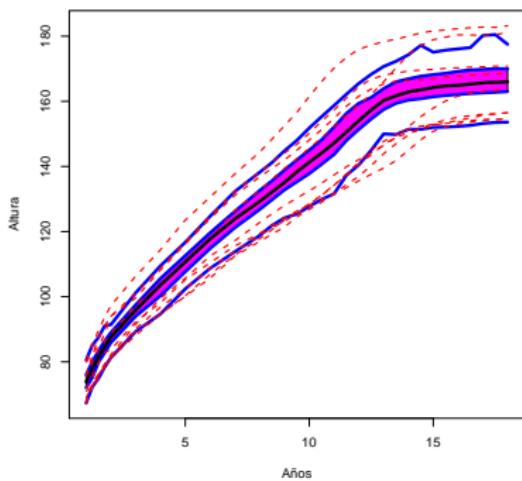
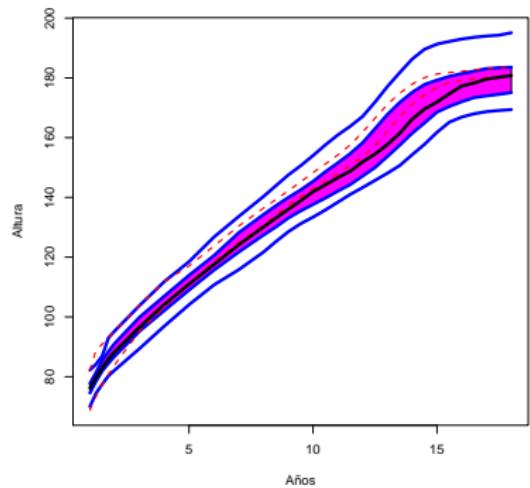
A 3x8 grid of circles, arranged in three rows and eight columns.

Boxplot Funcional

100

El boxplot funcional NO se obtiene de los boxplots puntuales en cada punto t .

Unión de boxplots puntuales



Estimación

10

Propiedades de $\widehat{\Gamma}$

Ejemplos

A diagram consisting of three rows of small circles. The top row contains 5 circles. The middle row contains 5 circles. The bottom row contains 7 circles.

Mediana

○○

Profundidad

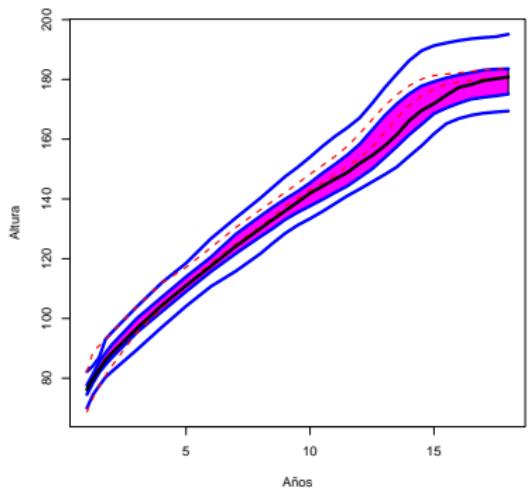
○○○○○○○○○○

Boxplot Funcional

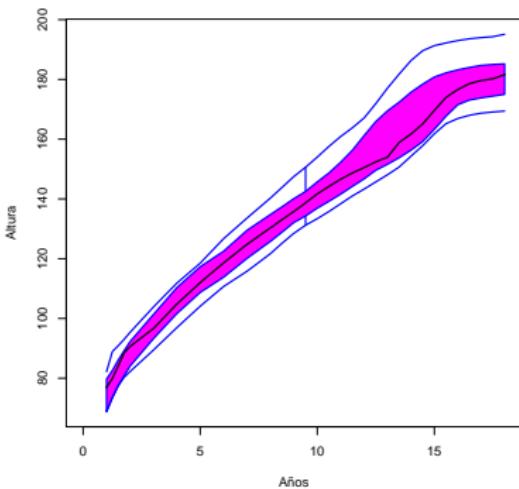
100

Varones

Boxplot puntual



Boxplot Funcional



Estimación

10

Propiedades de $\widehat{\Gamma}$

Ejemplos

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Mediana

○○

Profundidad

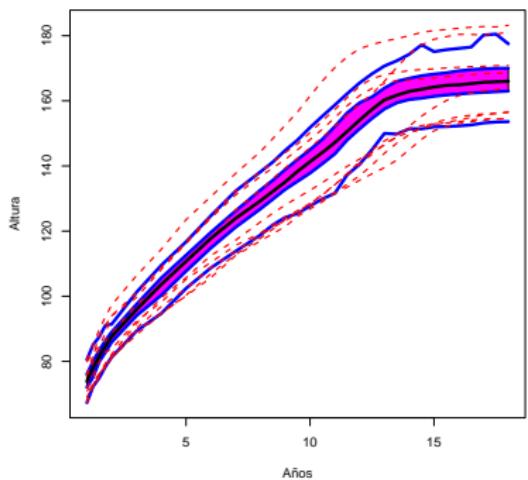
A 3x8 grid of 24 small circles, arranged in three rows and eight columns.

Boxplot Funcional

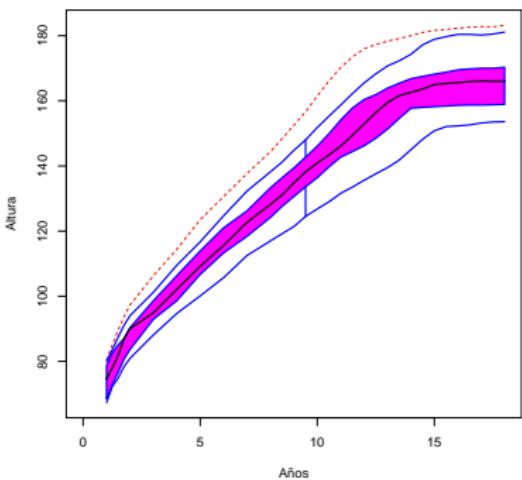
• 10

Mujeres

Boxplot puntual



Boxplot Funcional



Estimación

10

Propiedades de $\hat{\Gamma}$

○○○○
○○○○○○○○

Ejemplos

Mediana

OC

Profundidad

○○○○○○○○○○
○○○○○○
○○○○○○○○○○

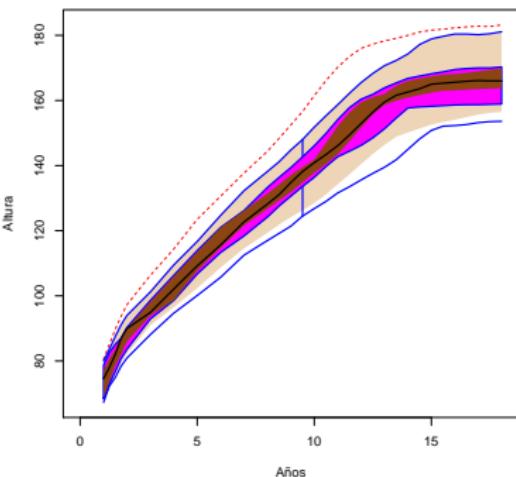
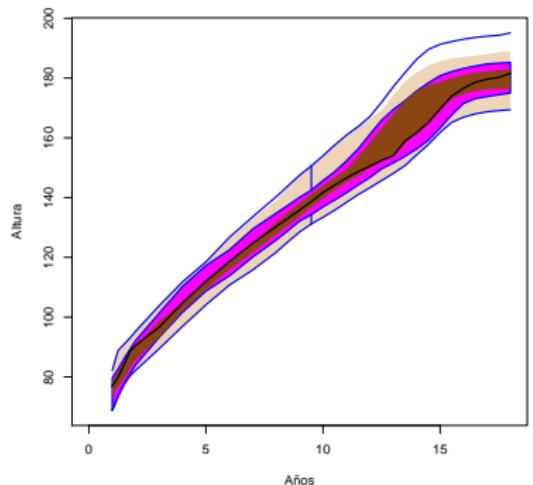
Boxplot Funcional

10

Boxplot funcional

Varones

Mujeres (con un outlier en rojo)



```
fbplot(growth$hgtf, x = growth$age, prob=c(0.9,0.5,0.2), color=c("bisque2", "magenta", "chocolate4"), ylab="Altura", xlab="Años", xlim=c(0,18), ylim=c(min(growth$hgtf), max(growth$hgtf)))
```

```
fbplot(growth$hgtm, x = growth$age, prob=c(0.9,0.5,0.2), color=c("bisque2", "magenta", "chocolate4"), ylab="Altura", xlab="Años", xlim=c(0,18), ylim=c(min(growth$hgtm), max(growth$hgtm)))
```

Estimación

Propiedades de Γ

Ejemplos

Mediana

Profundidad

Boxplot Funcional

10

The diagram consists of two rows of small circles. The top row contains 5 circles arranged horizontally. The bottom row contains 7 circles arranged horizontally, positioned directly below the top row.

OC

Detección de Outliers

Además de utilizar boxplot funcional se pueden detectar datos atípicos utilizando alguno de los procedimientos descriptos en

- Febrero, M., Galeano, P. & Gonzalez-Manteiga, W. (2007): A functional analysis of NOx levels: location and scale estimation and outlier detection. *Computational Statistics*, **22**, 411-427.
 - Febrero, M., Galeano, P. & Gonzalez-Manteiga, W. (2008): Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels. *Environmetrics*, **19**, 331-345.

función `foutliers` de la librería `rainbow` y
funciones `outliers.depth.trim`, `outliers.depth.pond`,
`outliers.thres.lrt` y `outliers.lrt` de la librería
`fda.usc`



Detección de Outliers

- Hyndman, R. & Ullah, S. (2007): Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, **51**, 4942-4956.
 - Hyndman, R. & Shang, H. (2010): Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, **19**, 29-45.

funciones `fboxplot` y `foutliers` de la librería `rainbow`

- Boente, G. & Salibián-Barrera, M. (2015): S -estimators for functional principal component analysis. *Journal of American Statistical Association*, **110**, 1100-1111.

<https://github.com/msalibian/S-FPCA>

Detección de Outliers: Febrero *et al.* (2007)

Procedimiento basado en un test de cociente de verosimilitud.

- Para una medida de profundidad dada sean

$$X_{(1)}, \dots, X_{(n)}$$

las curvas ordenadas de acuerdo a su profundidad, o sea, $X_{(1)}$ es el dato más profundo y $X_{(n)}$ el menos profundo.

- Definamos

$$O_\alpha(X_i) = \left\| \frac{X_i - \hat{\mu}_\alpha}{\hat{\sigma}_\alpha} \right\| \quad \Lambda = \max_{1 \leq i \leq n} O_\alpha(X_i)$$

donde $\|\cdot\|$ es la norma en L^1 , L^2 o L^∞

$$\widehat{\mu}_\alpha(t) = \frac{1}{n - [\alpha n]} \sum_{i=1}^{n - [\alpha n]} X_{(i)} \quad \widehat{\sigma}_\alpha^2(t) = \frac{1}{n - [\alpha n]} \sum_{i=1}^{n - [\alpha n]} (X_{(i)} - \widehat{\mu}_\alpha(t))^2$$

Detección de Outliers: Febrero *et al.* (2007)

- Procedimiento iterativo

- 1) Dada la muestra X_1, \dots, X_n , calcule Λ .
 - 2) Sea $X^{(M)}$ la observación que alcanza el máximo valor de Λ .

Si $\Lambda = O_\alpha(X^{(M)}) > C$, se declara a $X^{(M)}$ como outlier y se elimina de la muestra

Repita 1) y 2) hasta que no se detecten más outliers.

- Para determinar C se usa bootstrap suavizado

Detección de Outliers: Febrero *et al.* (2007)

- Para determinar C se usa bootstrap suavizado
 - a) Sea Y_1^b, \dots, Y_n^b , $1 \leq b \leq B$ las B muestras obtenidas por bootstrap suavizado.
 - b) Para cada b , calcule

$$\Lambda^b = \max_{1 \leq i \leq n - [\alpha]n} O_\alpha(Y_{(i)}^b)$$

donde

- $Y_{(1)}^b, \dots, Y_{(n)}^b$ son las curvas ordenadas de acuerdo a su profundidad
 - α es el valor usado para definir $\hat{\mu}_\alpha$ y $\hat{\sigma}_\alpha$

c) Defina

$$C = \max_{1 \leq b \leq B} \Lambda^b$$

Como se define la muestra por bootstrap suavizado

- Primero defina X_1^b, \dots, X_n^b , $1 \leq b \leq B$ las B muestras bootstrap usuales.
 - Supongamos que las curvas X_1, \dots, X_n se observan en la grilla de puntos t_1, \dots, t_p .
 - Defina

$$Y_i^b = X_i^b + Z_i^b$$

donde

- Z_1^b, \dots, Z_n^b son tales que

$$\mathbf{Z}_i^b = (Z_i^b(t_1), \dots, Z_i^b(t_p))^T \sim N_p(\mathbf{0}, \gamma \boldsymbol{\Sigma})$$

- Σ : matriz de covarianza muestral de $(X_i(t_1), \dots, X_i(t_p))^T$,
 $1 \leq i \leq n$
 - γ : parámetro de suavizado

Mujeres: **Outliers detectados**

- Boxplot Funcional: Observación 8
 - Febrero *et al.* (2007): Observación 8 ($\alpha = 0.1$, $\gamma = 0.05$, $B = 200$)
 - Febrero *et al.* (2008): Observación 8 (Profundidad Modal)
 - Hyndman, R. & Ullah, S. (2007):
 - Observaciones 1, 3, 6, 8, 10, 15, 17, 18, 25, 26, 29, 32, 37, 38, 42, 49, 53 (usando el default)
 - Observaciones 3, 8, 25, 32, 49, 53 (Usando $\lambda = 20$)
 - Hyndman, R. & Shang, H. (2010); ver gráficos
 - Boente, G. & Salibián-Barrera, M. (2015): Observación 3 (ver gráficos)

Estimación

6

Propiedades de $\widehat{\Gamma}$

○○○○○

Ejemplos

Mediana

○○

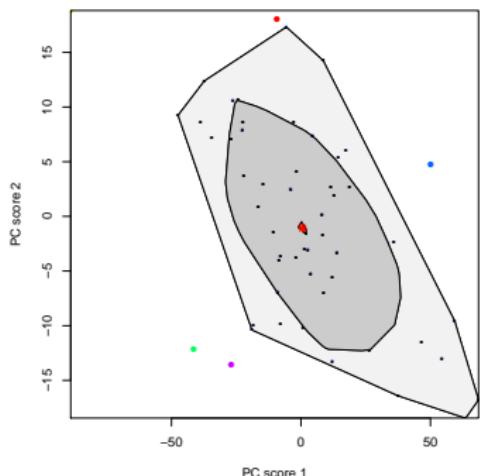
Profundidad

A diagram consisting of three rows of circles. The top row contains 8 circles, the middle row contains 5 circles, and the bottom row contains 8 circles, representing the numbers 8, 5, and 8 respectively.

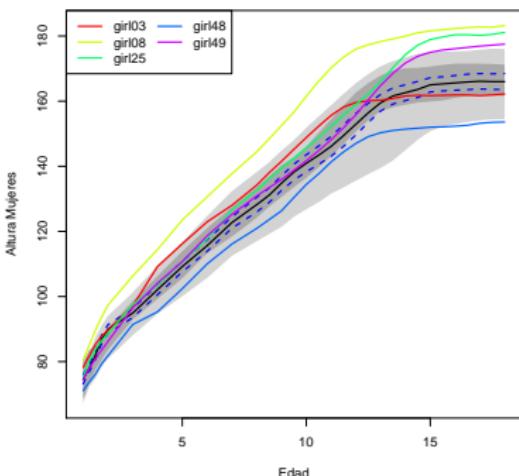
Boxplot Funcional

Mujeres: PCAPP

Bagplot



Bagplot Funcional



Estimación

10

Propiedades de $\widehat{\Gamma}$

A horizontal row of ten small, empty circles arranged in a single line.

Ejemplos

Mediana

○○

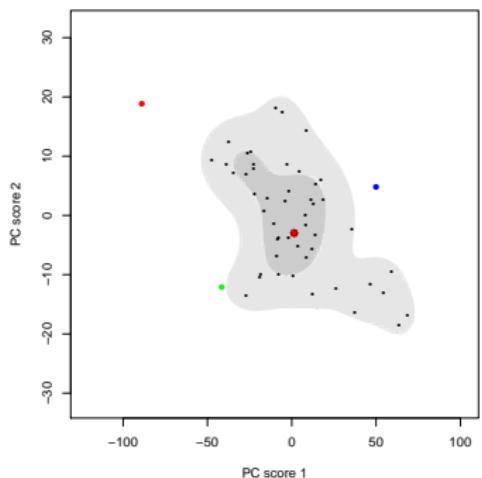
Profundidad

A horizontal row of ten small, light-colored circles arranged in a single line.

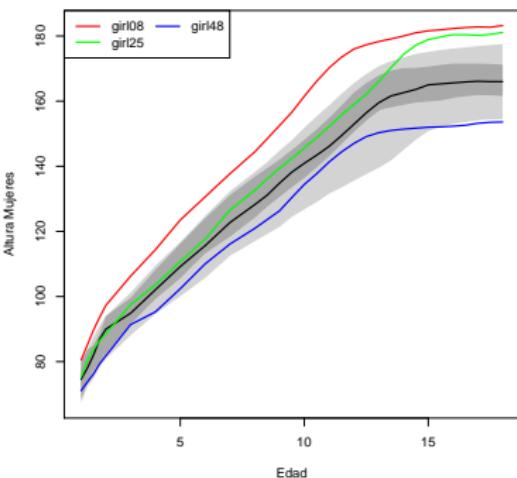
Boxplot Funcional

Mujeres: PCAPP

High Density Region



HDR Boxplot Funcional



Estimación

10

Propiedades de $\widehat{\Gamma}$

○○○○○

○○○○○○○○○○

Ejemplos

A 3x5 grid of 15 small circles, arranged in three rows and five columns.

Mediana

○○

Profundidad

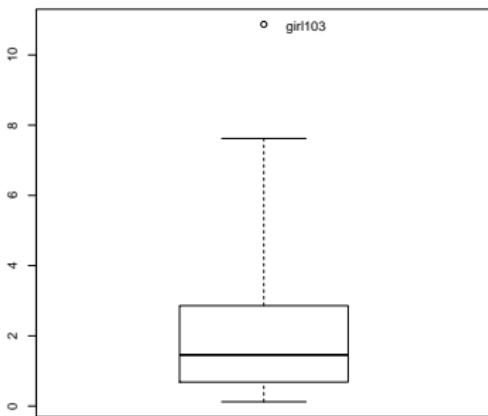
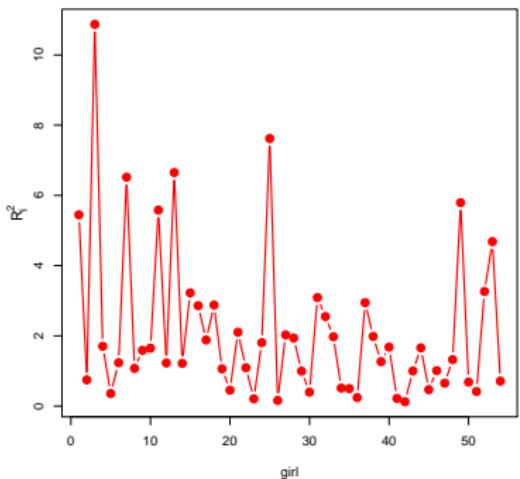
A 3x8 grid of circles, arranged in three rows and eight columns.

Boxplot Funcional

Mujeres: Boente & Salibián-Barrera (2015)

$$\|X_i - \hat{X}_i\|^2$$

Boxplot ajustado de R_i^2



Estimación

6

Propiedades de $\widehat{\Gamma}$

A diagram consisting of two rows of small circles. The top row contains four circles arranged horizontally. Below it, the bottom row contains seven circles arranged horizontally.

Ejemplos

100

Mediana

○○

Profundidad

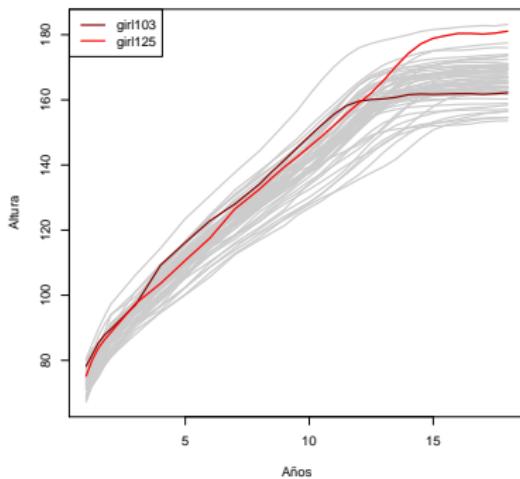
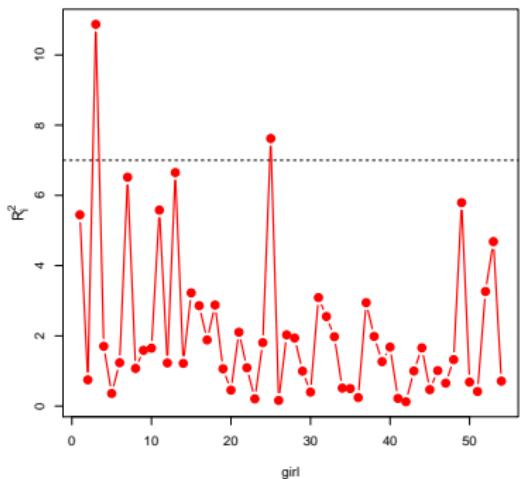
○○○○○○○○○○

Boxplot Funcional

• 10

Mujeres: Boente & Salibián-Barrera (2015)

$$R_i^2 = \|X_i - \hat{X}_i\|^2$$



Estimación

Propiedades de $\widehat{\Gamma}$

Ejemplos

Mediana

Profundidad

Boxplot Funcional

1

A diagram consisting of two rows of five circles each. The top row has circles at positions (1,1), (1,2), (1,3), (1,4), and (1,5). The bottom row has circles at positions (2,1), (2,2), (2,3), (2,4), and (2,5).

A 3x8 grid of small circles, arranged in three rows and eight columns.

○○

The diagram consists of three horizontal rows of small circles. The top row contains 8 circles. The middle row contains 5 circles. The bottom row contains 10 circles.

Mujeres: Punto Corte Hyndman & Ullah, S. (2007)

$$R_i^2 = \|X_i - \hat{X}_i\|^2$$

Alturas

