

Repaso Caso \mathbb{R}^P

oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooooooooooo
oooooooooooo

Estimación

ooo
oooooooooooo
ooooooo

Estimadores penalizados

oooooooooooo
oo

Componentes Principales Funcionales (FPCA)

Graciela Boente

Motivación

Sea $x \in \mathbb{R}^p$ tal que

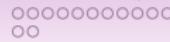
$$\mathbb{E}(\mathbf{x}) = \mathbf{0} \quad \text{VAR}(\mathbf{x}) = \Sigma$$

El método de componentes principales busca elegir q combinaciones lineales

$$\xi_1 = \phi_1^T \mathbf{x}, \quad \xi_2 = \phi_2^T \mathbf{x}, \quad \dots \quad \xi_q = \phi_q^T \mathbf{x}$$

de modo tal que si $\xi = (\xi_1, \dots, \xi_q)$ entonces, ξ explica una porción razonable de la dispersión total medida a través de traza(Σ).

Como ejemplo, tomemos las mediciones del tamaño de la cabeza del primer y segundo hijo.



Ejemplo 1

Teníamos como medidas

- x_1 = longitud de la cabeza,
- x_2 = ancho de la cabeza

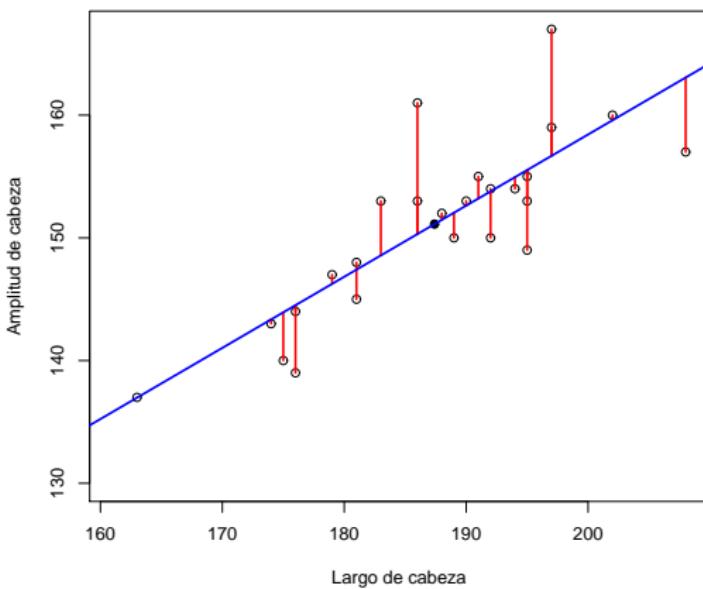
medidas sobre el primer y segundo hijo de 25 familias, dando origen a observaciones x_{ij} , $1 \leq j \leq 25$, $i = 1, 2$ donde $i = 1$ indica las observaciones correspondientes al hijo mayor e $i = 2$ aquellas correspondientes al segundo hijo.

$$\bar{x}_1 = \begin{pmatrix} 187.40 \\ 151.12 \end{pmatrix} \quad \bar{x}_2 = \begin{pmatrix} 183.32 \\ 149.36 \end{pmatrix}$$

Ejemplo 1

- En este ejemplo, ninguna de las dos variables x_1 o x_2 puede ser declarada como **independiente** o **dependiente**.
 - Esto constituye la diferencia esencial con el análisis de regresión.
 - La recta que buscamos **no es la recta de regresión** y se obtiene minimizando la distancia de los puntos a la recta pero **en lugar de medir la distancia verticalmente como en regresión, la medimos en forma ortogonal a la recta**.
 - Es el principio de mínimos cuadrados ortogonales de Pearson (1901).

Hijo Mayor: Recta de regresión



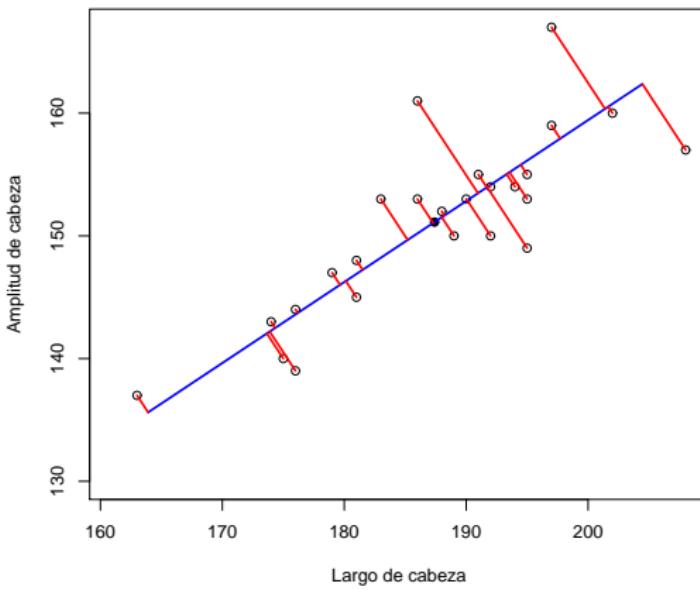
5

Caso Espacio Hilbert

Estimación

Estimadores penalizados
oooooooooooo
oo

Hijo Mayor: Componente principal



Definición

Sea $\mathbf{x} \in \mathbb{R}^p$ tal que

$$\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu} \quad \text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$$

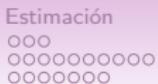
Sean

- $\lambda_1 \geq \dots \geq \lambda_p$ los autovalores de $\boldsymbol{\Sigma}$
- ϕ_1, \dots, ϕ_p los autovectores de $\boldsymbol{\Sigma}$ asociados a $\lambda_1 \geq \dots \geq \lambda_p$
- $\Phi = (\phi_1, \dots, \phi_p)$, $\Phi\Phi^T = \mathbf{I}_p$
- $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_p)$

$$\boldsymbol{\Phi}^T \boldsymbol{\Sigma} \boldsymbol{\Phi} = \boldsymbol{\Lambda}, \quad \boldsymbol{\Sigma} = \sum_{j=1}^p \lambda_j \phi_j \phi_j^T = \boldsymbol{\Phi} \boldsymbol{\Lambda} \boldsymbol{\Phi}^T$$

Luego, podemos escribir a \mathbf{x} como

$$\mathbf{x} = \boldsymbol{\mu} + \sum_{j=1}^p \{\phi_j^T(\mathbf{x} - \boldsymbol{\mu})\} \phi_j$$



Definición

Sea el vector $\xi = \Phi^T(x - \mu)$. Las coordenadas ξ_1, \dots, ξ_p de ξ se llaman **las componentes principales de x** .

La j -ésima componente principal es, por lo tanto,

$$\xi_j = \phi_j^T(\mathbf{x} - \boldsymbol{\mu}),$$

corresponde a la proyección ortogonal de $(\mathbf{x} - \boldsymbol{\mu})$ en la dirección ϕ_j .

Se llama *j*-ésima componente principal estandarizada a la variable

$$z_j = \lambda_j^{-\frac{1}{2}} \xi_j = \lambda_j^{-\frac{1}{2}} \phi_j^T(\mathbf{x} - \mu)$$

Propiedad 1. La componentes principales ξ_1, \dots, ξ_p son no correlacionadas y $\text{VAR}(\xi_j) = \lambda_j$, o sea,

$$\text{VAR}(\boldsymbol{\xi}) = \boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_p)$$

Predicción

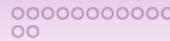
Supongamos que $E\mathbf{x} = \boldsymbol{\mu}$ y $\text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$.

Podemos escribir a \mathbf{x} como

$$\mathbf{x} = \boldsymbol{\mu} + \sum_{j=1}^p \xi_j \phi_j$$

Supongamos que predecimos a \mathbf{x} usando un predictor lineal basado en $\boldsymbol{\xi}_q = (\xi_1, \dots, \xi_q)^T$. El mejor predictor lineal de \mathbf{x} basado en $\boldsymbol{\xi}_q$ es

$$\mathbf{x}^* = \boldsymbol{\mu} + \sum_{j=1}^q \xi_j \phi_j$$



Propiedades de optimalidad

Propiedad 1. (Pearson, 1901) Sea $\mathbf{x} \in \mathbb{R}^P$ tal que

$$\mathbb{E}(\mathbf{x}) = \mathbf{0} \quad \text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma}$$

- $\lambda_1 \geq \dots \geq \lambda_p$ los autovalores de $\boldsymbol{\Sigma}$ y
- ϕ_1, \dots, ϕ_p los autovectores de $\boldsymbol{\Sigma}$ asociados a $\lambda_1 \geq \dots \geq \lambda_p$.
- \mathcal{H}_0 el subespacio generado por ϕ_1, \dots, ϕ_q donde $\lambda_q > \lambda_{q+1}$.

Indiquemos por $\pi(\mathbf{x}, \mathcal{H})$ a la proyección ortogonal de \mathbf{x} sobre el subespacio \mathcal{H} . Entonces, se tiene que para todo subespacio \mathcal{H} de dimensión q

$$\mathbb{E}\|\mathbf{x} - \pi(\mathbf{x}, \mathcal{H}_0)\|^2 \leq \mathbb{E}\|\mathbf{x} - \pi(\mathbf{x}, \mathcal{H})\|^2$$

o sea, las componentes principales dan el mejor ajuste lineal de dimensión q .

Propiedades de optimalidad

Propiedad 2. Sea $x \in \mathbb{R}^p$ tal que

$$\mathbb{E}(\mathbf{x}) = \boldsymbol{\mu} \quad \text{VAR}(\mathbf{x}) = \boldsymbol{\Sigma} > 0$$

Sean $\lambda_1 \geq \dots \geq \lambda_p$ los autovalores de Σ y ϕ_1, \dots, ϕ_p los autovectores de Σ asociados a $\lambda_1 \geq \dots \geq \lambda_p$. Entonces,

- a) $\max_{\|\mathbf{a}\|=1} \text{VAR}(\mathbf{a}^T \mathbf{x}) = \text{VAR}(\xi_1)$, o sea, el máximo se alcanza en ϕ_1 .

$$\text{b) } \max_{\|\mathbf{a}\|=1} \text{VAR}(\mathbf{a}^T \mathbf{x}) = \text{VAR}(\xi_{k+1}),$$

$$\text{COV}(\mathbf{a}^T \mathbf{x}, \xi_j) = 0 \quad 1 \leq j \leq k$$

es decir, el máximo se alcanza en ϕ_{k+1} .

La condición $\text{Cov}(\mathbf{a}^T \mathbf{x}, \xi_j) = 0$ asegura que no se repite información.

c) $\sum_{j=1}^p \text{VAR}(\xi_j) = \text{traza}(\Sigma).$

Notación

- \mathcal{H} espacio de Hilbert separable, producto interno $\langle \cdot, \cdot \rangle$ y norma $\|\alpha\|^2 = \langle \alpha, \alpha \rangle$.
 - $u \otimes v : \mathcal{H} \rightarrow \mathcal{H}$ $(u \otimes v)w = \langle v, w \rangle u$.
 - Sea $\mu = \mathbb{E}(X)$. Si $\mathbb{E}(\|X\|^2) < \infty$, el operador de covarianza Γ de X es tal que $\Gamma : \mathcal{H} \rightarrow \mathcal{H}$

$$\text{Cov}(\langle X, \alpha \rangle, \langle X, \beta \rangle) = \text{Cov}(\langle X - \mu, \alpha \rangle, \langle X - \mu, \beta \rangle) = \langle \alpha, \Gamma \beta \rangle$$
 - Γ es un operador lineal, auto-adjunto, continuo, definido no-negativo y Hilbert-Schmidt.
 - $\Gamma = \mathbb{E}\{(X - \mu) \otimes (X - \mu)\}$ con $\mu = \mathbb{E}(X)$.
 - Las autofunciones ϕ_j de Γ cumplen $\Gamma \phi_j = \lambda_j \phi_j$

Caso L^2

- Cuando $\mathcal{H} = L^2(\mathcal{I})$, tenemos el núcleo de covarianza

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

- $(\Gamma\alpha)(t) = \int_{\mathcal{I}} \gamma(t, s)\alpha(s) ds$
 - Las autofunciones cumplen

$$(\Gamma \phi_j)(t) = \lambda_j \phi_j(t) \quad \text{o sea}$$

$$\int_{\mathcal{I}} \gamma(t,s) \phi_j(s) \, ds = \lambda_j \phi_j(t)$$

Teorema de Mercer

- ϕ_ℓ : autofunciones de Γ con autovalor asociado λ_ℓ

$$\langle \phi_k, \phi_\ell \rangle = \delta_{k\ell}$$

-

$$\Gamma = \sum_{\ell \geq 1} \lambda_\ell \phi_\ell \otimes \phi_\ell \quad \lambda_j \geq \lambda_{j+1}$$

Teorema de Mercer: $\mathcal{H} = L^2([0, 1])$

- Si $\gamma(s, t)$ es **continuo** sobre $[0, 1]^2$, existe una sucesión decreciente de números positivos λ_j y una sucesión de funciones continuas ϕ_j tales que

$$\int_0^1 \gamma(t, s) \phi_j(s) ds = \lambda_j \phi_j(t) \quad \forall j \in \mathbb{N} \quad \int_0^1 \phi_\ell(s) \phi_j(s) ds = \delta_{\ell j}.$$

Más aún,

$$\gamma(t, s) = \sum_{\ell \geq 1} \lambda_\ell \phi_\ell(t) \phi_\ell(s) \quad s, t \in [0, 1]$$

y la serie converge uniformemente en $[0, 1]^2$, es decir,

$$\lim_{p \rightarrow \infty} \sup_{(s, t) \in [0, 1]^2} \left| \sum_{\ell=1}^p \lambda_\ell \phi_\ell(t) \phi_\ell(s) - \gamma(t, s) \right| = 0$$

Karhunen-Loèv

- Teorema de Mercer

$$\Gamma = \sum_{\ell \geq 1} \lambda_\ell \phi_\ell \otimes \phi_\ell \quad \quad \lambda_j \geq \lambda_{j+1}$$

$$\langle \phi_k, \phi_\ell \rangle = \delta_{k\ell}$$

- Desarrollo de Karhunen-Loëve

$$X = \mu + \sum_{\ell=1}^{\infty} \xi_\ell \phi_\ell = \mu + \sum_{\ell=1}^{\infty} \lambda_\ell^{\frac{1}{2}} z_\ell \phi_\ell$$

$$\xi_\ell = \lambda_\ell^{\frac{1}{2}} z_\ell = \langle X - \mu, \phi_\ell \rangle$$

$$\mathbb{E}(z_\ell) = 0 \quad \text{Cov}(z_\ell, z_k) = \delta_{k\ell}$$

$$\mathbb{E}(\xi_\ell) = 0 \quad \text{Cov}(\xi_\ell, \xi_k) = 0 \text{ si } k \neq \ell \quad \text{VAR}(\xi_\ell) = \lambda_\ell$$

Karhunen-Loève

Veamos que

$$\mathbb{E}(z_\ell) = 0 \quad \text{Cov}(z_\ell, z_k) = \delta_{k\ell}$$

- $\mathbb{E}(z_\ell) = \lambda_\ell^{-\frac{1}{2}} \mathbb{E}\langle X - \mu, \phi_\ell \rangle = \lambda_\ell^{-\frac{1}{2}} \langle \mathbb{E}X - \mu, \phi_\ell \rangle = 0$
- $\text{Cov}(z_\ell, z_k) = \lambda_\ell^{-\frac{1}{2}} \lambda_k^{-\frac{1}{2}} \text{Cov}(\langle X - \mu, \phi_\ell \rangle, \langle X - \mu, \phi_k \rangle)$
 $= \lambda_\ell^{-\frac{1}{2}} \lambda_k^{-\frac{1}{2}} \langle \phi_\ell, \Gamma \phi_k \rangle$
 $= \lambda_\ell^{-\frac{1}{2}} \lambda_k^{-\frac{1}{2}} \lambda_k \langle \phi_\ell, \phi_k \rangle$
 $= \begin{cases} \lambda_\ell^{-\frac{1}{2}} \lambda_k^{-\frac{1}{2}} \lambda_k & k = \ell \\ \lambda_\ell^{-\frac{1}{2}} \lambda_k^{-\frac{1}{2}} \lambda_k \times 0 & k \neq \ell \end{cases}$
 $= \delta_{\ell k}$

Karhunen-Loève

En qué sentido vale

$$X = \mu + \sum_{\ell=1}^{\infty} \xi_{\ell} \phi_{\ell} = \mu + \sum_{\ell=1}^{\infty} \langle X - \mu, \phi_{\ell} \rangle \phi_{\ell}$$

a) $\lim_{p \rightarrow \infty} \mathbb{E} \|X - \mu - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}\|^2 = 0$

$$\begin{aligned} \mathbb{E} \left\| X - \mu - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell} \right\|^2 &= \mathbb{E} \langle X - \mu - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}, X - \mu - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell} \rangle \\ &= \mathbb{E} \|X - \mu\|^2 - 2 \sum_{\ell=1}^p \mathbb{E} \langle X - \mu, \xi_{\ell} \phi_{\ell} \rangle + \mathbb{E} \left\langle \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}, \sum_{j=1}^p \xi_j \phi_j \right\rangle \\ &= \sum_{j \geq 1} \lambda_j - 2 \sum_{\ell=1}^p \mathbb{E} \{ \langle X - \mu, \phi_{\ell} \rangle \langle X - \mu, \phi_{\ell} \rangle \} + \mathbb{E} \left\langle \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}, \sum_{j=1}^p \xi_j \phi_j \right\rangle \end{aligned}$$

$$\lim_{p \rightarrow \infty} \mathbb{E} \|X - \mu - \sum_{\ell=1}^p \xi_\ell \phi_\ell\|^2 = 0$$

$$\begin{aligned}
 & \mathbb{E} \left\| X - \mu - \sum_{\ell=1}^p \xi_\ell \phi_\ell \right\|^2 \\
 &= \sum_{j \geq 1} \lambda_j - 2 \sum_{\ell=1}^p \mathbb{E} \{ \langle X - \mu, \phi_\ell \rangle \langle X - \mu, \phi_\ell \rangle \} + \mathbb{E} \left\langle \sum_{\ell=1}^p \xi_\ell \phi_\ell, \sum_{j=1}^p \xi_j \phi_j \right\rangle \\
 &= \sum_{j \geq 1} \lambda_j - 2 \sum_{\ell=1}^p \text{Cov} (\langle X - \mu, \phi_\ell \rangle, \langle X - \mu, \phi_\ell \rangle) + \sum_{\ell=1}^p \sum_{j=1}^p \mathbb{E} (\xi_\ell \xi_j) \langle \phi_\ell, \phi_j \rangle \\
 &= \sum_{j \geq 1} \lambda_j - 2 \sum_{\ell=1}^p \langle \phi_\ell, \Gamma \phi_\ell \rangle + \sum_{\ell=1}^p \sum_{j=1}^p \lambda_j \delta_{j\ell} \langle \phi_\ell, \phi_j \rangle \\
 &= \sum_{j \geq 1} \lambda_j - 2 \sum_{\ell=1}^p \langle \phi_\ell, \lambda_\ell \phi_\ell \rangle + \sum_{j=1}^p \lambda_j \langle \phi_j, \phi_j \rangle = \sum_{j \geq 1} \lambda_j - 2 \sum_{\ell=1}^p \lambda_\ell + \sum_{j=1}^p \lambda_j \\
 &= \sum_{j \geq 1} \lambda_j - \sum_{\ell=1}^p \lambda_\ell \rightarrow 0 \text{ cuando } p \rightarrow \infty
 \end{aligned}$$

Karhunen-Loève

En qué sentido vale

$$X(t) = \mu(t) + \sum_{\ell=1}^{\infty} \xi_{\ell} \phi_{\ell}(t) \quad (1)$$

- a) $\lim_{p \rightarrow \infty} \mathbb{E} \|X - \mu - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}\|^2 = 0$ con $\xi_{\ell} = \langle X - \mu, \phi_{\ell} \rangle$
- b) Si $\mathcal{H} = L^2([0, 1])$ y el núcleo de covarianza $\gamma(s, t)$ es **continuo** sobre $[0, 1]^2$, entonces vale (1) para una sucesión de variables aleatorias $\{\xi_{\ell}\}$ tal que $\mathbb{E}(\xi_{\ell}) = 0$, $\text{Cov}(\xi_{\ell}, \xi_k) = \lambda_{\ell} \delta_{\ell k}$

La serie en (1) converge uniformemente en $L^2(\Omega)$ el sentido que

$$\lim_{p \rightarrow \infty} \sup_{t \in [0, 1]} \mathbb{E} \left(X(t) - \mu(t) - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}(t) \right)^2 = 0$$

Karhunen-Loève b)

Sea

$$\xi_\ell = \int_0^1 (X(t) - \mu(t)) \phi_\ell(t) dt = \langle X - \mu, \phi_\ell \rangle$$

que vimos cumple

$$\mathbb{E}(\xi_\ell) = 0 \quad \text{Cov}(\xi_\ell, \xi_k) = \lambda_\ell \delta_{\ell,k}$$

Por otra parte,

$$\begin{aligned} \mathbb{E}\{(X(t) - \mu(t)) \xi_\ell\} &= \mathbb{E}\left\{(X(t) - \mu(t)) \int_0^1 (X(s) - \mu(s)) \phi_\ell(s) ds\right\} \\ &= \int_0^1 \mathbb{E}\{(X(t) - \mu(t)) (X(s) - \mu(s))\} \phi_\ell(s) ds \\ &= \int_0^1 \gamma(t, s) \phi_\ell(s) ds = \lambda_\ell \phi_\ell(t) \end{aligned}$$

de donde

Repaso Caso \mathbb{R}^P

○○○○○
○○○○○

Caso Espacio Hilbert

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

Estimación

○○○
○○○○○○○○○○
○○○○○○

Estimadores penalizados

○○○○○○○○○○
○○

Karhunen-Loève b)

$$\begin{aligned}
 & \mathbb{E} \left\{ (X(t) - \mu(t)) - \sum_{\ell=1}^p \xi_\ell \phi_\ell(t) \right\}^2 \\
 &= \mathbb{E} (X(t) - \mu(t))^2 - 2 \sum_{\ell=1}^p \mathbb{E} \{(X(t) - \mu(t)) \xi_\ell \phi_\ell(t)\} + \mathbb{E} \left\{ \sum_{\ell=1}^p \xi_\ell \phi_\ell(t) \right\}^2 \\
 &= \mathbb{E} (X(t) - \mu(t))^2 - 2 \sum_{\ell=1}^p \phi_\ell(t) \mathbb{E} \{(X(t) - \mu(t)) \xi_\ell\} + \sum_{\ell=1}^p \sum_{j=1}^p \mathbb{E} \{\xi_\ell \phi_\ell(t) \xi_j \phi_j(t)\} \\
 &= \mathbb{E} (X(t) - \mu(t))^2 - 2 \sum_{\ell=1}^p \phi_\ell(t) \lambda_\ell \phi_\ell(t) + \sum_{\ell=1}^p \sum_{j=1}^p \phi_\ell(t) \phi_j(t) \mathbb{E} \{\xi_\ell \xi_j\} \\
 &= \mathbb{E} (X(t) - \mu(t))^2 - 2 \sum_{\ell=1}^p \lambda_\ell \phi_\ell^2(t) + \sum_{\ell=1}^p \phi_\ell^2(t) \lambda_\ell \\
 &= \gamma(t, t) - \sum_{\ell=1}^p \lambda_\ell \phi_\ell^2(t)
 \end{aligned}$$

Repaso Caso \mathbb{R}^P

○○○○○
○○○○○

Caso Espacio Hilbert

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

Estimación

○○○
○○○○○○○○○○
○○○○○○

Estimadores penalizados

○○○○○○○○○○
○○

Karhunen-Loève b)

Por el Teorema de Mercer, obtenemos que

$$\sup_{t \in [0,1]} \mathbb{E} \left\{ (X(t) - \mu(t)) - \sum_{\ell=1}^p \xi_\ell \phi_\ell(t) \right\}^2 \leq \sup_{t \in [0,1]} \left| \gamma(t, t) - \sum_{\ell=1}^p \lambda_\ell \phi_\ell^2(t) \right| \rightarrow 0$$

Karhunen-Loève: $X(t) = \mu(t) + \sum_{\ell=1}^{\infty} \xi_{\ell} \phi_{\ell}(t)$ (1)

- a) $\lim_{p \rightarrow \infty} \mathbb{E} \|X - \mu - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}\|^2 = 0$ con $\xi_{\ell} = \langle X - \mu, \phi_{\ell} \rangle$
- b) Si $\mathcal{H} = L^2([0, 1])$ y el núcleo de covarianza $\gamma(s, t)$ es **continuo** sobre $[0, 1]^2$, entonces la serie en (1) converge uniformemente en $L^2(\Omega)$ el sentido que

$$\lim_{p \rightarrow \infty} \sup_{t \in [0, 1]} \mathbb{E} \left(X(t) - \mu(t) - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}(t) \right)^2 = 0$$

- c) Si $\mathcal{H} = L^2([0, 1])$, X es un **proceso Gaussiano con trayectorias continuas**, o sea, tal que

$$\mathbb{P} \left(\lim_{s \rightarrow t} X(s) = X(t) \quad \text{para todo } t \in [0, 1] \right) = 1$$

y el núcleo de covarianza $\gamma(s, t)$ es **continuo**, entonces la serie en (1) converge uniformemente en $[0, 1]$ con probabilidad 1, es decir,

$$\mathbb{P} \left(\lim_{p \rightarrow \infty} \sup_{t \in [0, 1]} \left| X(t) - \mu(t) - \sum_{\ell=1}^p \xi_{\ell} \phi_{\ell}(t) \right| = 0 \right) = 1$$

Adler, R. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. IMS Lecture Notes Monograph Series, Vol 12, Institute Mathematical Statistics.

○○○○○
○○○○○

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

○○○
○○○○○○○○○○
○○○○○○○○

○○○○○○○○○○
○○

Caso L^2

- Cuando $\mathcal{H} = L^2(\mathcal{I})$, tenemos el núcleo de covarianza

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

- $(\Gamma\alpha)(t) = \int_{\mathcal{I}} \gamma(t, s)\alpha(s) ds$
- Las autofunciones cumplen

$$(\Gamma\phi_j)(t) = \lambda_j \phi_j(t) \quad \text{o sea}$$

$$\int_{\mathcal{I}} \gamma(t, s)\phi_j(s) ds = \lambda_j \phi_j(t)$$

Repaso Caso \mathbb{R}^P

○○○○○
○○○○○

Caso Espacio Hilbert

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

Estimación

○○○
○○○○○○○○○○
○○○○○○

Estimadores penalizados

○○○○○○○○○○
○○

Observación

En el caso multivariado, si $\mathbf{Z} \in \mathbb{R}^P$ es tal que

$$\mathbf{Z} \sim N(\mathbf{0}, \mathbf{I}_P)$$

entonces

- Z_i i.i.d. $\mathbb{E}(Z_i) = 0$ y $\text{VAR}(Z_i) = 1$
- Por lo tanto, podemos escribir

$$\mathbf{Z} = \sum_{j=1}^P Z_j \mathbf{e}_j$$

donde \mathbf{e}_j es la base canónica.

Observación

En el caso infinito-dimensional, si tomamos como operador $\Gamma = \mathbf{I}$, o sea,

$$\Gamma x = x$$

entonces

- Γ es lineal, auto-adjunto, acotado
- No es un operador compacto, ni Hilbert–Schmidt (si $\dim(\mathcal{H}) = \infty$)
- Por lo tanto, $\Gamma = \mathbf{I}$ no define un operador de covarianza.
- **No podemos definir un elemento aleatorio con operador de covarianza I**

Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

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

Estimación

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

Estimadores penalizados

○○○○○○○○○○
○○

Observación

- Z_i i.i.d. $\mathbb{E}(Z_i) = 0$ y $\text{VAR}(Z_i) = 1$
 - ϕ_j una base ortonormal de $L^2([0, 1])$.
 - **Porqué**

$$X = \sum_{j \geq 1} Z_j \phi_j$$

no define un elemento aleatorio de $L^2([0, 1])$

Observación

- En este caso,

$$\mathbb{E} \sum_{j=1}^n Z_j^2 = n$$

- Por lo tanto,

$$\mathbb{E} \sum_{j \geq 1} Z_j^2 = \mathbb{E} \lim_{n \rightarrow \infty} \sum_{j=1}^n Z_j^2 = \lim_{n \rightarrow \infty} \mathbb{E} \sum_{j=1}^n Z_j^2 = \infty$$

- Luego, $\|X\|^2 = \infty$ casi seguramente y por eso, $X \notin L^2([0, 1])$
- No hay análogo a la $N(0, I_p)$ en el caso funcional.**

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

- Sea Γ un operador compacto, auto-adjunto, definido no-negativo sobre un espacio de Hilbert separable \mathcal{H}
- Sean $\lambda_1 \geq \lambda_2 \geq \dots$ los autovalores positivos de Γ , cada uno de ellos repetidos tantas veces como su multiplicidad.
- Sea ϕ_j la autofunción asociada a λ_j elegidas de modo que $\{\phi_j, j \geq 1\}$ es un conjunto ortonormal en \mathcal{H} .

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

- Entonces, se tiene que

$$\begin{aligned}\lambda_1 &= \max_{\alpha \neq 0} \frac{\langle \Gamma \alpha, \alpha \rangle}{\|\alpha\|^2} \\ \lambda_{k+1} &= \min_{\mathcal{L}_k} \max_{\alpha \in \mathcal{L}_k^\perp} \frac{\langle \Gamma \alpha, \alpha \rangle}{\|\alpha\|^2},\end{aligned}\tag{2}$$

donde \mathcal{L}_k es cualquier subespacio de dimensión k

\mathcal{L}_k^\perp es el subespacio ortogonal a \mathcal{L}_k , o sea,

$$\mathcal{L}_k^\perp = \{u \in \mathcal{H} : \langle u, y \rangle, \quad \forall y \in \mathcal{L}_k\}$$

El mínimo en (2) se alcanza cuando \mathcal{L}_k está generado por ϕ_1, \dots, ϕ_k .

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

- En particular

$$\lambda_1 = \max_{\alpha \neq 0} \frac{\langle \Gamma \alpha, \alpha \rangle}{\|\alpha\|^2} = \max_{\|\alpha\|=1} \langle \Gamma \alpha, \alpha \rangle \quad (3)$$

$$\begin{aligned} \lambda_{k+1} &= \max_{\substack{\alpha \neq 0 \\ \langle \alpha, \phi_j \rangle = 0, 1 \leq j \leq k}} \frac{\langle \Gamma \alpha, \alpha \rangle}{\|\alpha\|^2} \\ &= \max_{\substack{\|\alpha\|=1 \\ \langle \alpha, \phi_j \rangle = 0, 1 \leq j \leq k}} \langle \Gamma \alpha, \alpha \rangle, \end{aligned} \quad (4)$$

Repaso Caso \mathbb{R}^P

○○○○
○○○○

Caso Espacio Hilbert

Estimación

A diagram consisting of three rows of small circles. The top row contains 3 circles. The middle row contains 8 circles. The bottom row contains 6 circles.

Estimadores penalizados

○○○○○○○○○○
○○

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

Probaremos (3). Dado $\alpha \in \mathcal{H}$, como

$$\Gamma = \sum_{j \geq 1} \lambda_j \phi_j \otimes \phi_j$$

tenemos,

$$\Gamma\alpha = \sum_{j \geq 1} \lambda_j \phi_j \otimes \phi_j \quad \alpha = \sum_{j \geq 1} \lambda_j \langle \alpha, \phi_j \rangle \phi_j$$

de donde

$$\langle \Gamma\alpha, \alpha \rangle = \left\langle \sum_{j \geq 1} \lambda_j \langle \alpha, \phi_j \rangle \phi_j, \alpha \right\rangle = \sum_{j \geq 1} \lambda_j \langle \alpha, \phi_j \rangle \langle \phi_j, \alpha \rangle = \sum_{j \geq 1} \lambda_j \langle \alpha, \phi_j \rangle^2$$

Luego,

$$\max_{\alpha \neq 0} \frac{\langle \Gamma \alpha, \alpha \rangle}{\|\alpha\|^2} = \max_{\alpha \neq 0} \frac{\sum_{j \geq 1} \lambda_j \langle \alpha, \phi_j \rangle^2}{\|\alpha\|^2}$$

Repaso Caso \mathbb{R}^P

oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooo●oooo
oooooooooooo

Estimación

ooo
oooooooooooo
ooooooo

Estimadores penalizados

oooooooooooo
oo

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

Por la igualdad de Parseval

$$\|\alpha\|^2 = \sum_{j \geq 1} \langle \alpha, \phi_j \rangle^2$$

Por lo tanto, como $\lambda_1 \geq \lambda_2 \geq \dots$

$$\langle \Gamma\alpha, \alpha \rangle = \sum_{j \geq 1} \lambda_j \langle \alpha, \phi_j \rangle^2 \leq \lambda_1 \sum_{j \geq 1} \langle \alpha, \phi_j \rangle^2 = \lambda_1 \|\alpha\|^2$$

Por otra parte, si $\alpha = \phi_1$,

$$\langle \Gamma\alpha, \alpha \rangle = \lambda_1$$

es decir, para $\alpha \neq 0$,

$$\frac{\langle \Gamma\alpha, \alpha \rangle}{\|\alpha\|^2} \leq \lambda_1 = \langle \Gamma\phi_1, \phi_1 \rangle$$

lo que concluye la demostración de (3).

Observemos que la demostración es análoga a la del caso \mathbb{R}^P .



Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

A horizontal row of 18 small circles. The 8th circle from the left is filled black, while all other circles are unfilled.

Estimación

A diagram consisting of three rows of small circles. The top row contains 3 circles. The middle row contains 7 circles. The bottom row contains 5 circles.

Estimadores penalizados

○○○○○○○○○○
○○

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

Probaremos (4). Dado $\alpha \in \mathcal{H}$, tal que $\langle \alpha, \phi_j \rangle = 0$, para $1 \leq j \leq k$ se cumple

$$\Gamma\alpha = \sum_{j \geq 1} \lambda_j \langle \alpha, \phi_j \rangle \phi_j = \sum_{j > k+1} \lambda_j \langle \alpha, \phi_j \rangle \phi_j$$

de donde

$$\langle \Gamma\alpha, \alpha \rangle = \left\langle \sum_{j \geq k+1} \lambda_j \langle \alpha, \phi_j \rangle \phi_j, \alpha \right\rangle = \sum_{j \geq k+1} \lambda_j \langle \alpha, \phi_j \rangle^2.$$

Luego,

$$\max_{\substack{\alpha \neq 0 \\ \langle \alpha, \phi_j \rangle = 0, 1 \leq j \leq k}} \frac{\langle \Gamma \alpha, \alpha \rangle}{\|\alpha\|^2} = \max_{\substack{\alpha \neq 0 \\ \langle \alpha, \phi_j \rangle = 0, 1 \leq j \leq k}} \frac{\sum_{j \geq k+1} \lambda_j \langle \alpha, \phi_j \rangle^2}{\|\alpha\|^2}$$

Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

Estimación

A diagram consisting of three rows of small circles. The top row contains 3 circles, the middle row contains 7 circles, and the bottom row contains 5 circles, arranged horizontally.

Estimadores penalizados

○○○○○○○○○○
○○

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

Por lo tanto, como $\lambda_{k+1} \geq \lambda_{k+2} \geq \dots$

$$\langle \Gamma\alpha, \alpha \rangle = \sum_{j>k+1} \lambda_j \langle \alpha, \phi_j \rangle^2 \leq \lambda_{k+1} \sum_{j>k+1} \langle \alpha, \phi_j \rangle^2$$

$$\langle \Gamma\alpha, \alpha \rangle \leq \lambda_{k+1} \sum_{j>1} \langle \alpha, \phi_j \rangle^2 = \lambda_{k+1} \|\alpha\|^2$$

Por otra parte, si $\alpha = \phi_{k+1}$,

$$\langle \Gamma\alpha, \alpha \rangle = \lambda_{k+1}$$

es decir, para $\alpha \neq 0$, tal que $\langle \alpha, \phi_j \rangle = 0$, para $1 \leq j \leq k$, tenemos

$$\frac{\langle \Gamma\alpha, \alpha \rangle}{\|\alpha\|^2} \leq \lambda_{k+1} = \langle \Gamma\phi_{k+1}, \phi_{k+1} \rangle$$

lo que concluye la demostración de (4).

Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

○○○○○○○○○○○○○○○○●○○○○○○○○○○○○○○○○

Estimación

A diagram consisting of three rows of circles. The top row contains 3 circles, the middle row contains 7 circles, and the bottom row contains 6 circles, totaling 16 circles.

Estimadores penalizados

○○○○○○○○○○
○○

Principio min-max de Courant–Fisher: Dunford & Schwartz (1965), Cap. 10

Observemos que si $\lambda_{k+1} \geq \lambda_{k+2}$, entonces

$$\langle \Gamma\alpha, \alpha \rangle = \sum_{i \geq k+1} \lambda_i \langle \alpha, \phi_i \rangle^2 < \lambda_{k+1} \|\alpha\|^2$$

para α tal que $\langle \alpha, \phi_j \rangle^2 \neq 0$ para algún $j \geq k + 2$.

Por lo tanto, el valor que maximiza

$$\frac{\langle \Gamma\alpha, \alpha \rangle}{\|\alpha\|^2}$$

sujeto a $\alpha \neq 0$, $\langle \alpha, \phi_j \rangle = 0$, para $1 \leq j \leq k$, es único salvo signo, o sea, es $\phi_{k+1} \circ -\phi_{k+1}$.

Componentes Principales Funcionales (FPCA)

- Sea Γ el operador de covarianza de X
- Sean $\lambda_1 \geq \lambda_2 \geq \dots$ los autovalores positivos de Γ .
- Sea ϕ_j la autofunción asociada a λ_j elegidas de modo que $\{\phi_j, j \geq 1\}$ es un conjunto ortonormal en \mathcal{H} .
- **La j -ésima componente principal funcional de X se define como**

$$\xi_j = \langle X - \mu, \phi_j \rangle$$

y ϕ_j se llama la j -ésima dirección principal.

Las variables aleatorias ξ_j se llaman los escores de X

Propiedades FPCA

- **Propiedad 1.**

La primer dirección principal maximiza $\text{VAR}(\langle \alpha, X \rangle)$ sobre $\mathcal{S}_1 = \{\alpha : \|\alpha\| = 1\}$.

- **Propiedad 2.**

La primer dirección principal corresponde a la autofunción de Γ asociada al mayor autovalor y $\text{VAR}(\xi_1) = \lambda_1$.

- **Propiedad 3.**

Sea $Y = X - \mu$. Si \mathcal{L}_0 es el subespacio generado por ϕ_1, \dots, ϕ_p , con $\lambda_p > \lambda_{p+1}$, dado \mathcal{L} de dimensión p ,

$$\mathbb{E}(\|Y - \pi(Y, \mathcal{L}_0)\|^2) \leq \mathbb{E}(\|Y - \pi(Y, \mathcal{L})\|^2)$$

$\pi(X, \mathcal{L})$ es la proyección ortogonal de X sobre \mathcal{L} .

Propiedad 3: Caso \mathbb{R}^P

Recordemos que en el caso multivariado para probar la **Propiedad 3** usamos el

Teorema de separación de Poincaré. Sea $\Sigma \in \mathbb{R}^{P \times P}$ una matriz simétrica definida no-negativa. Sean

- $\lambda_1 \geq \dots \geq \lambda_p$ los autovalores de Σ y
- ϕ_1, \dots, ϕ_p los autovectores de Σ asociados a $\lambda_1 \geq \dots \geq \lambda_p$.

Entonces, si $\mathbf{B} \in \mathbb{R}^{P \times k}$ es tal que $\mathbf{B}^T \mathbf{B} = \mathbf{I}_k$, se tiene que

$$\lambda_j(\mathbf{B}^T \Sigma \mathbf{B}) \leq \lambda_j = \lambda_j(\Sigma) \quad 1 \leq j \leq k$$

donde $\lambda_j(\mathbf{A})$ indica el j -ésimo autovalor de \mathbf{A} .

Reparo Caso \mathbb{R}^P

oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooooooo
oooo●oooooooooooo

Estimación

ooo
oooooooooooo
ooooooo

Estimadores penalizados

oooooooooooo
oo

Caso Hilbert

Sean

- \mathcal{H} un espacio de Hilbert separable
- $\Gamma : \mathcal{H} \rightarrow \mathcal{H}$ un operador auto-adjunto, definido no-negativo y compacto
- $\lambda_1 \geq \lambda_2 \geq \dots$ sus autovalores positivos.
- Sea ϕ_j la autofunción asociada a λ_j , $j \geq 1$.

Repaso Caso \mathbb{R}^P oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooooooooooo
oooo●oooooooooooo

Estimación

ooo
oooooooooooo
ooooooo

Estimadores penalizados

oooooooooooo
oo

Caso Hilbert

Sean

- \mathcal{H}_1 un espacio de Hilbert separable
- $B : \mathcal{H} \rightarrow \mathcal{H}_1$ es un operador lineal sub-unitario, o sea,

$$BB^* = \mathbf{I}_{\mathcal{G}}$$

donde $\mathcal{G} = \text{rango}(B) = B(\mathcal{H})$

Entonces, se cumple que

$$\lambda_j(B\Gamma B^*) \leq \lambda_j(\Gamma) = \lambda_j$$

donde $\lambda_j(\Upsilon)$ indica el j -ésimo mayor autovalor de Υ .

Boente, G. Salibián-Barrera, M. & Tyler, D. (2014) . A characterization of elliptical distributions and some optimality properties of principal components for functional data. *Journal of Multivariate Analysis*, 131, 254–264.

Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

A 3x10 grid of circles. The bottom row has a black dot at the fifth position from the left.

Estimación

A diagram consisting of three rows of small circles. The top row contains 3 circles. The middle row contains 7 circles. The bottom row contains 6 circles.

Estimadores penalizados

○○○○○○○○○○
○○

$$\mathbb{E}(\|Y - \pi(Y, \mathcal{L}_0)\|^2) \leq \mathbb{E}(\|Y - \pi(Y, \mathcal{L})\|^2)$$

Para todo subespacio \mathcal{L} tenemos que

$$\mathbb{E}(\|Y - \pi(Y, \mathcal{L})\|^2) = \mathbb{E}\|Y\|^2 - \mathbb{E}(\|\pi(Y, \mathcal{L})\|^2)$$

Por lo tanto, basta mostrar que

$$\mathbb{E} (\|\pi(Y, \mathcal{L})\|^2) \leq \mathbb{E} (\|\pi(Y, \mathcal{L}_0)\|^2)$$

Observemos que si $\gamma_1, \dots, \gamma_p$ es una base ortonormal de \mathcal{L} ,

$$\pi(Y, \mathcal{L}) = \sum_{j=1}^p \langle Y, \gamma_j \rangle \gamma_j$$

Por lo tanto,

$$\mathbb{E} (\|\pi(Y, \mathcal{L})\|^2) = \sum_{j=1}^p \mathbb{E} (\langle Y, \gamma_j \rangle^2) = \sum_{j=1}^p \text{VAR} (\langle Y, \gamma_j \rangle)$$

$$\mathbb{E}(\|\pi(Y, \mathcal{L}_0)\|^2) = \sum_{j=1}^p \text{VAR}(\langle Y, \phi_j \rangle) = \sum_{j=1}^p \text{VAR}(\xi_j) = \sum_{j=1}^p \lambda_j$$

Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

A horizontal row of fifteen small circles. The circle in the seventh position from the left is filled with black, while all other circles are unfilled.

Estimación

A diagram consisting of three rows of small circles. The top row contains 3 circles. The middle row contains 8 circles. The bottom row contains 5 circles.

Estimadores penalizados

○○○○○○○○○○
○○

$$\mathbb{E}(\|Y - \pi(Y, \mathcal{L}_0)\|^2) \leq \mathbb{E}(\|Y - \pi(Y, \mathcal{L})\|^2)$$

Luego, basta mostrar que

$$\sum_{j=1}^p \text{VAR}(\langle Y, \gamma_j \rangle) \leq \sum_{j=1}^p \lambda_j$$

Sea $B : \mathcal{H} \rightarrow \mathbb{R}^p$ el operador lineal definido por

$$Bu = (\langle u, \gamma_1 \rangle, \dots, \langle u, \gamma_p \rangle)^T \quad \forall u \in \mathcal{H}$$

y B^* el adjunto de B , es decir,

$$(Bu)^T \mathbf{y} = \langle B u, \mathbf{y} \rangle_{\mathbb{R}^p} = \langle u, B^* \mathbf{y} \rangle \quad \forall u \in \mathcal{H}, \mathbf{y} \in \mathbb{R}^p$$

Repaso Caso \mathbb{R}^p

oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooooooo
oooooooo●oooooooo

Estimación

ooo
oooooooooooo
ooooooo

Estimadores penalizados

oooooooooooo
oo

Observemos que en este caso $B^* \mathbf{y} = \sum_{j=1}^p y_j \gamma_j$ pues

$$\langle B u, \mathbf{y} \rangle_{\mathbb{R}^p} = \begin{pmatrix} \langle u, \gamma_1 \rangle \\ \vdots \\ \langle u, \gamma_p \rangle \end{pmatrix}^T \mathbf{y} = \sum_{j=1}^p y_j \langle u, \gamma_j \rangle = \langle u, \sum_{j=1}^p y_j \gamma_j \rangle = \langle u, B^* \mathbf{y} \rangle.$$

Por lo tanto, para todo $\mathbf{y} \in \mathbb{R}^p$

$$BB^* \mathbf{y} = B \left(\sum_{j=1}^p y_j \gamma_j \right) = \begin{pmatrix} \left\langle \sum_{j=1}^p y_j \gamma_j, \gamma_1 \right\rangle \\ \vdots \\ \left\langle \sum_{j=1}^p y_j \gamma_j, \gamma_p \right\rangle \end{pmatrix}^T = \mathbf{y}$$

de donde obtenemos que $B : \mathcal{H} \rightarrow \mathcal{H}_1$ es un operador lineal sub-unitario.

Reparo Caso \mathbb{R}^P

Caso Espacio Hilbert

Estimación

Estimadores penalizados

Por el Teorema de Poincaré,

$$\lambda_j(B\Gamma B^*) \leq \lambda_j$$

Definamos $\mathbf{x} = B(X - \mu) = B Y$, luego la matriz de covarianza Σ de \mathbf{x} es igual a

$$\Sigma = B\Gamma B^*$$

de donde

$$\sum_{j=1}^p \text{VAR}(\langle Y, \gamma_j \rangle) = \sum_{j=1}^p \text{VAR}(x_j) = \text{traza}(\Sigma)$$

es decir,

$$\sum_{j=1}^p \text{VAR}(\langle Y, \gamma_j \rangle) = \sum_{j=1}^p \lambda_j(B\Gamma B^*) \leq \sum_{j=1}^p \lambda_j \quad \square$$

Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

Estimación

○○○
○○○○○○○○○○
○○○○○○○

Estimadores penalizados

○○○○○○○○○○
○○

Bondad de ajuste

Observemos que si \mathcal{L}_0 es el subespacio generado por ϕ_1, \dots, ϕ_p

$$\mathbb{E}(\|Y - \pi(Y, \mathcal{L}_0)\|^2) = \sum_{j \geq p+1} \lambda_j$$

Por lo tanto, el porcentaje de varianza explicada

$$\frac{\sum_{j=1}^p \lambda_j}{\sum_{j \geq 1} \lambda_j}$$

da una medida de la bondad del ajuste logrado.

Componentes Principales Funcionales (FPCA)

Observemos que si $\{\phi_\ell\}_{\ell \geq 1}$ es una base ortonormal de \mathcal{H} y definimos

$$U = \sum_{\ell=1}^p \xi_\ell \phi_\ell \quad 1 \leq i \leq n$$

donde las variables ξ_ℓ cumplen

- $\mathbb{E}\xi_\ell = 0$
- $\text{VAR}(\xi_\ell) = \lambda_\ell$
- $\text{Cov}(\xi_\ell, \xi_j) = 0$

Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

A horizontal row of 15 small circles. The 13th circle from the left is filled black, while the others are white with black outlines.

Estimación

○○○○○○○○○○○○○○

Estimadores penalizados

○○○○○○○○○○
○○

Componentes Principales Funcionales (FPCA)

Entonces

- El operador de covarianza de U es

$$\sum_{\ell=1}^p \lambda_\ell \phi_\ell \otimes \phi_\ell$$

- $\xi = (\xi_1, \dots, \xi_p)^T$ tiene matriz de covarianza $\text{diag}(\lambda_1, \dots, \lambda_p)$

Componentes Principales Funcionales (FPCA)

En este sentido, FPCA es una técnica fundamental a ser considerada en datos funcionales.

- FPCA puede usarse para reconstruir las curvas, en el caso de registros esparsos.
- FPCA provee una reducción de dimensión interpretable y natural ya que las curvas se pueden representar por sus escores en una base *pequeña*.
- Provee una reducción de un espacio de dimensión infinita a uno de dimensión finita.
- Si se utiliza un procedimiento robusto de estimación los escores permiten identificar outliers.

Componentes Principales Funcionales (FPCA)

Si estimamos las autofunciones y autovalores, y definimos

$$\hat{\xi}_{i,\ell} = \langle X_i - \hat{\mu}, \hat{\phi}_\ell \rangle$$

Podemos predecir la trayectoria X_i por

$$\hat{X}_i = \hat{\mu} + \sum_{\ell=1}^p \hat{\xi}_{i,\ell} \hat{\phi}_\ell \quad 1 \leq i \leq n$$

- FPCA ofrece un medio efectivo de reducción de dimensión, proveyendo “predicciones” naturales/ trayectorias reconstruidas
- Las curvas se representan por sus escores en una base “pequeña”

Repaso Caso \mathbb{R}^P

oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooooooo
oooooooooooo

Estimación

●○○
○○○○○○○○○○
○○○○○○

Estimadores penalizados

oooooooooooo
oo

Componentes Principales Funcionales Empíricas

Las direcciones principales ϕ_j pueden estimarse por las autofunciones $\hat{\phi}_j$ de

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

y los escores ξ_j pueden predecirse por

$$\hat{\xi}_{i,j} = \langle X_i - \hat{\mu}, \hat{\phi}_j \rangle$$

dando origen a la j -ésima componente principal de X_i .

Repaso Caso \mathbb{R}^P

oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooooooo
oooooooooooo

Estimación

●●○
oooooooooooo
oooooooo

Estimadores penalizados

oooooooooooo
oo

Componentes Principales Funcionales Empíricas

Sean \hat{X}_i son las trayectorias predichas

$$\hat{X}_i = \hat{\mu} + \sum_{\ell=1}^p \hat{\xi}_{i,\ell} \hat{\phi}_{\ell}$$

Las primeras p componentes principales dan el mejor ajuste empírico a subespacios de dimensión p en el sentido que

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \|X_i - \hat{X}_i\|^2 &= \frac{1}{n} \sum_{i=1}^n \|(X_i - \hat{\mu}) - \sum_{\ell=1}^p \hat{\xi}_{i,\ell} \hat{\phi}_{\ell}\|^2 \\ &\leq \frac{1}{n} \sum_{i=1}^n \|(X_i - \hat{\mu}) - \sum_{\ell=1}^p \langle X_i - \hat{\mu}, u_{\ell} \rangle u_{\ell}\|^2 \end{aligned}$$

donde u_1, \dots, u_p son ortonormales.

Repaso Caso \mathbb{R}^P

10

Caso Espacio Hilbert

The diagram consists of three horizontal rows of small circles. The top row contains 10 circles. The middle row contains 6 circles. The bottom row contains 10 circles. All circles are identical in size and color.

Estimación

A diagram consisting of three rows of circles. The top row contains 3 circles, the middle row contains 8 circles, and the bottom row contains 5 circles.

Estimadores penalizados

A horizontal row of 10 small circles, followed by a vertical column of 2 small circles.

Componentes Principales Funcionales Empíricas

Si $\widehat{\lambda}_j$ es el autovalor de $\widehat{\Gamma}$ asociado a $\widehat{\phi}_j$, entonces

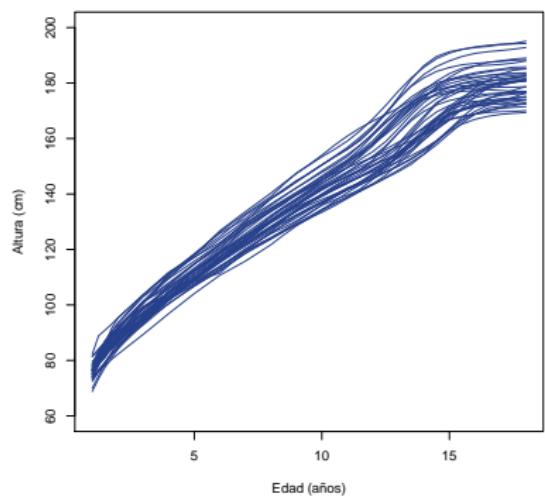
- $\hat{\lambda}_j = 0$ si $j > n$
 - $(1/n) \sum_{i=1}^n \|X_i - \hat{X}_i\|^2 = \sum_{j=p+1}^n \hat{\lambda}_j$
 - El porcentaje de varianza explicado por las primeras p componentes principales empíricas es

$$VE(p) = \frac{\sum_{j=1}^p \hat{\lambda}_j}{\sum_{j=1}^n \hat{\lambda}_j}$$

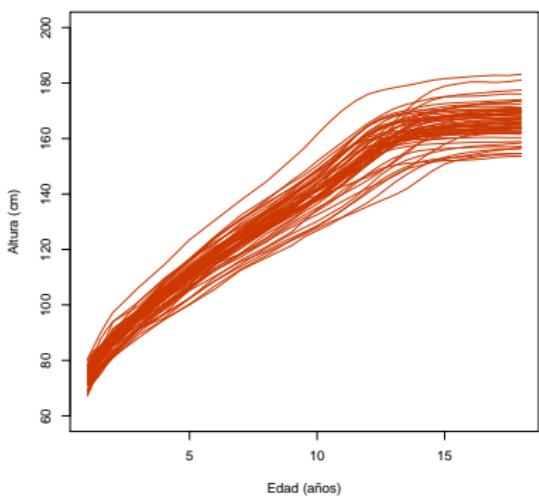
- Elegimos p de modo que $VE(p)$ supere un porcentaje dado, como por ejemplo, 85%.

Ejemplo: Alturas niños

Curvas para 39 varones



Curvas para 54 mujeres



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

Estimación

Estimadores penalizados

10

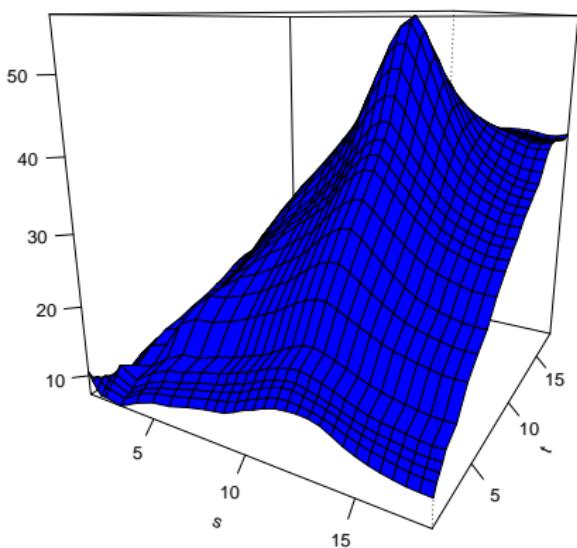
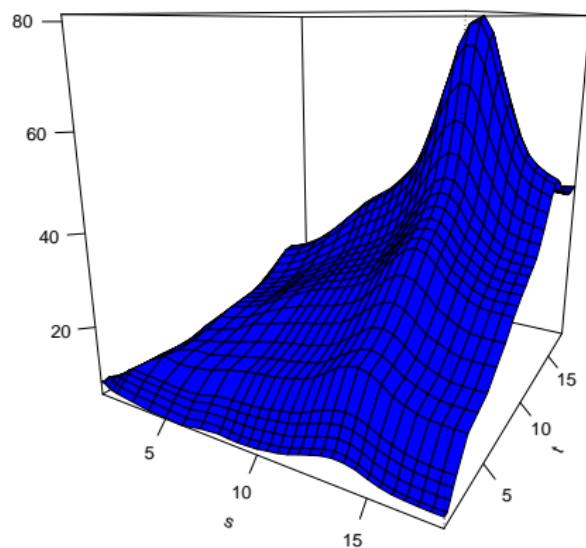
The diagram consists of three horizontal rows of small circles. The top row contains 10 circles arranged in a single line. The middle row contains 5 circles arranged in a single line. The bottom row contains 10 circles arranged in a single line.

10

$$\widehat{\gamma}(s, t)$$

Varones

Mujeres



Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

The diagram consists of three horizontal rows of small circles. The top row contains 15 circles arranged in a single line. The middle row contains 6 circles arranged in a single line. The bottom row contains 12 circles arranged in a single line.

Estimación

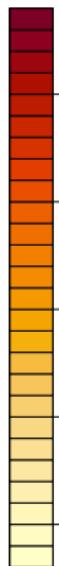
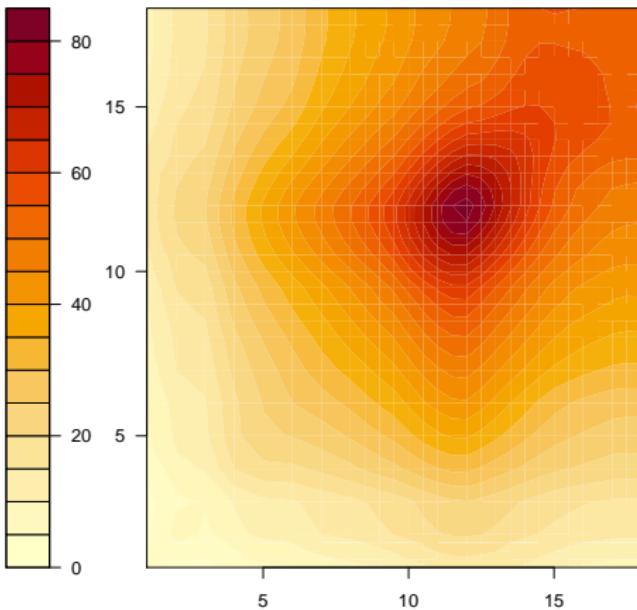
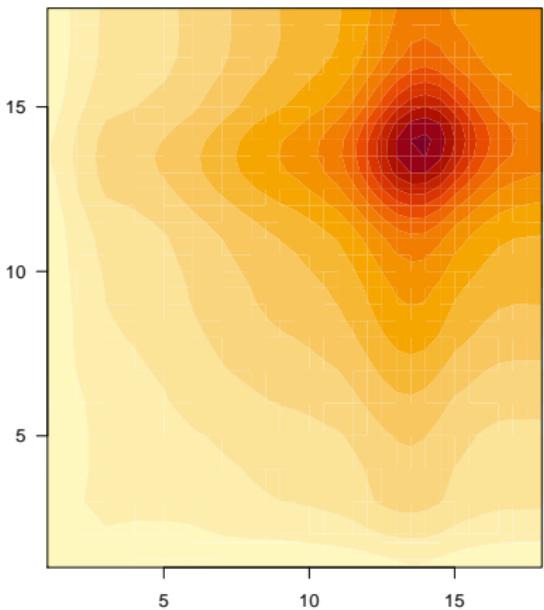
Estimadores penalizados

10

$$\widehat{\gamma}(s, t)$$

Varones

Mujeres



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

Estimación

Estimadores penalizados

10

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.

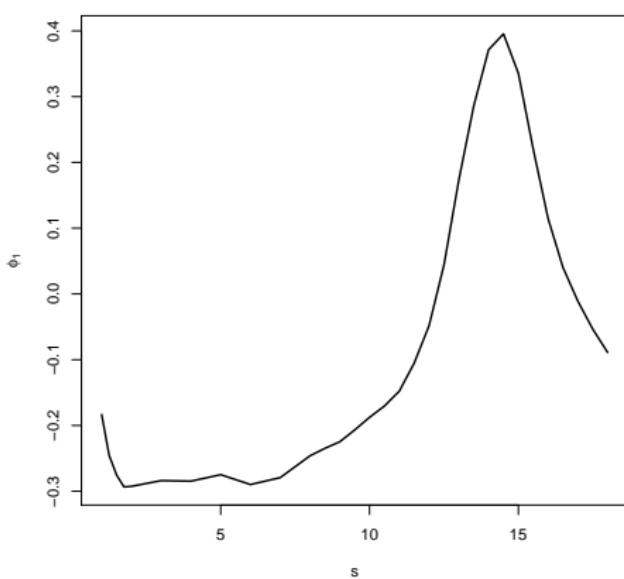
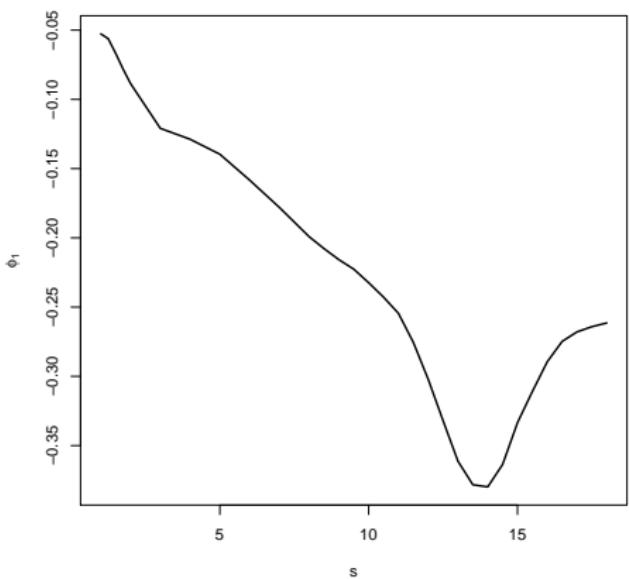
○○○○○○○○○○

○○

$\hat{\phi}_j(t)$ Varones ($EV(2) = 93.6\%$)

$$j = 1$$

j = 2



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

Estimación

Estimadores penalizados

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

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.

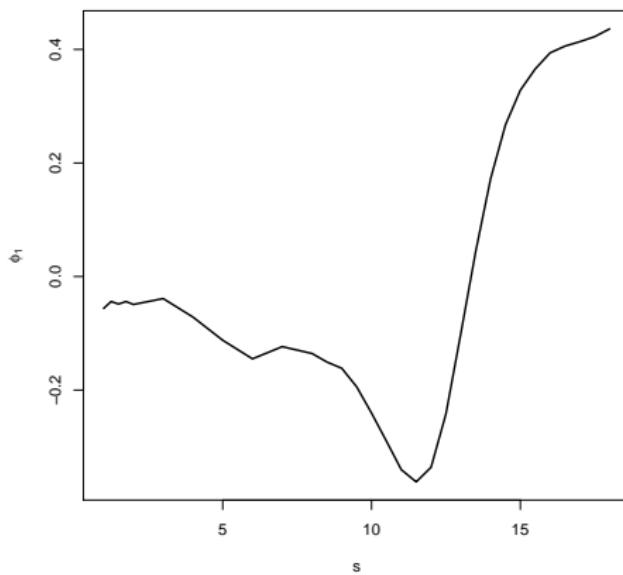
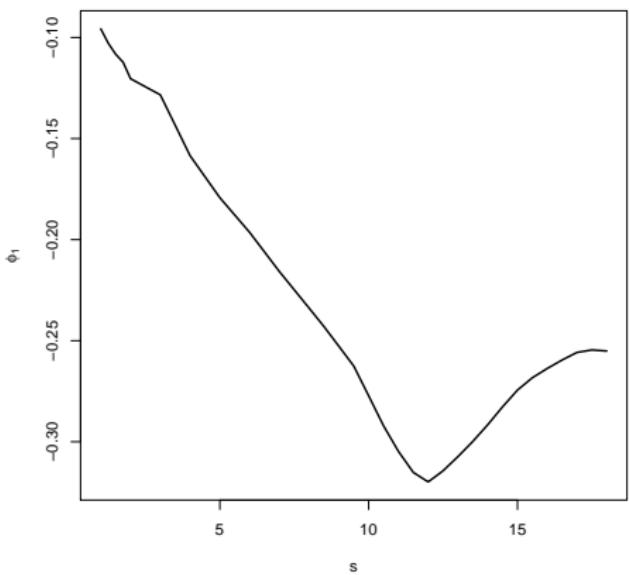
○○○○○○○○○○

○○

$\hat{\phi}_j(t)$ Mujeres ($EV(2) = 95.1\%$)

$$j = 1$$

j = 2



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

Estimación

Estimadores penalizados

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

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.

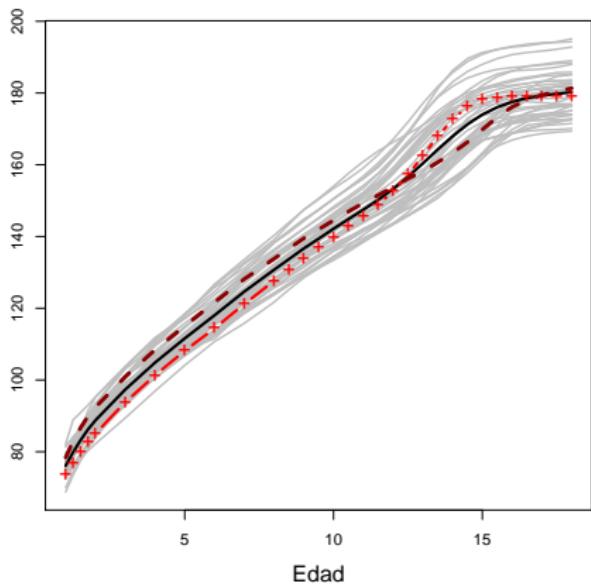
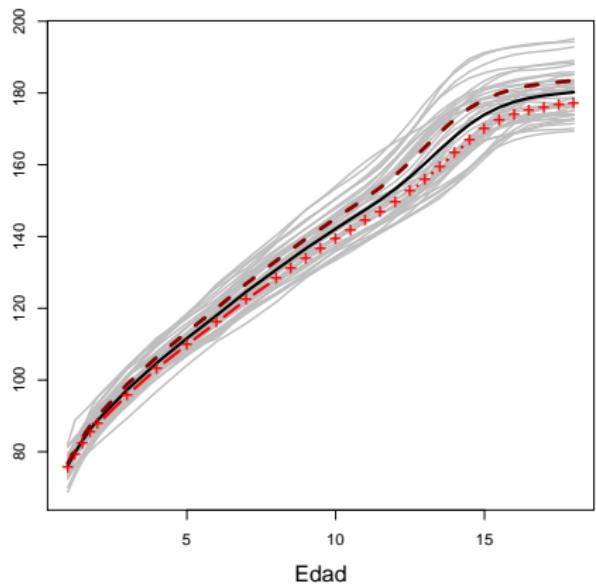
10

100

$$\hat{\mu} \pm \alpha \hat{\phi}_j \text{ Varones}$$

$$j = 1$$

$$j = 2$$



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

Estimación

Estimadores penalizados

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

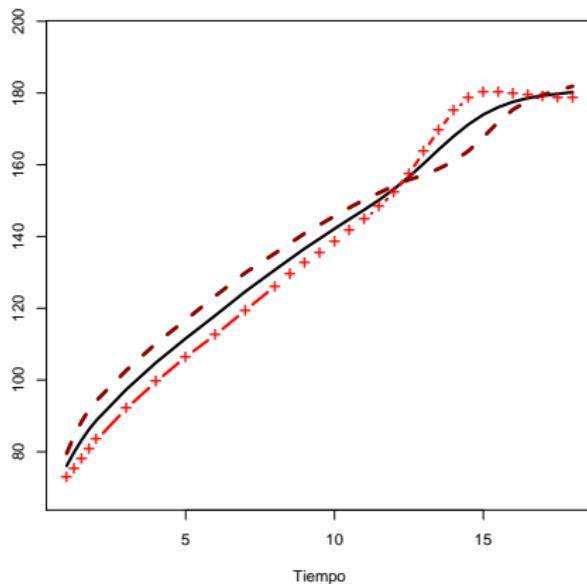
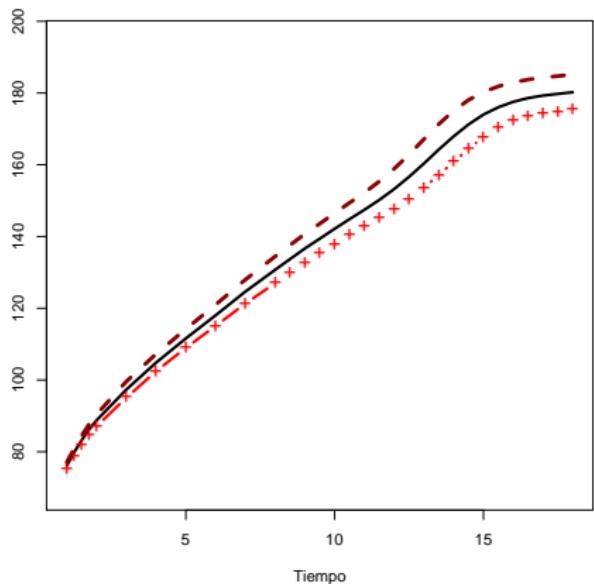
The diagram consists of three horizontal rows of small circles. The top row contains 10 circles arranged in a single line. The middle row contains 5 circles arranged in a single line. The bottom row contains 10 circles arranged in a single line.

A horizontal row of ten small circles, followed by a vertical stack of two small circles.

$$\hat{\mu} \pm \alpha \hat{\phi}_j \text{ Varones}$$

$$j = 1$$

j = 2



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

Estimación

Estimadores penalizados

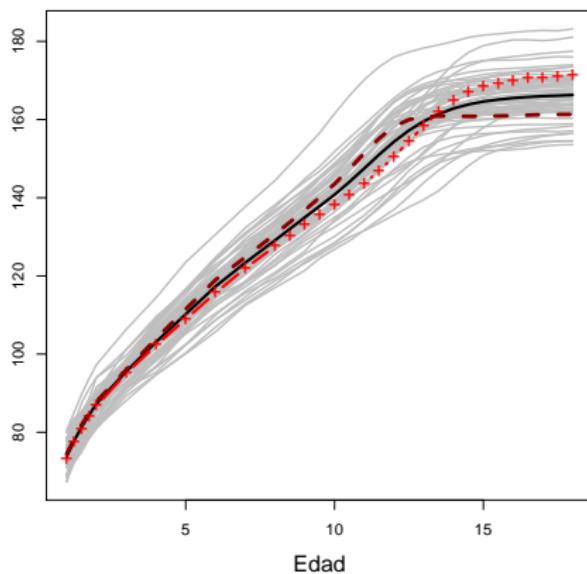
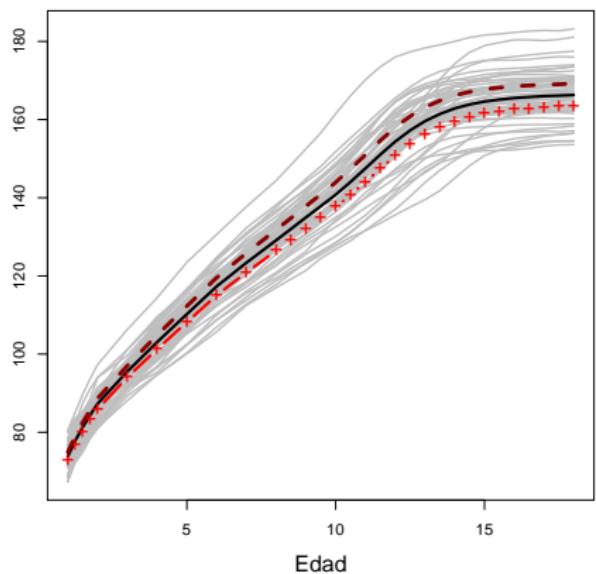
A 2x5 grid of five circles in each row.

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.

$\hat{\mu} \pm \alpha \hat{\phi}_j$ Mujeres

$$j = 1$$

j = 2



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

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 6 circles arranged in a single line. The bottom row contains 8 circles arranged in a single line.

Estimación

10

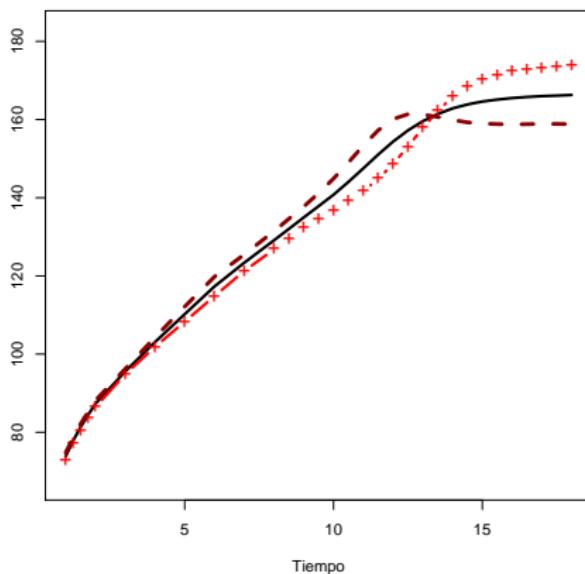
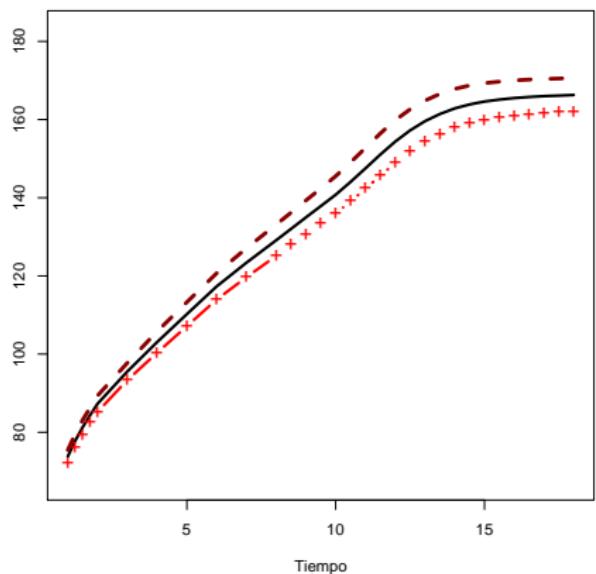
Estimadores penalizados

10

$$\hat{\mu} \pm \alpha \hat{\phi}_j \text{ Mujeres}$$

$$j = 1$$

j = 2



Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

Estimación

○○○
○○○○○○○○○○●
○○○○○○○

Estimadores penalizados

A horizontal row of ten small circles, followed by a vertical stack of two small circles.

Cómo elegir α

Supongamos $X_i \in L^2(\mathcal{I})$ y sea $L = \text{largo}(\mathcal{I})$.

Definamos

$$\bar{\hat{\mu}} = \frac{1}{L} \int_{\mathcal{T}} \hat{\mu}(t) dt$$

$$C^2 = \frac{1}{L} \int_{\mathcal{I}} \left(\widehat{\mu}(t) - \bar{\widehat{\mu}}(t) \right)^2 dt = \frac{1}{L} \left\| \widehat{\mu} - \bar{\widehat{\mu}} \right\|^2$$

Ramsay & Silverman (2005) sugieren elegir

$$\alpha = 0.2 C$$

En el ejemplo anterior elegimos $\alpha = 0.6 C$

Repaso Caso \mathbb{R}^P

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

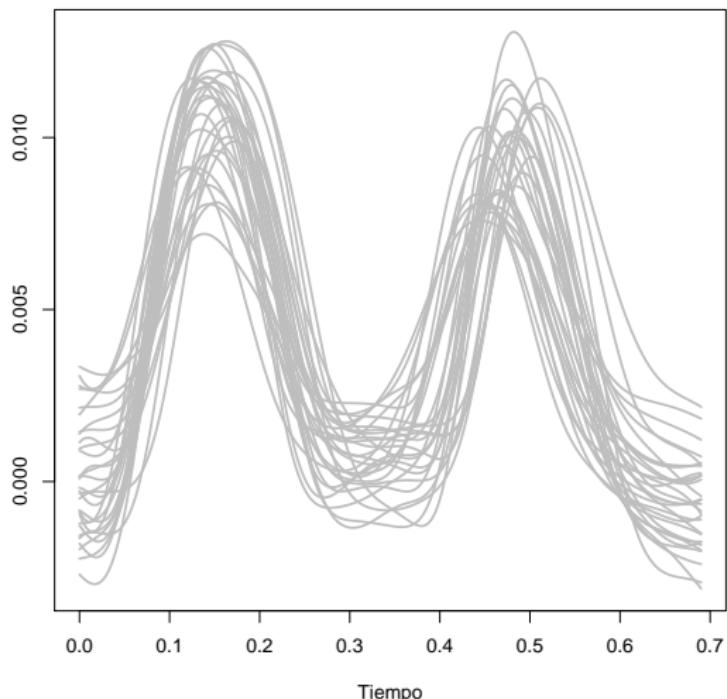
Caso Espacio Hilbert

Estimación

Estimadores penalizados

A horizontal row of 10 small circles, followed by a vertical column of 2 small circles.

Ejemplo: Movimiento Labio Inferior



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

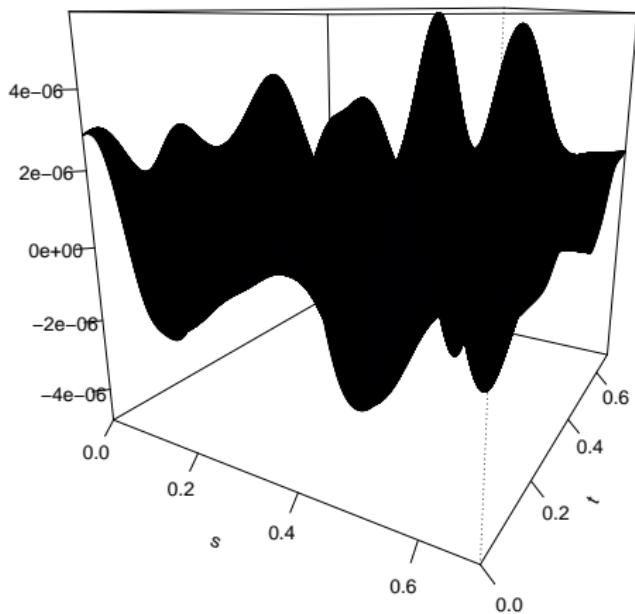
The diagram consists of three horizontal rows of small circles. The top row contains 10 circles arranged in a single line. The middle row contains 6 circles arranged in a single line. The bottom row contains 10 circles arranged in a single line.

Estimación

Estimadores penalizados

○○○○○○○○○○
○○

$$\widehat{\gamma}(s, t)$$



Repaso Caso \mathbb{R}^P

10

Caso Espacio Hilbert

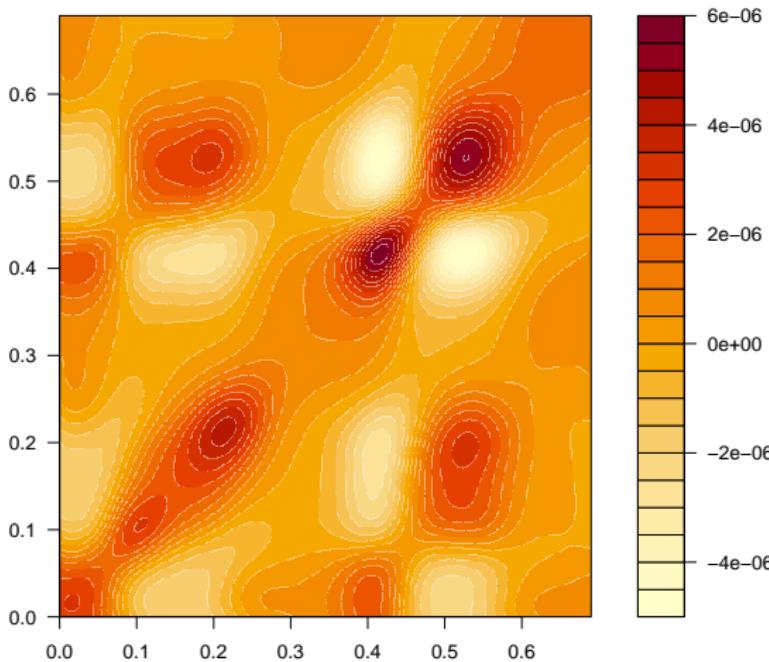
The diagram consists of three horizontal rows of small circles. The top row contains 10 circles arranged in a single line. The middle row contains 6 circles arranged in a single line. The bottom row contains 10 circles arranged in a single line.

Estimación

Estimadores penalizados

100

$$\widehat{\gamma}(s, t)$$



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

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

The diagram consists of three horizontal rows of small circles. The top row contains 10 circles. The middle row contains 6 circles. The bottom row contains 10 circles.

Estimación

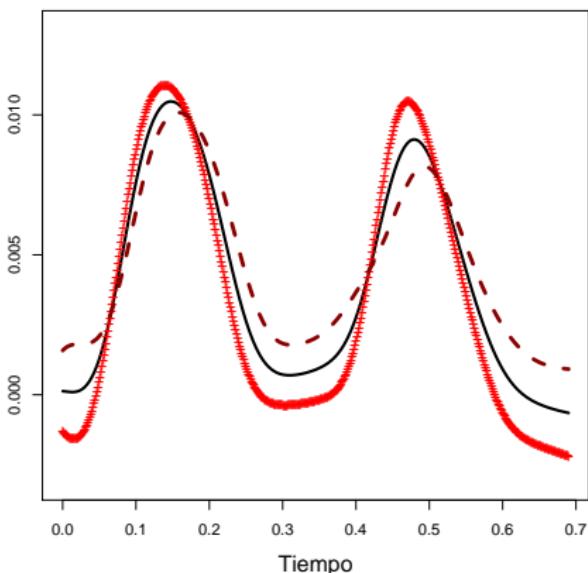
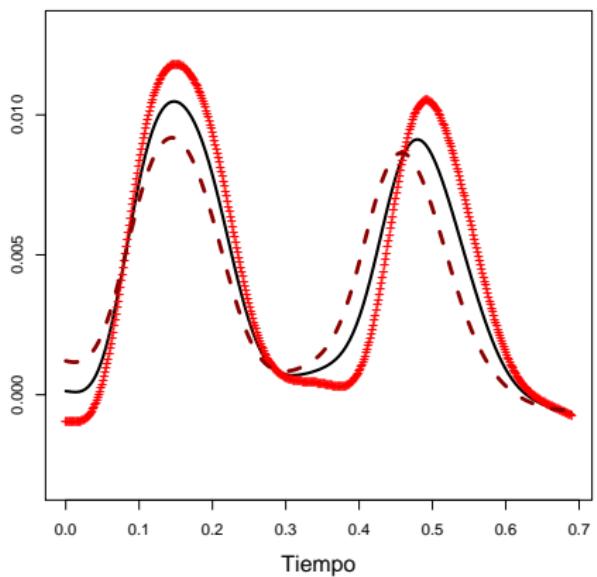
Estimadores penalizados

○○○○○○○○○○
○○

$\hat{\mu} \pm \alpha \hat{\phi}_j$: $EV(4) = 86.5\%$, $\alpha = 0.7$

$$j = 1$$

$$j=2$$



Repaso Caso \mathbb{R}^P

Caso Espacio Hilbert

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

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.

Estimación

A sequence of 10 circles arranged in three rows: top row has 3 circles, middle row has 6 circles, bottom row has 1 circle in the center.

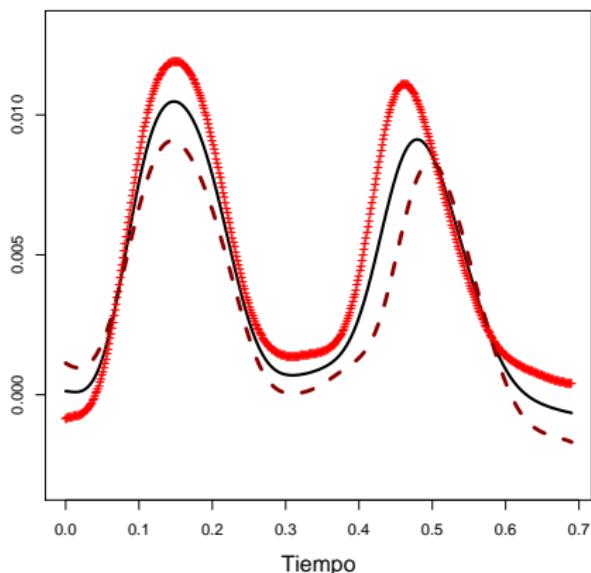
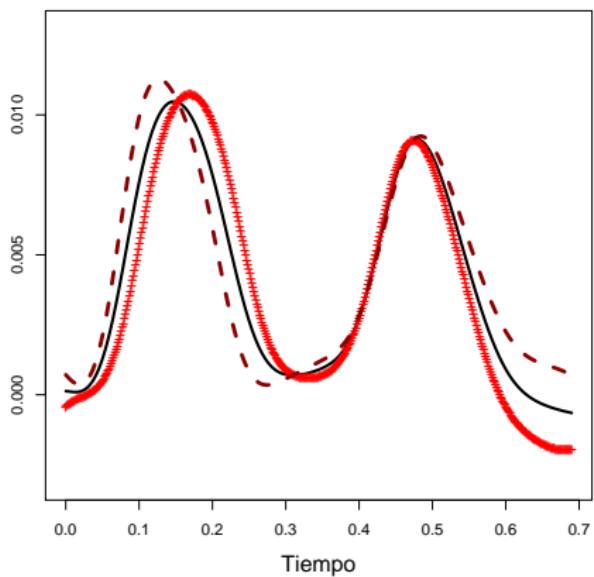
Estimadores penalizados

10

$\hat{\mu} \pm \alpha \hat{\phi}_j$: $EV(4) = 86.5\%$, $\alpha = 0.7$

$$j = 3$$

$j = 4$



Repaso Caso \mathbb{R}^P

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

Caso Espacio Hilbert

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.

Estimación

○○○○○○○○○○

Estimadores penalizados

○○○○○○○○○○
○○

El j -ésimo modo de variación es el conjunto de funciones

$$\mu(t) \pm \alpha\sqrt{\lambda_j}\phi_j(t) \quad t \in \mathcal{I} \quad \alpha \in [-A, A]$$

que se analizan simultáneamente sobre un rango de valores de α , usualmente para $A = 2$

En la práctica se reemplazan los valores desconocidos por sus estimadores obteniendo los modos de variación empíricos

$$\hat{\mu}(t) \pm \alpha \sqrt{\hat{\lambda}_j} \hat{\phi}_j(t) \quad t \in \mathcal{I} \quad \alpha \in [-A, A]$$

Repaso Caso \mathbb{R}^P

10

Caso Espacio Hilbert

The diagram consists of three horizontal rows of small circles. The top row contains 15 circles arranged in a single line. The middle row contains 6 circles arranged in a single line. The bottom row contains 10 circles arranged in a single line.

Estimación

Estimadores penalizados

10

Primer Modo de variación

Segundo Modo de variación

Tercer Modo de variación

Cuarto Modo de variación

Componentes Principales Suavizadas

- Algunas veces en lugar de las direcciones principales empíricas, interesa obtener direcciones principales suaves.
- Uno de los argumentos fundamentales es que el suavizado es una herramienta de regularización que puede revelar comportamientos más interpretables e interesantes de los modos de variación para datos funcionales.
- Una opción es usar un **Suavizador basado en núcleos**
- Otra posibilidad es usar **Penalización**

Componentes Principales Penalizadas

- $D : \mathcal{H}_S \rightarrow \mathcal{H}$ un operador lineal (“diferenciador”)
 - Por ejemplo, si $\mathcal{H} = L^2(\mathcal{I})$

$$D\alpha \equiv \alpha''$$

- D permite definir un producto interno de penalización

$$[\alpha, \beta] = \langle D\alpha, D\beta \rangle$$

- Por ejemplo, si $\mathcal{H} = L^2(\mathcal{I})$

$$[\alpha, \beta] = \int_{\mathcal{H}} \alpha''(t) \beta''(t) dt$$

Componentes Principales Penalizadas

- “Operador de penalización” $\Psi : \mathcal{H}_S \rightarrow \mathbb{R}$,

$$\Psi(\alpha) = [\alpha, \alpha]$$

- Si $\mathcal{H} = L^2(\mathcal{I})$ y $D\alpha = \alpha''$ entonces

$$\Psi(\alpha) = [\alpha, \alpha] = \int_{\mathcal{I}} (\alpha''(t))^2 dt$$

- Producto interno penalizado

$$\langle \alpha, \beta \rangle_\tau = \langle \alpha, \beta \rangle + \tau [\alpha, \beta]$$

- Norma penalizada

$$\|\alpha\|_\tau^2 = \|\alpha\|^2 + \tau \Psi(\alpha)$$

Componentes Principales Penalizadas

- Si $\mathcal{H} = L^2(\mathcal{I})$ y $D\alpha = \alpha''$ entonces

$$\langle \alpha, \beta \rangle_{\tau} = \int_{\mathcal{I}} \alpha(t) \beta(t) dt + \tau \int_{\mathcal{I}} \alpha''(t) \beta''(t) dt$$

$$\|\alpha\|_{\tau}^2 = \int_{\mathcal{I}} \alpha^2(t) dt + \tau \int_{\mathcal{I}} (\alpha''(t))^2 dt$$

Componentes Principales Penalizadas

Dos enfoques

- Rice & Silverman (1991) proponen penalizar por rugosidad la variabilidad $s_n^2(\alpha)$
- Silverman (1996) propone penalizar por rugosidad la norma
- Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, **24**, 1-24.
- Rice, J. & Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B*, **53**, 233-243.

Ver función `fdata2pc` de la librería `fda.usc`

Repaso Caso \mathbb{R}^P

10

Caso Espacio Hilbert

Estimación

Estimadores penalizados

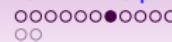
Componentes Principales Penalizadas

Sean

$$s_n^2(\alpha) = \frac{1}{n} \sum_{i=1}^n (\langle X_i - \bar{X}, \alpha \rangle)^2$$

$$L_{n,\tau}(\alpha) = s_n^2(\alpha) - \tau \Psi(\alpha)$$

$$= s_n^2(\alpha) - \tau \int_{\mathcal{I}} (\alpha''(t))^2 dt$$



FPCA Penalizadas: Rice & Silverman (1991)

- Penalizan la variabilidad por rugosidad

$$\begin{aligned}\widehat{\phi}_{1,\tau_1} &= \underset{\|\alpha\|=1}{\operatorname{argmax}} \{L_{n,\tau_1}(\alpha)\} \\ &= \underset{\|\alpha\|=1}{\operatorname{argmax}} \left\{ s_n^2(\alpha) - \tau_1 \int_{\mathcal{I}} (\alpha''(t))^2 dt \right\}\end{aligned}$$

- Los estimadores de las siguientes direcciones se definen como

$$\begin{aligned}\widehat{\phi}_{k+1,\tau_{k+1}} &= \underset{\substack{\|\alpha\|=1 \\ \langle \alpha, \widehat{\phi}_j \rangle = 0, 1 \leq j \leq k}}{\operatorname{argmax}} \{L_{n,\tau_{k+1}}(\alpha)\} \\ &= \underset{\substack{\|\alpha\|=1 \\ \langle \alpha, \widehat{\phi}_j \rangle = 0, 1 \leq j \leq k}}{\operatorname{argmax}} \left\{ s_n^2(\alpha) - \tau_{k+1} \int_{\mathcal{I}} (\alpha''(t))^2 dt \right\}\end{aligned}$$

Diversos estudios muestran que a medida que k aumenta es conveniente usar valores más chicos de τ_k .

FPCA Penalizadas: Rice & Silverman (1991)

$$L_{n,\tau}(\alpha) = s_n^2(\alpha) - \tau \Psi(\alpha)$$

- Si aumentamos la rugosidad de α para un τ fijo, aumenta $\Psi(\alpha)$ y por lo tanto decrece la función objetivo $L_{n,\tau}(\alpha)$.
- Observemos que, si $\tau \rightarrow 0$, $L_{n,\tau}(\alpha)$ converge a la varianza muestral $s_n^2(\alpha)$.

Por lo tanto, $\hat{\phi}_{1,\tau}$ se acerca a la dirección principal empírica.

Repaso Caso \mathbb{R}^P

oooooo
oooooo

Caso Espacio Hilbert

oooooooooooooooooooo
oooooooo
oooooooooooo

Estimación

ooo
oooooooooooo
ooooooo

Estimadores penalizados

oooooooo●ooo
oo

FPCA Penalizadas: Rice & Silverman (1991)

$$L_{n,\tau}(\alpha) = s_n^2(\alpha) - \tau \Psi(\alpha)$$

- Por otra parte, cuanto mayor sea τ , más penalizada se verá $s_n^2(\alpha)$ por la rugosidad de α
- En el límite cuando $\tau \rightarrow \infty$, lo que domina es la penalidad y por lo tanto, la dirección principal deberá tener

$$\Psi(\widehat{\phi}_{1,\tau}) = 0$$

Repaso Caso \mathbb{R}^P

10

Caso Espacio Hilbert

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 6 circles arranged in a single line. The bottom row contains 7 circles arranged in a single line.

Estimación

Estimadores penalizados

10

FPCA Penalizadas: Silverman (1996)

- Propone penalizar por rugosidad la norma

$$\hat{\phi}_{1,\tau} = \operatorname{argmax}_{\|\alpha\|_\tau=1} s_n^2(\alpha) = \operatorname{argmax}_{\alpha \neq 0} \frac{s_n^2(\alpha)}{\|\alpha\|^2 + \tau \Psi(\alpha)}$$

$$= \operatorname{argmax}_{\alpha \neq 0} \frac{s_n^2(\alpha)}{\int_{\mathcal{T}} \alpha^2(t) dt + \tau \int_{\mathcal{T}} (\alpha''(t))^2 dt}$$

- Las siguientes direcciones se definen como

$$\hat{\phi}_{k+1,\tau} = \operatorname{argmax}_{\|\alpha\|_\tau=1} s_n^2(\alpha) = \operatorname{argmax}_{\alpha \neq 0} \frac{s_n^2(\alpha)}{\|\alpha\|^2 + \tau \Psi(\alpha)}$$

$$\langle \alpha, \hat{\phi}_j \rangle_\tau = 0, \forall 1 \leq j \leq k$$

Las direcciones $\hat{\phi}_j$ son ortonormales respecto del producto interno $\langle \cdot, \cdot \rangle_\tau$

FPCA Penalizadas: Silverman (1996)

- Silverman (1996) muestra que los estimadores $\hat{\phi}_{j,\tau}$ son consistentes si $\tau = \tau_n \rightarrow 0$, es decir,

$$\left\| s_j \frac{\hat{\phi}_{j,\tau}}{\|\hat{\phi}_{j,\tau}\|} - \phi_j \right\| \xrightarrow{a.s.} 0$$

donde $s_j = \text{signo}(\langle \hat{\phi}_{j,\tau}, \phi_j \rangle)$

- Qi & Zhao (2011) obtienen distribución asintótica de los estimadores.

Qi, X. & Zhao, H. (2011). Some theoretical properties of Silverman's method for Smoothed functional principal component analysis. *Journal of Multivariate Analysis*, **102**, 741-767.

Elección de τ

La elección de τ juega un papel importante en la práctica porque determina un balance entre rugosidad y buen ajuste predictivo.

Supongamos que nos interesa un número fijo, ℓ , de componentes.

1. Centre los datos, o sea, defina $X_i^{(c)} = X_i - \hat{\mu}$ donde

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

2. Para cada $1 \leq m \leq \ell$ y $0 < \tau$, sea

$$\hat{\phi}_{m,\tau}^{(-i)}$$

el estimador de la m -ésima dirección principal calculado sin la observación i .

Elección de τ

3. Sea $\widehat{\mathcal{L}}_{\ell}^{(-i)}$ el subespacio generado por $\widehat{\phi}_{1,\tau}^{(-i)}, \dots, \widehat{\phi}_{\ell,\tau}^{(-i)}$.

Defina para $1 \leq i \leq n$

$$\widehat{X}_{i,\tau} = \pi(X_i^{(c)}; \widehat{\mathcal{L}}_\ell^{(-i)}) = \sum_{m=1}^{\ell} \left\langle X_i^{(c)}, \widehat{\phi}_{m,\tau}^{(-i)} \right\rangle \widehat{\phi}_{m,\tau}^{(-i)}$$

$$X_i^\perp(\tau) = X_i^{(c)} - \hat{X}_{i,\tau}$$

- #### 4. Sea

$$CV_\ell(\tau) = \sum_{i=1}^n \|X_i^\perp(\tau)\|^2$$

Elegimos $\hat{\tau}$ como el valor τ que minimiza $CV_\ell(\tau)$.