# Functional Data Analysis: Estimación y Boxplot

**Graciela Boente**

# Estimación

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

- Sea $\mu = \mathbb{E}X$ y $\Gamma = \mathbb{E}\left\{(X-\mu)\otimes(X-\mu)\right\}$.

- $\mu$ puede estimarse usando la media muestral

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

- $\Gamma$ puede estimarse usando el operador de covarianza muestral

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

# 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 $\Gamma$ 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 $\gamma(s,t) = min(s,t) - t\,s$ entonces

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

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

# Propiedades

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

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

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

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

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

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

- Sea $Z_i = X_i - \mu$ entonces

$$\mathbb{E}(\|\widehat{\mu}-\mu\|^2 = \mathbb{E}(\|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\|^2 = \mathbb{E}(\|\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\|^2$$

$$= \frac{1}{n^2}\mathbb{E}\langle\sum_{i=1}^{n} Z_i, \sum_{j=1}^{n} Z_j\rangle = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbb{E}\langle Z_i, Z_j\rangle$$

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

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

pues

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

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

$$\mathbb{E}(\|\widehat{\mu}-\mu\|^2 = \frac{1}{n^2}\sum^{n}\mathbb{E}\langle Z_i, Z_i\rangle = \frac{1}{n^2}\sum^{n}\mathbb{E}\|Z_i\|^2 = \frac{1}{n}\mathbb{E}\|X-\mu\|^2$$

# Propiedades

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

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

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

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

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

- Sea $\mu = \mathbb{E}X$ y $\Gamma = \mathbb{E}\left\{(X-\mu)\otimes(X-\mu)\right\}$.

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

# Espacios de operadores

- $\mathcal{H}$ espacio de Hilbert separable

- Let $\Upsilon : \mathcal{H} \to \mathcal{H}$ un operador lineal acotado ($\Upsilon \in \mathcal{L}$)

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

- El operador $\Upsilon$ 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 $\Upsilon$ 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 $\Upsilon_1$ y $\Upsilon_2$ 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

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

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, omando como base la base ortonormal de las autofunciones de $\Upsilon$,

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

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 $\Gamma$ es el operador de covarianza con núcleo de covarianza

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

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 $\widehat{\Gamma}$

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

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

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

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

# Propiedades de $\widehat{\Gamma}$

- $X_i : \Omega \to \mathcal{H}$, $1 \leq i \geq n$, 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 $\Gamma$ con autovalores asociados $\{\lambda_\ell : \ell \geq 1\}$ tales que $\lambda_\ell \geq \lambda_{\ell+1}$.
- $\xi_m$ las coordenadas estandarizadas de $X - \mu$ en la base $\{\phi_{i,\ell} : \ell \geq 1\}$, o sea,

$$\xi_m = \lambda_m^{-1/2} \langle X_i - \mu_i, \phi_{i,m} \rangle$$

- Por lo tanto,

$$\mathbb{E}(\xi_m) = 0 \qquad \mathbb{E}(\xi_m^2) = 1$$

$$\mathbb{E}(\xi_m \xi_s) = 0 \quad \text{para} \quad m \neq s$$

# Propiedades de $\widehat{\Gamma}$

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

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

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

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

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

# Propiedades de $\widehat{\phi}_j$

- $X_i : \Omega \to \mathcal{H}$, $1 \leq i \geq n$, 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 $\lambda_1 > \lambda_2 > \cdots > \lambda_p > \lambda_{p+1}$
- $\widehat{\lambda}_j \xrightarrow{a.s.} \lambda_j$, $j = 1, \ldots p$
- signo $\left(\langle \widehat{\phi}_j, \phi_j \rangle\right) \widehat{\phi}_j \xrightarrow{a.s.} \phi_j$, , $j = 1, \ldots p$

# Propiedades de $\widehat{\phi}_j$

- $X_i : \Omega \to \mathcal{H}$, $1 \leq i \geq n$, 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 $\widehat{\lambda}_j$.

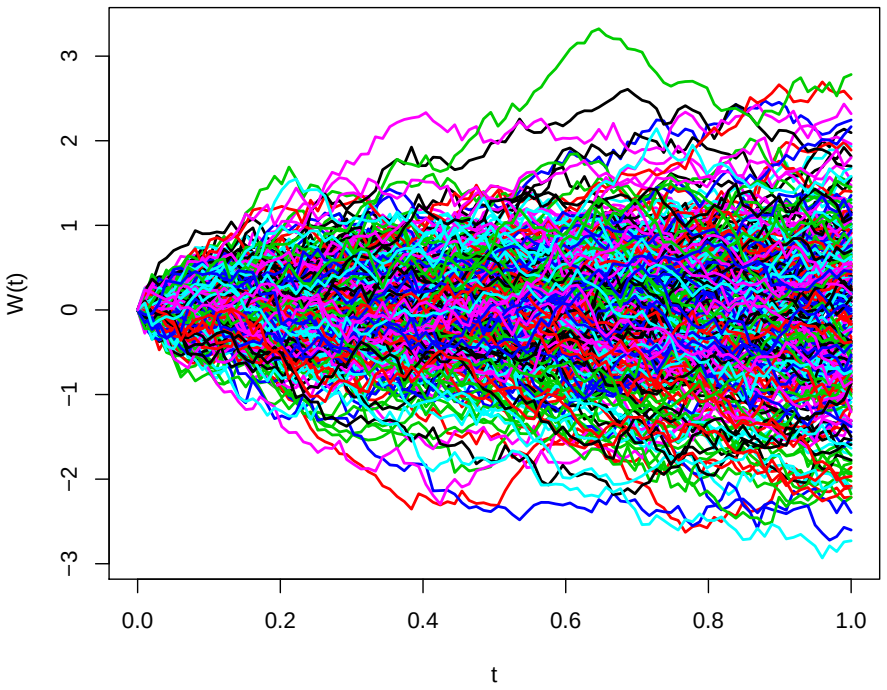
$$\limsup n\,\mathbb{E}\left\{\left(\widehat{\lambda}_j - \lambda_j\right)^2\right\} < \infty$$

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

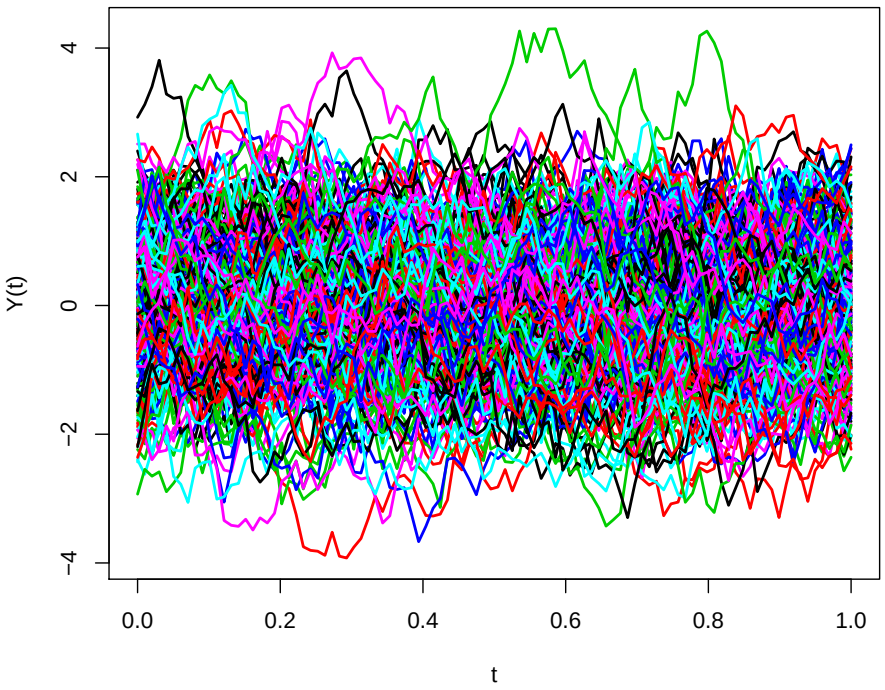
- Dauxois *et al.* (1982) obtuvieron la distribución asintótica de $\sqrt{n}\left(\widehat{\phi}_j - \phi_j\right)$

# Ejemplos

Wiener

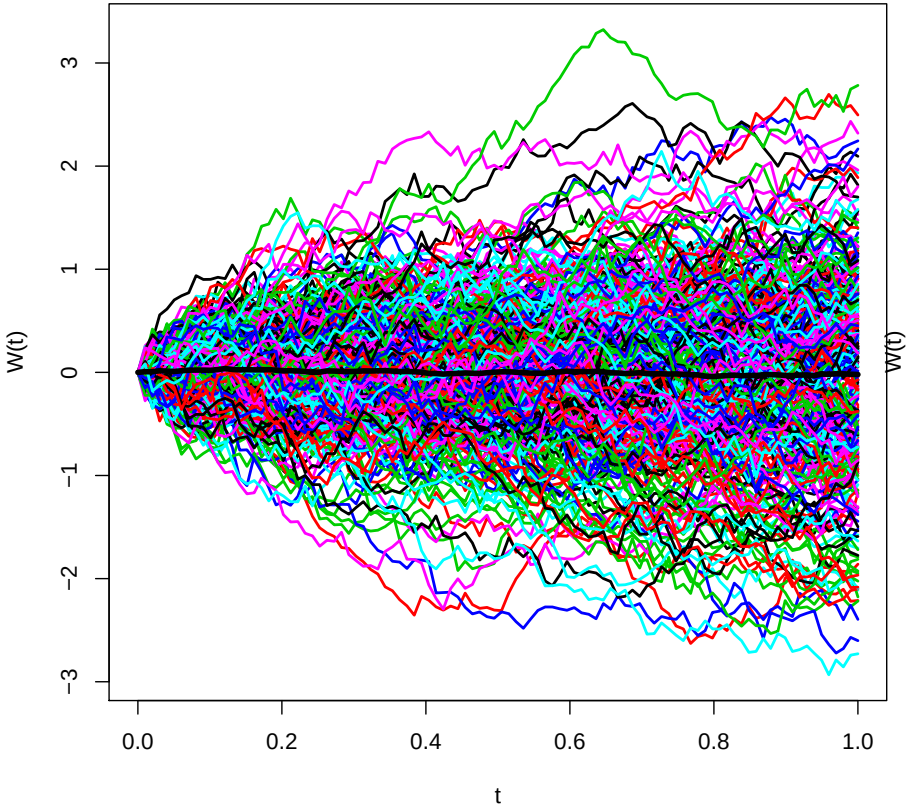



$Y(t)$

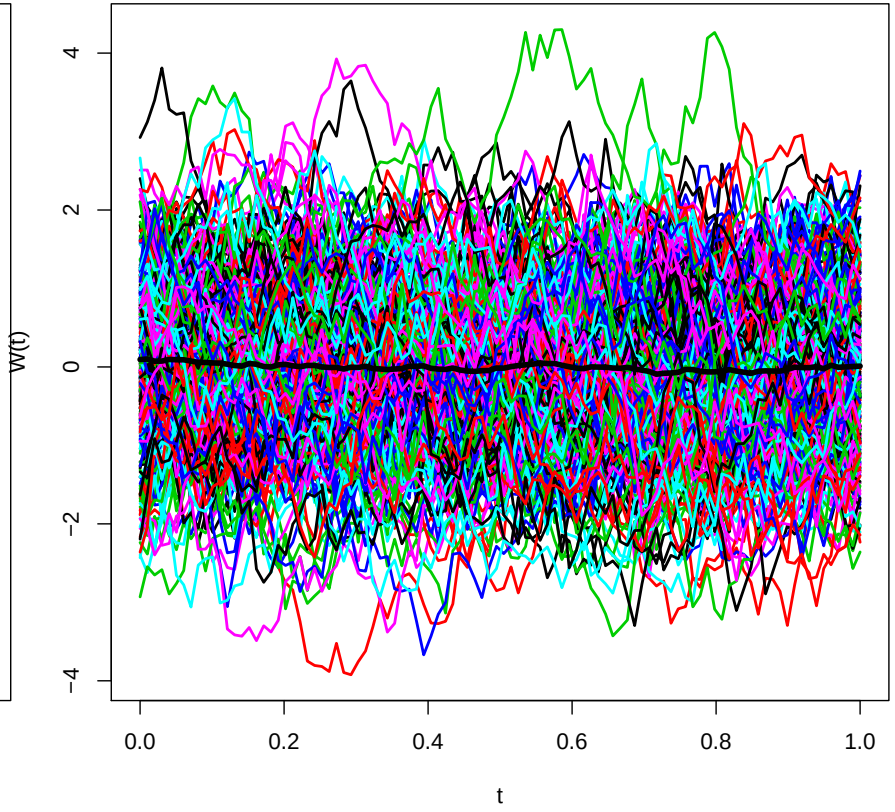


# Ejemplos

Wiener

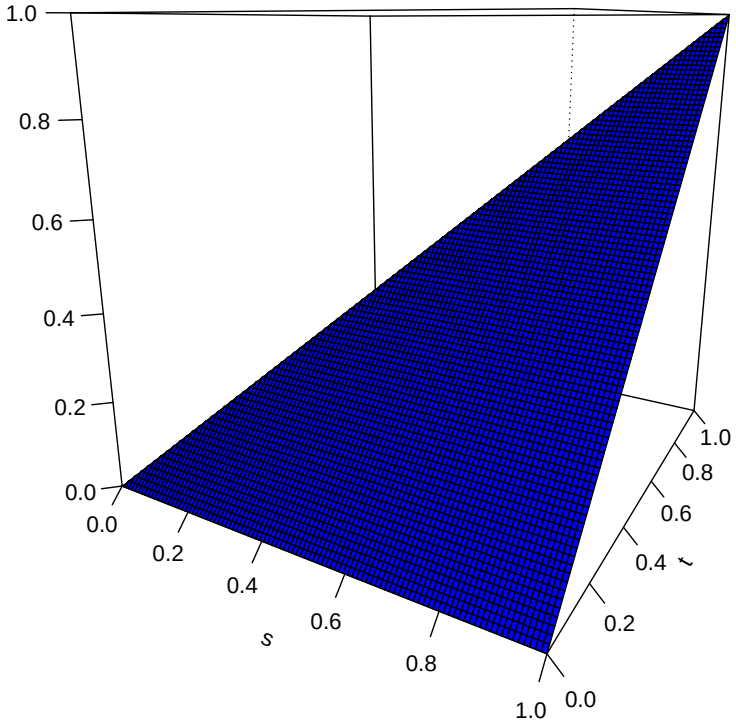



$Y(t)$

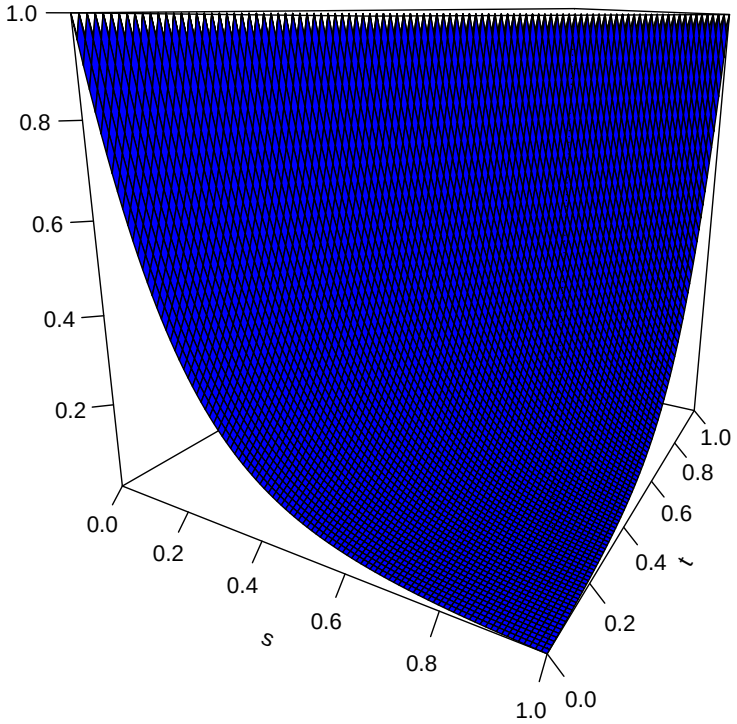


# Ejemplos: $\Gamma$

Wiener

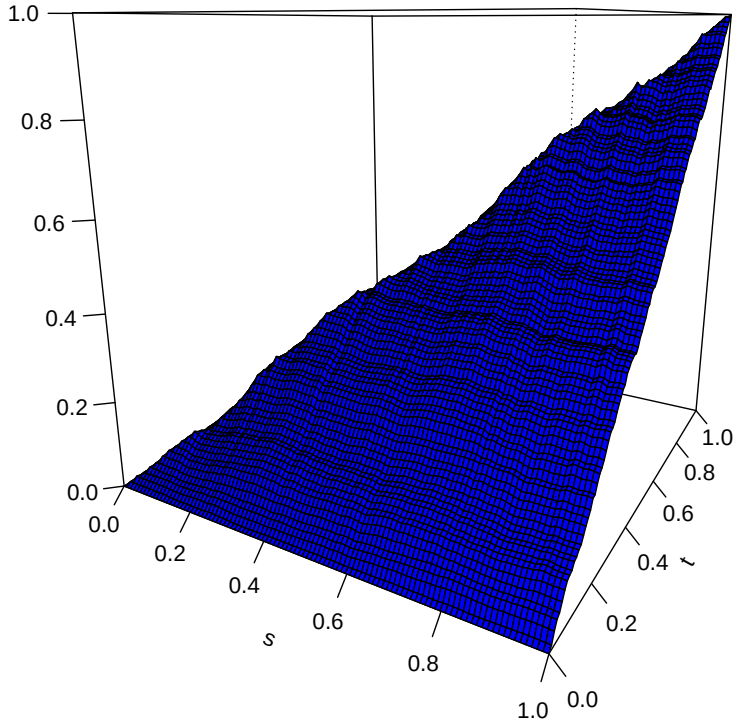



$Y(t)$

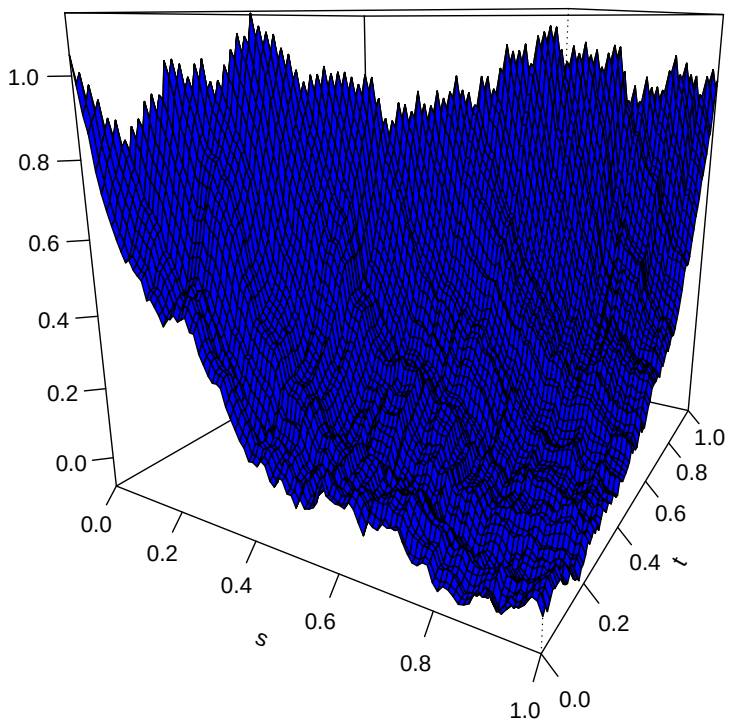


# Ejemplos: $\widehat{\Gamma}$

Wiener

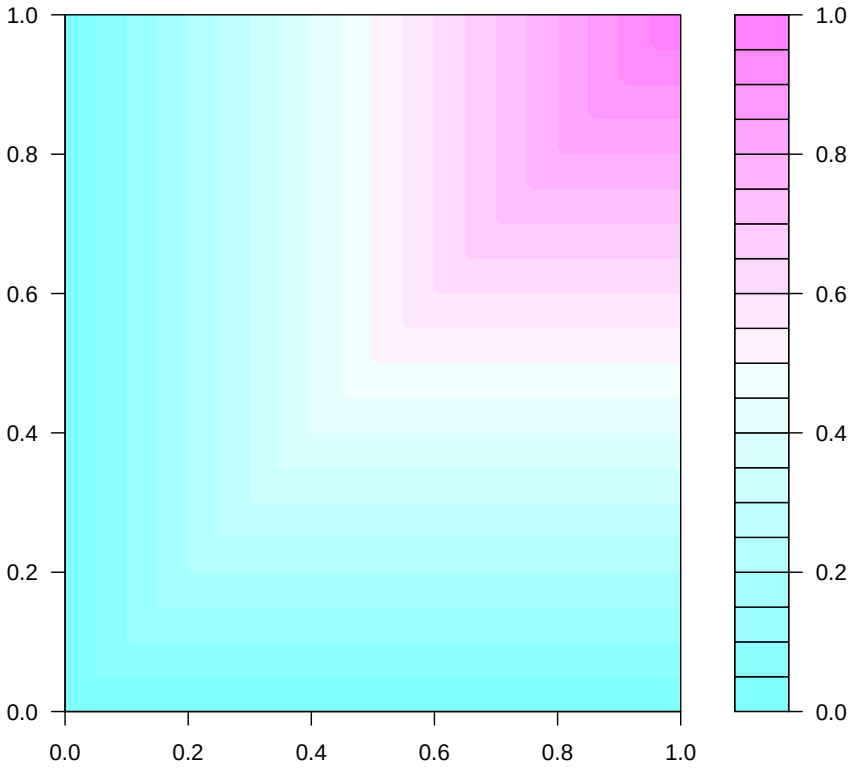



$Y(t)$

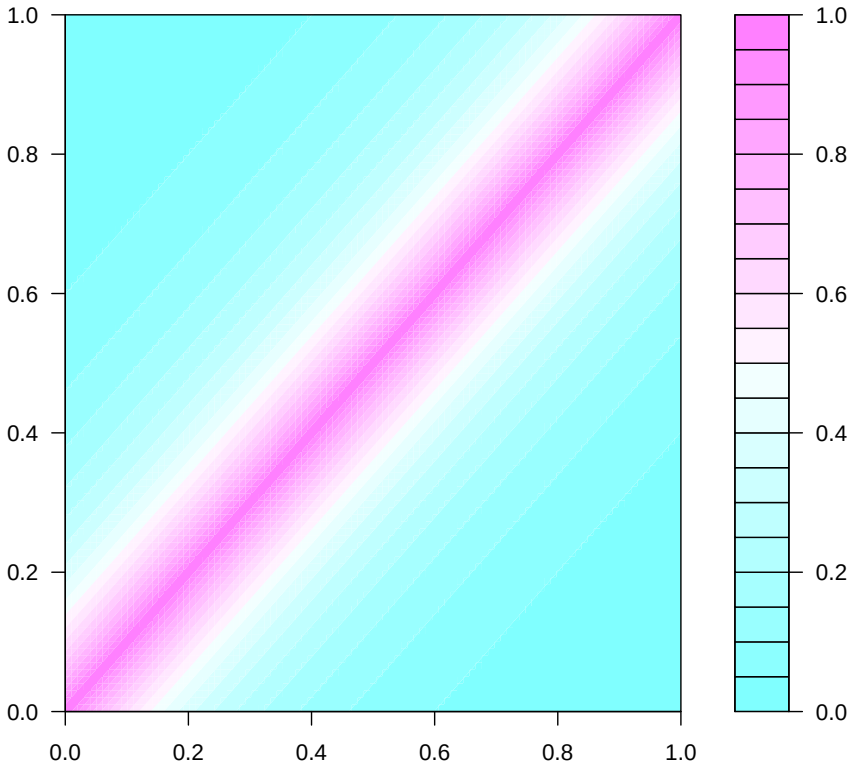


# Ejemplos: $\Gamma$

Wiener

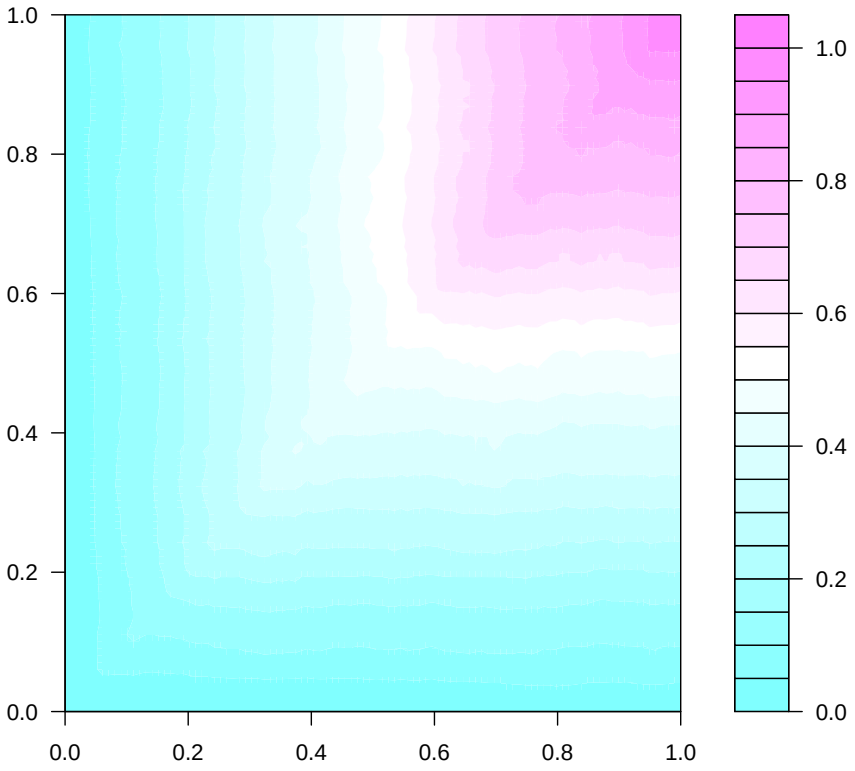



$Y(t)$

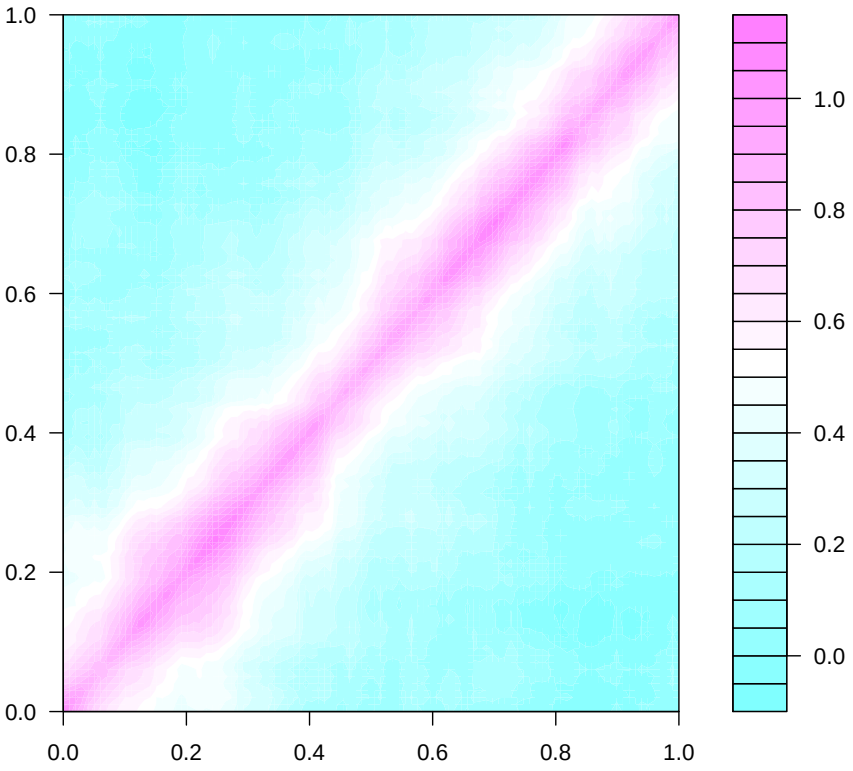


# Ejemplos: $\widehat{\Gamma}$

Wiener

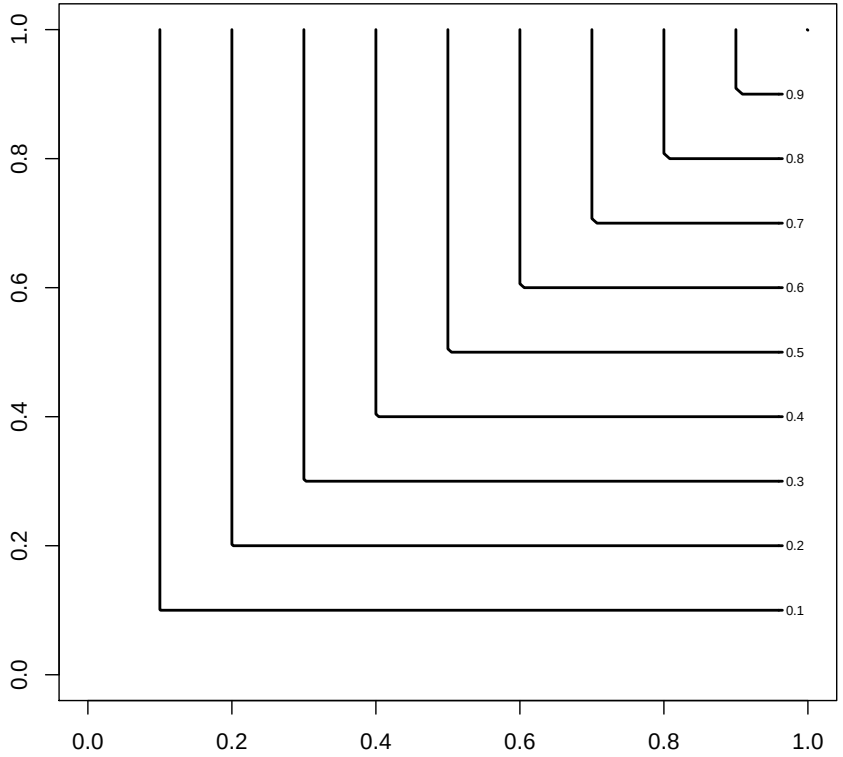



$Y(t)$

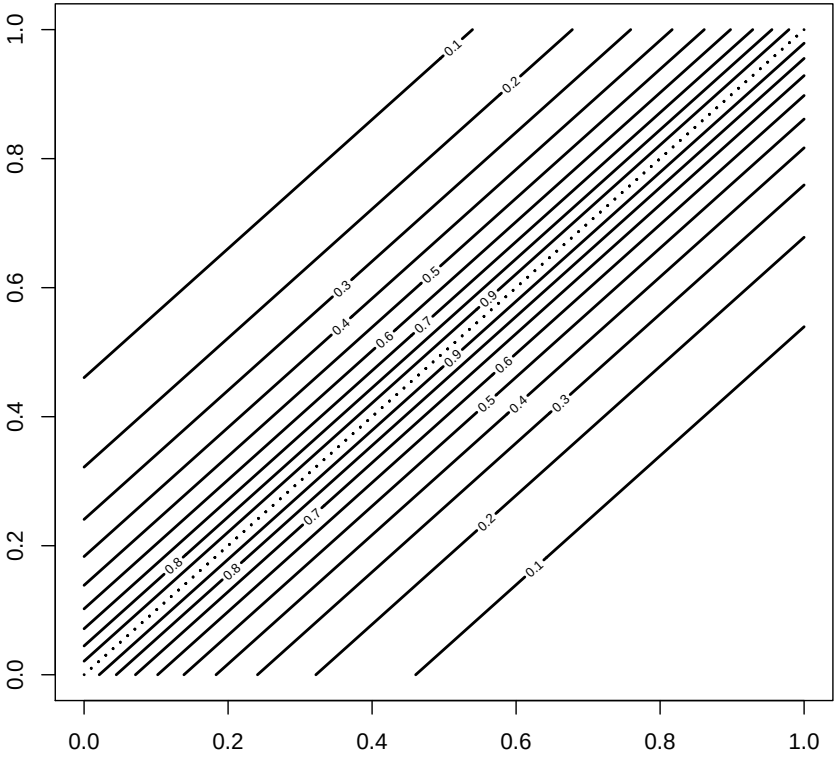


# Ejemplos: $\Gamma$

Wiener

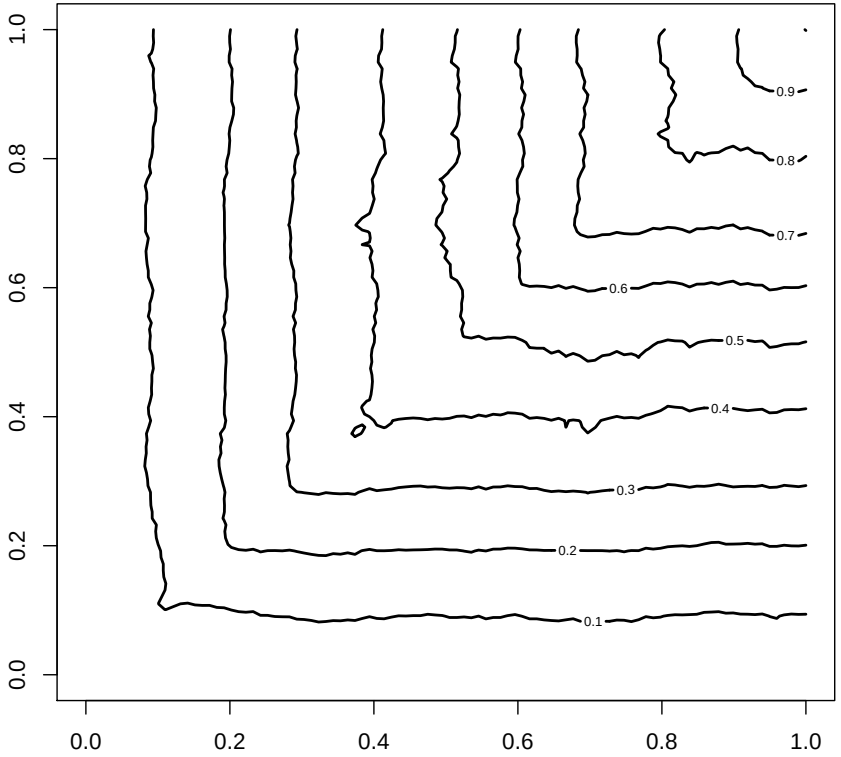



$Y(t)$

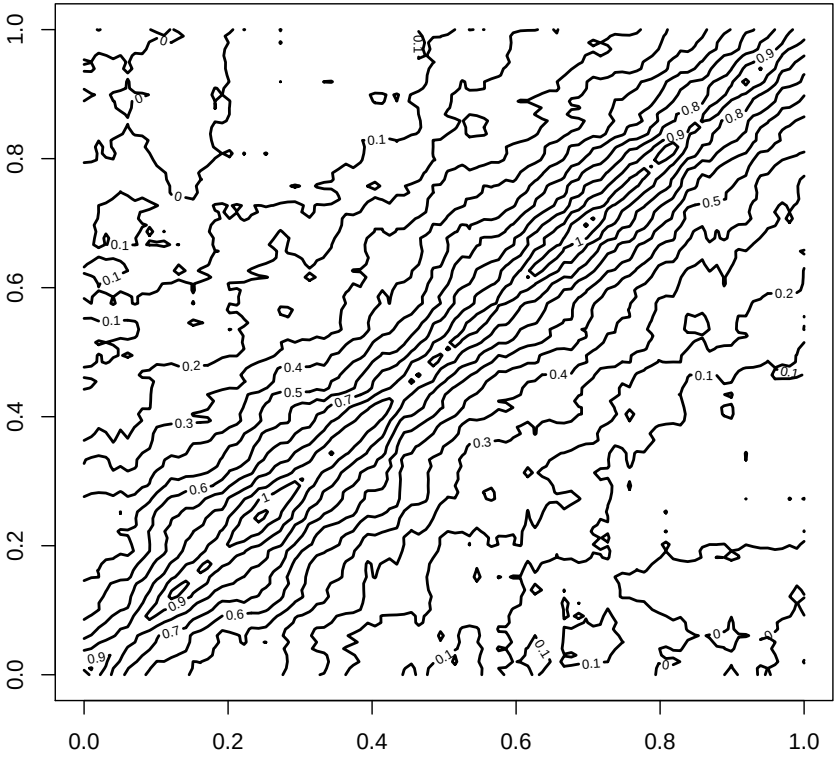


# Ejemplos: $\widehat{\Gamma}$

Ornstein-Uhlenbeck




$Y(t)$



# Estimación

- Podemos estimar las autofunciones de $\Gamma$ 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

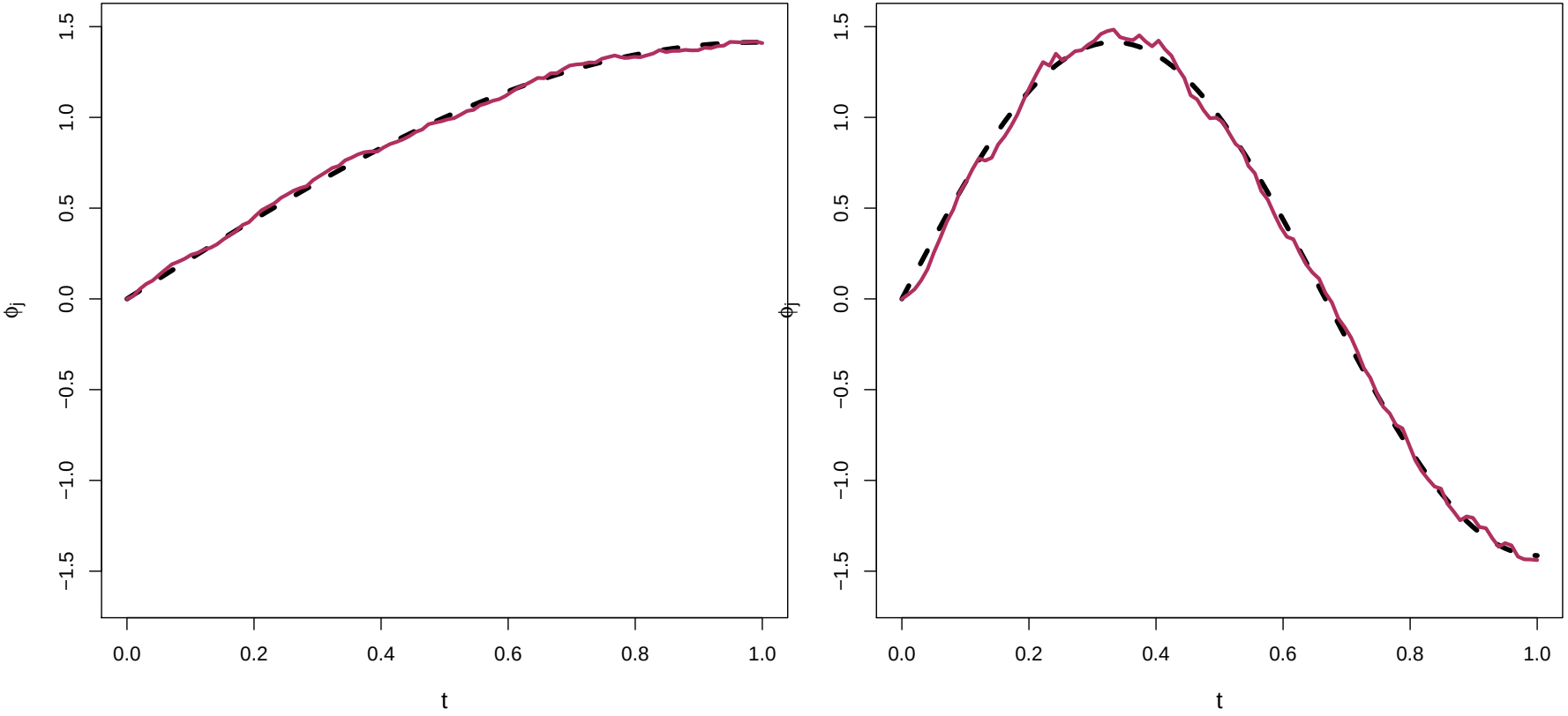
$$\gamma(s,t) = \min(s,t)$$

entonces

$$\lambda_j = \pi^{-2}(j-0.5)^{-2}$$
$$\phi_j(t) = \sqrt{2}\sin\left(\left(j-\frac{1}{2}\right)\pi\, t\right)$$

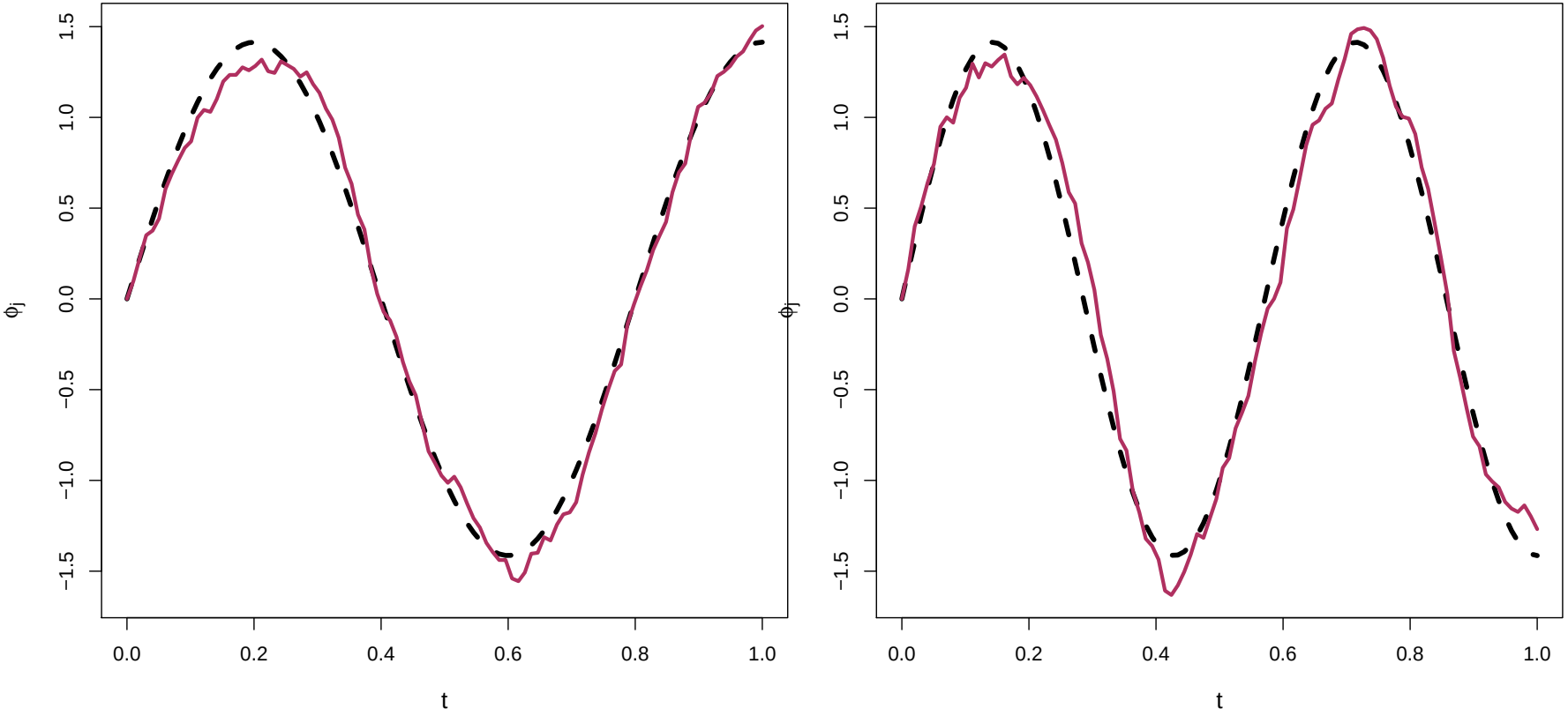
# Ejemplos: $\widehat{\phi}$ para $W(t)$

$j = 1$ $j = 2$



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

$j = 3$ $j = 4$



# Estimación

- Podemos estimar las autofunciones de $\Gamma$ usando las de $\widehat{\Gamma}$ asociadas a los autovalores no nulos.

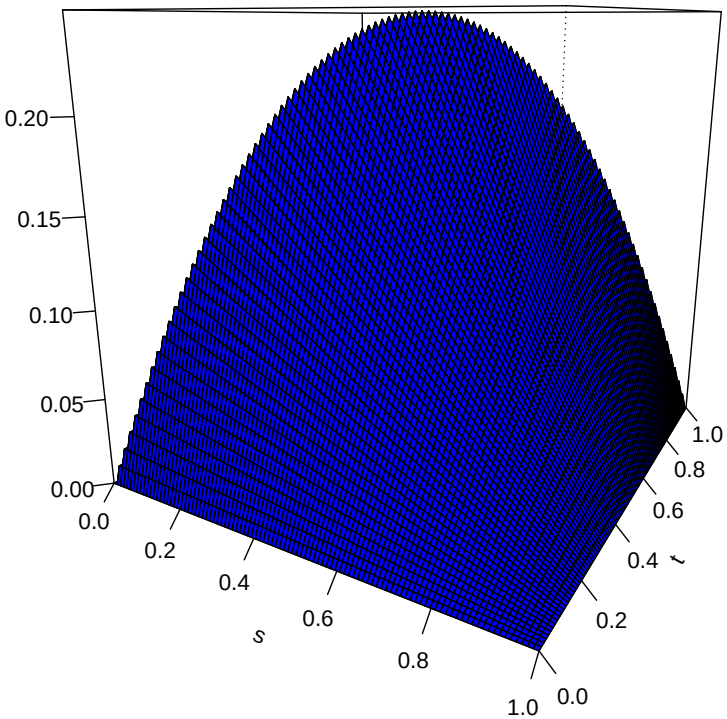
- Si $\mathcal{H} = L^2(\mathcal{I})$, y tenemos un puente Browniano $X(t)$ con

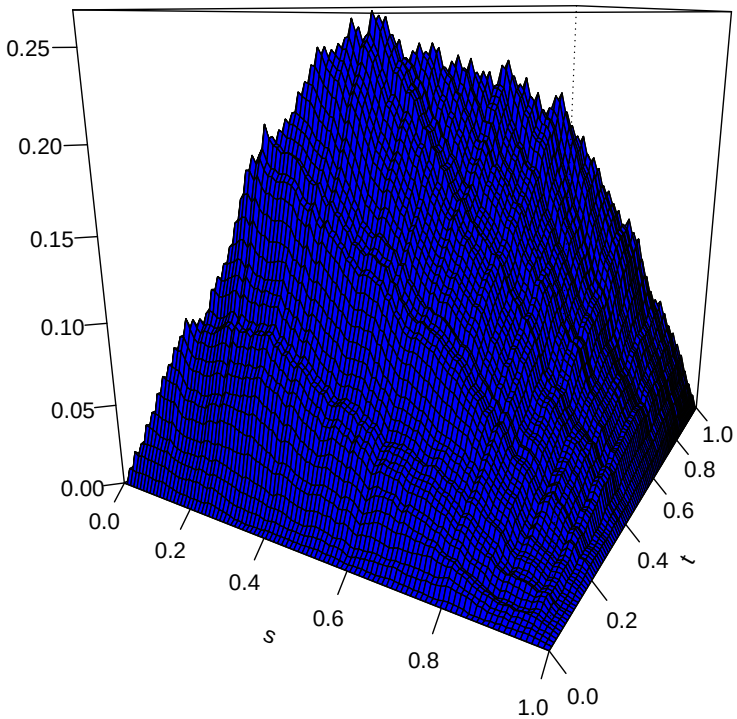
$$\gamma(s,t) = \min(s,t) - t\,s$$

entonces

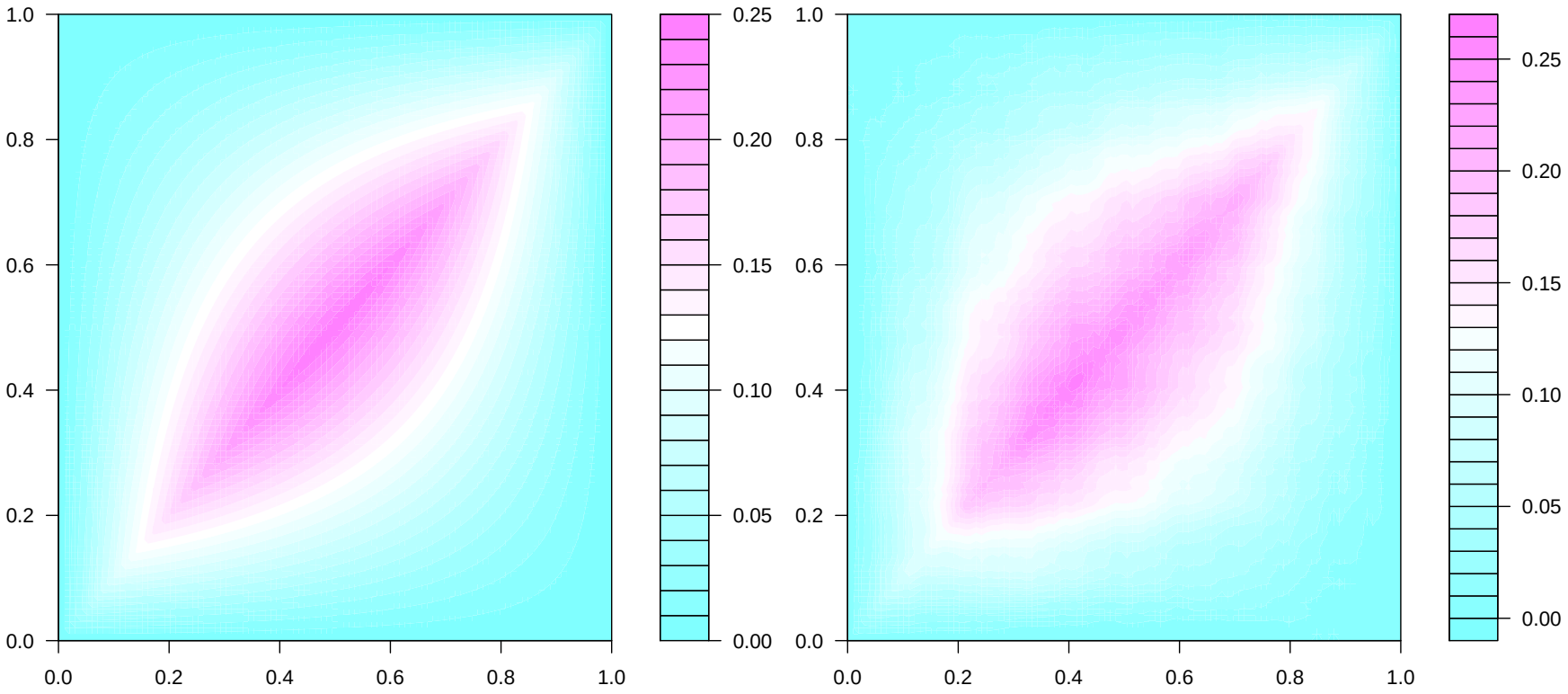
$$\lambda_j = \pi^{-2} j^{-2}$$
$$\phi_j(t) = \sqrt{2}\sin(j\,\pi\,t)$$

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



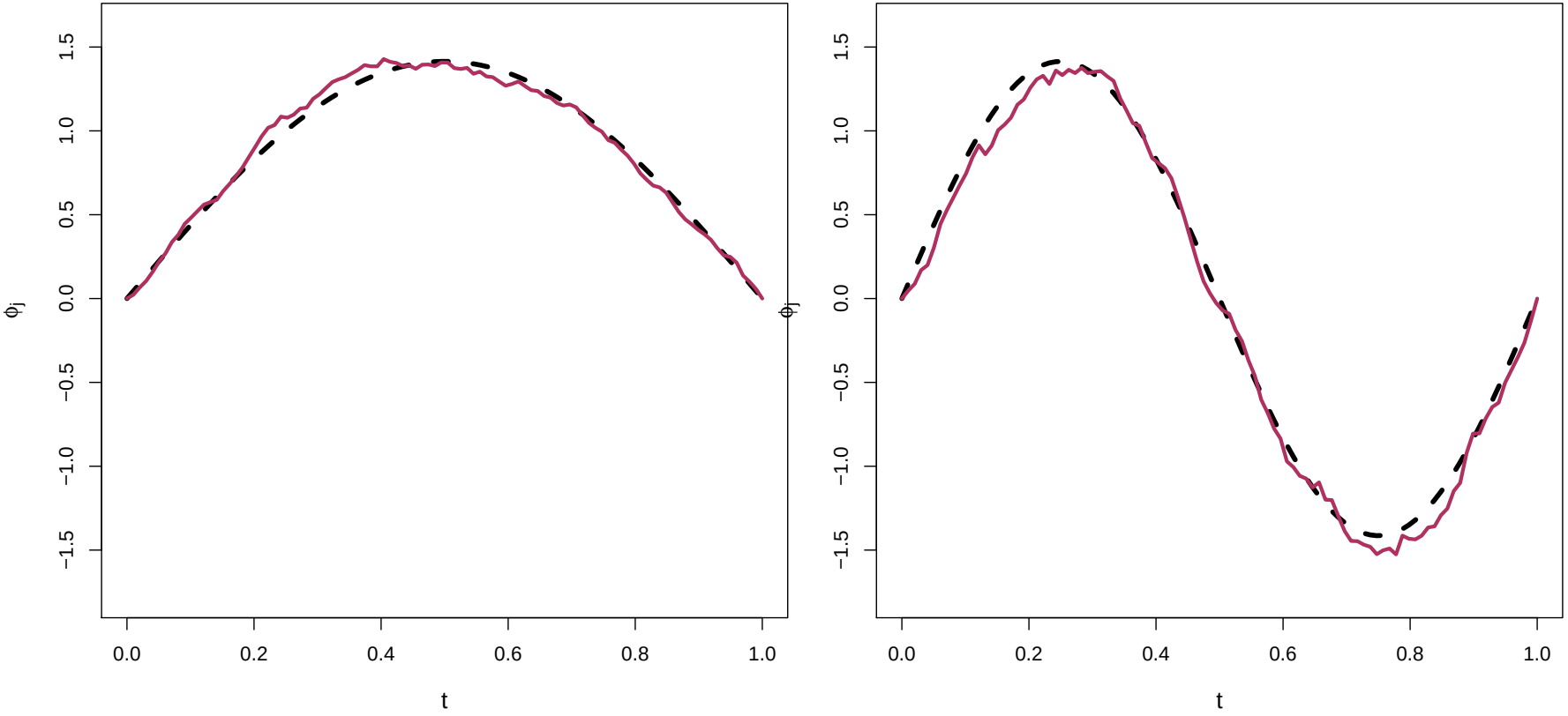



# Ejemplos: $\widehat{\Gamma}$ para $X(t)$



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

$j = 1$ $j = 2$



# Estimadores suaves

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

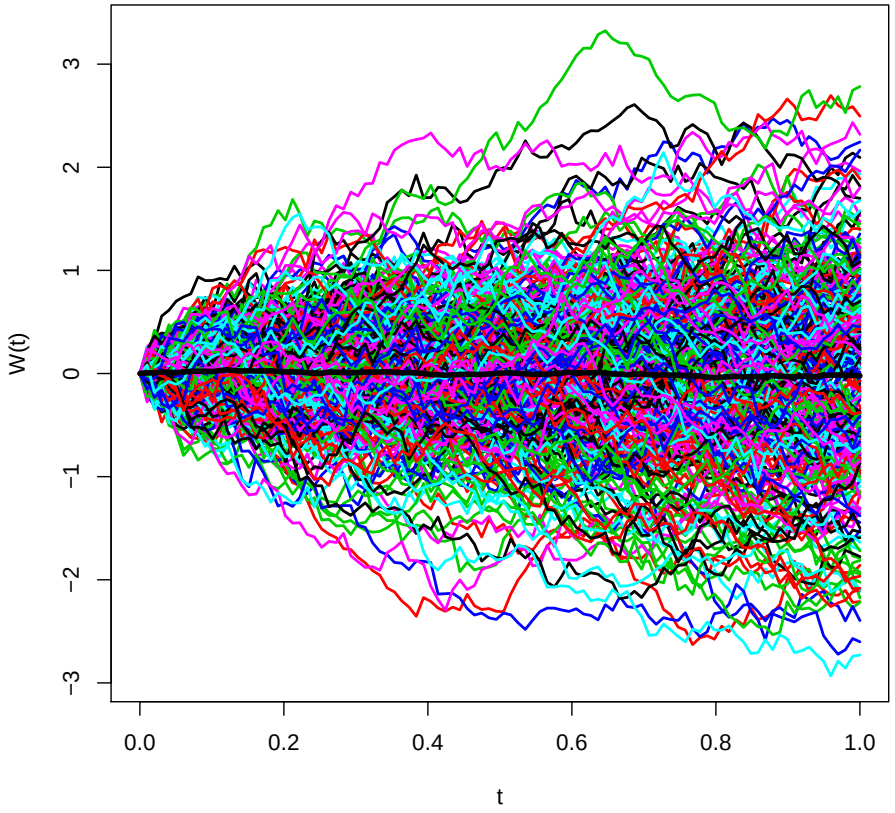
- $X_i(t) \approx \widehat{X}_i(t) = \sum_{j=1}^{p} a_{ij}\phi_j(t)$
- Si la trayectoria $X_i(t)$ se observa en los puntos $t_1 < \ldots t_m$, los coeficientes $a_{ij}$ se pueden elegir de modo a minimizar

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

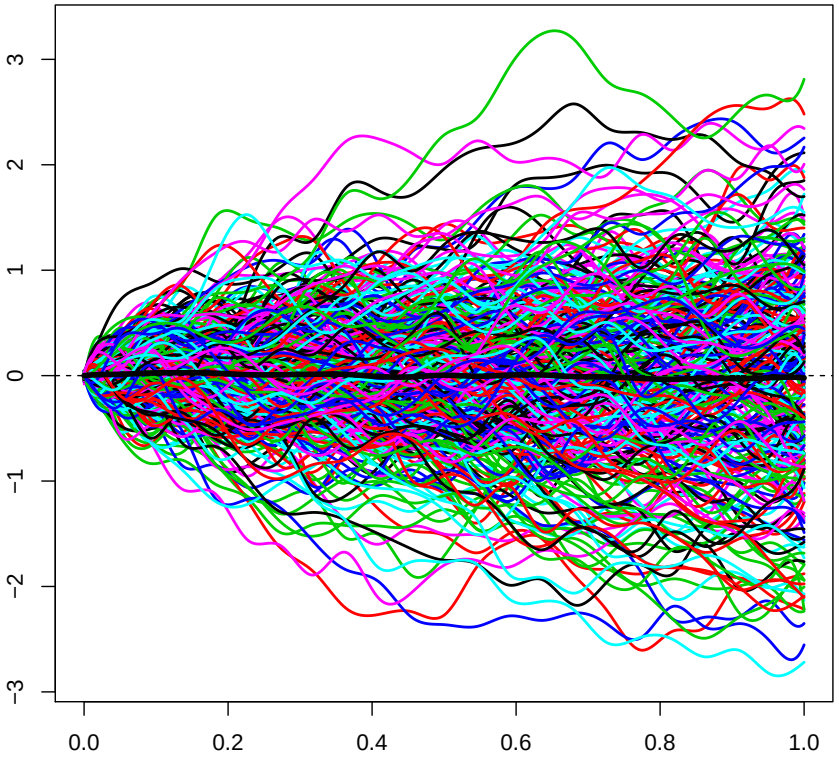
  ver detalles en Ramsay y Silverman (2005), Sección 3.2
- Se calcula luego el operador de covarianza muestral de $\widehat{X}_i(t)$ y sus autofunciones que dan estimadores de las autofunciones de $\Gamma$.

# Ejemplo: Wiener

Wiener $W_i(t)$

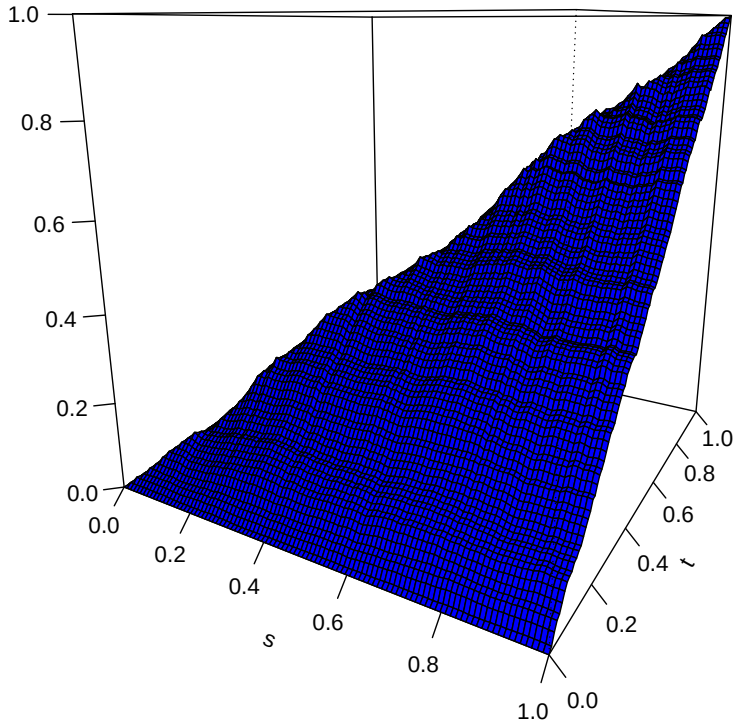



$\widehat{W}_i(t)$

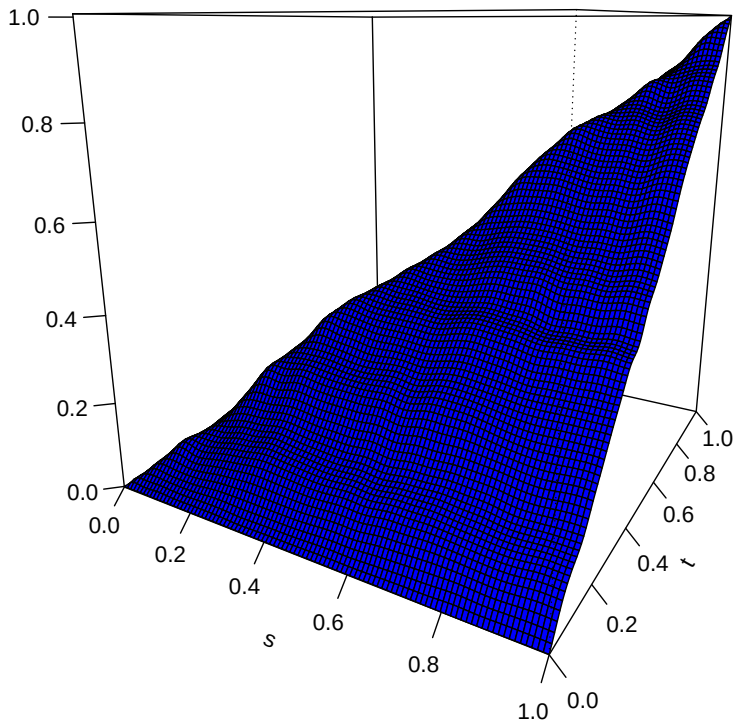


# Ejemplos: $\widehat{\Gamma}$

$\widehat{\Gamma}_W$

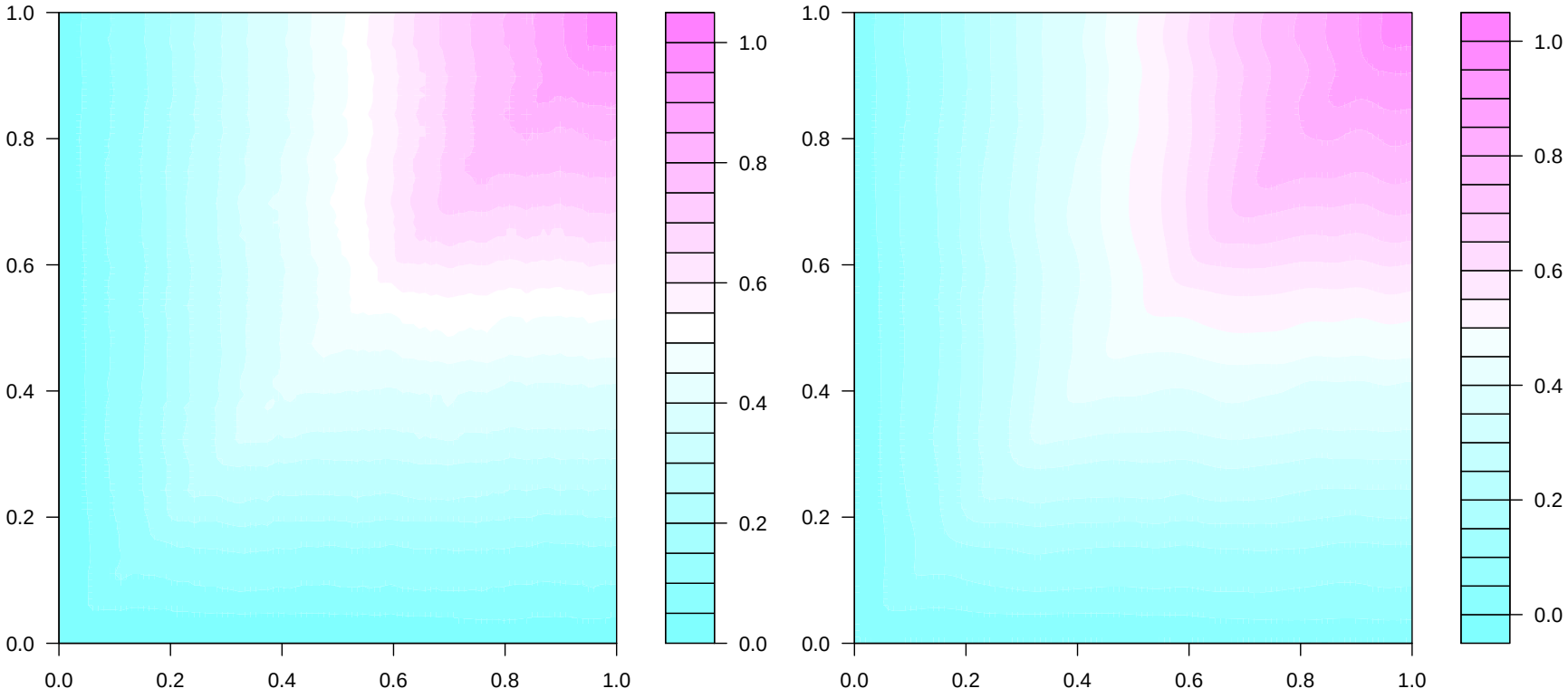



$\widehat{\Gamma}_{\widehat{W}}$



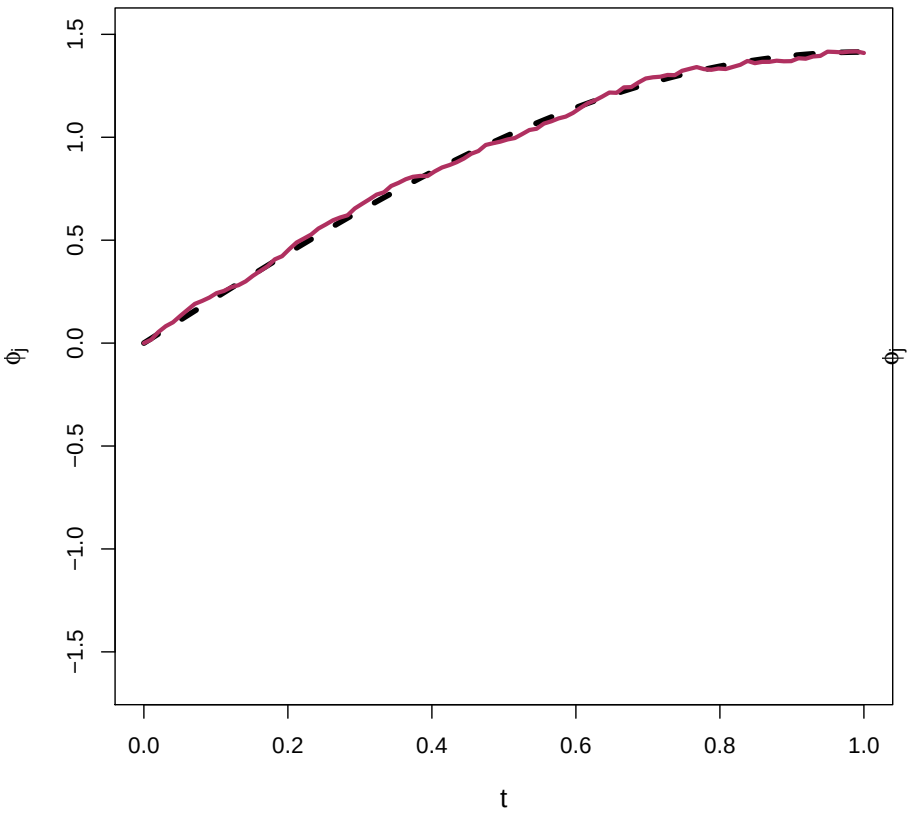
# Ejemplos: $\widehat{\Gamma}$

$\widehat{\Gamma}_W$ $\widehat{\Gamma}_{\widehat{W}}$

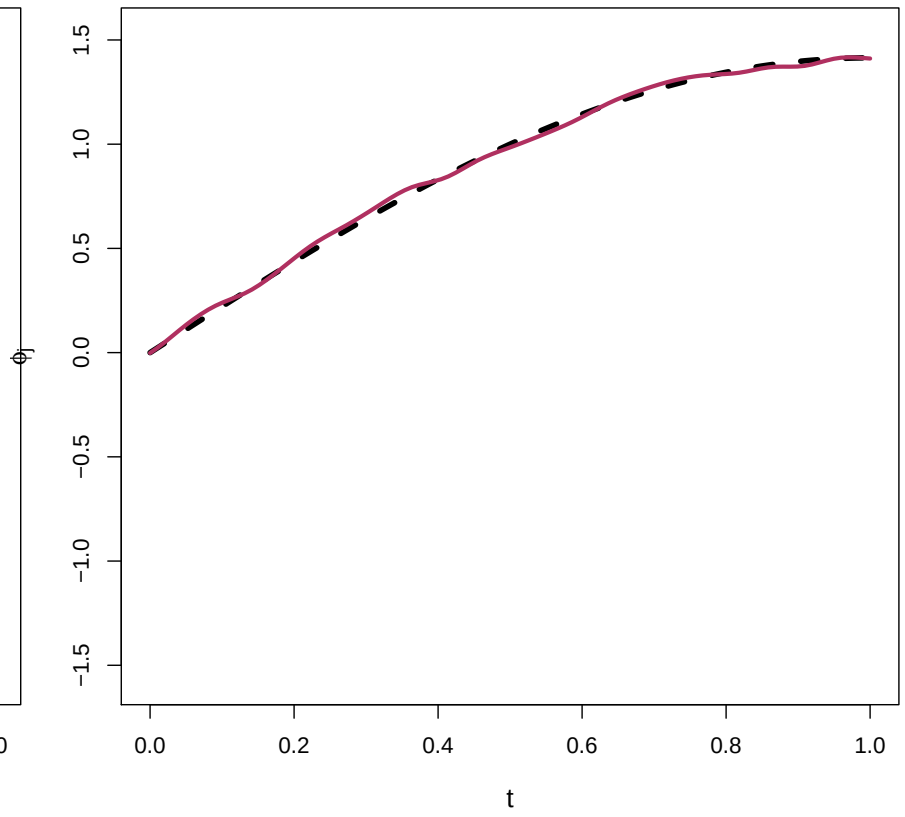


# $\widehat{\phi}_j$ para $W(t)$ y $\widehat{W}(t)$: $j = 1$

$\widehat{\phi}_{1,W}$

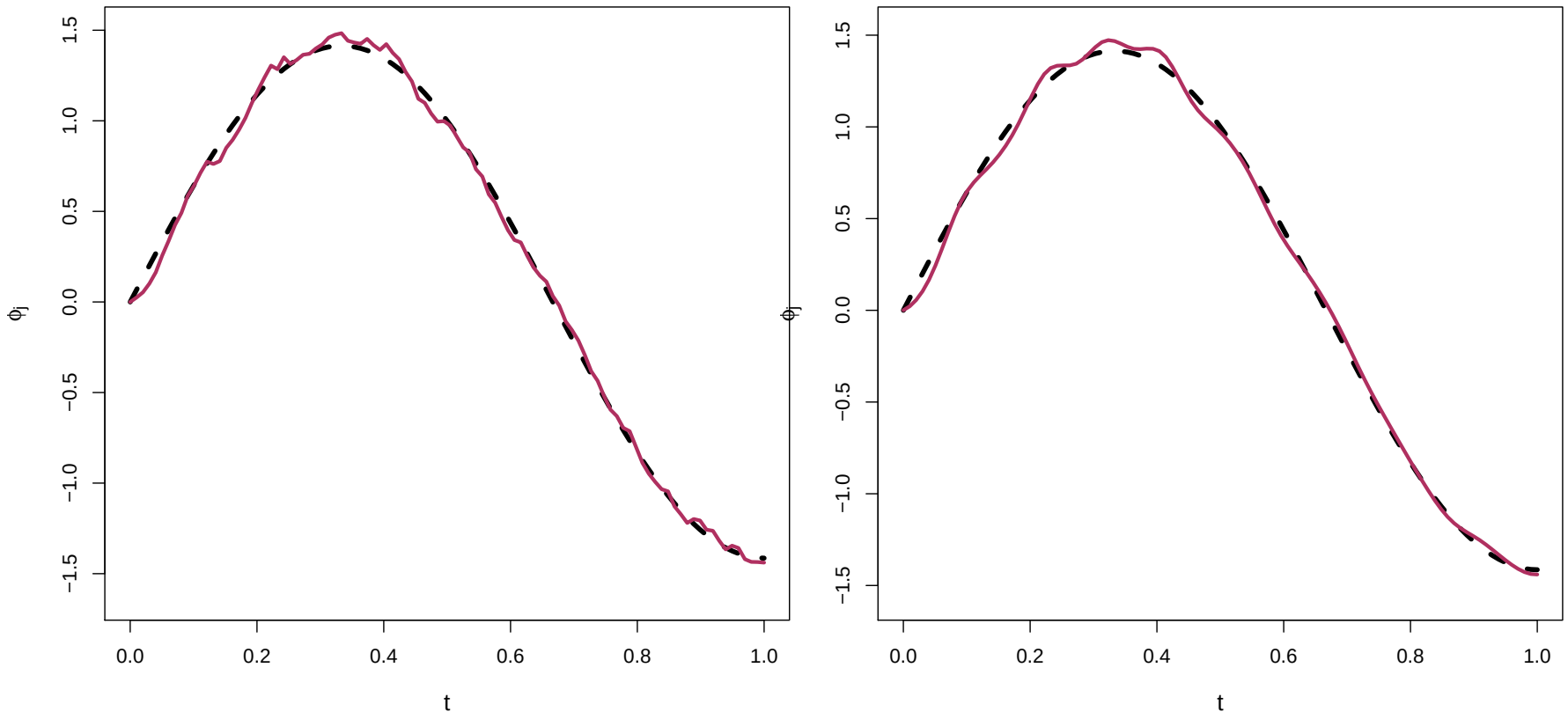



$\widehat{\phi}_{1,\widehat{W}}$



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

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



# 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 $t_1 < \ldots t_m$,

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

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

# 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 $t_1 < \ldots t_m$,

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

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

# Estimadores suaves

- El estimador del operador de covarianza se define como

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

donde

$$\widehat{\gamma}_h(t,s) = \frac{1}{n}\sum_{i=1}^{n}\left(X_{i,h}(t) - \overline{X}_h(t)\right)\left(X_{i,h}(s) - \overline{X}_h(s)\right) .$$

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

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

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

# 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 $m \geq 2$,

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

entonces, si $h \to 0$,

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

## 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

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

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

  b) En particular, si $m_j = 1$,

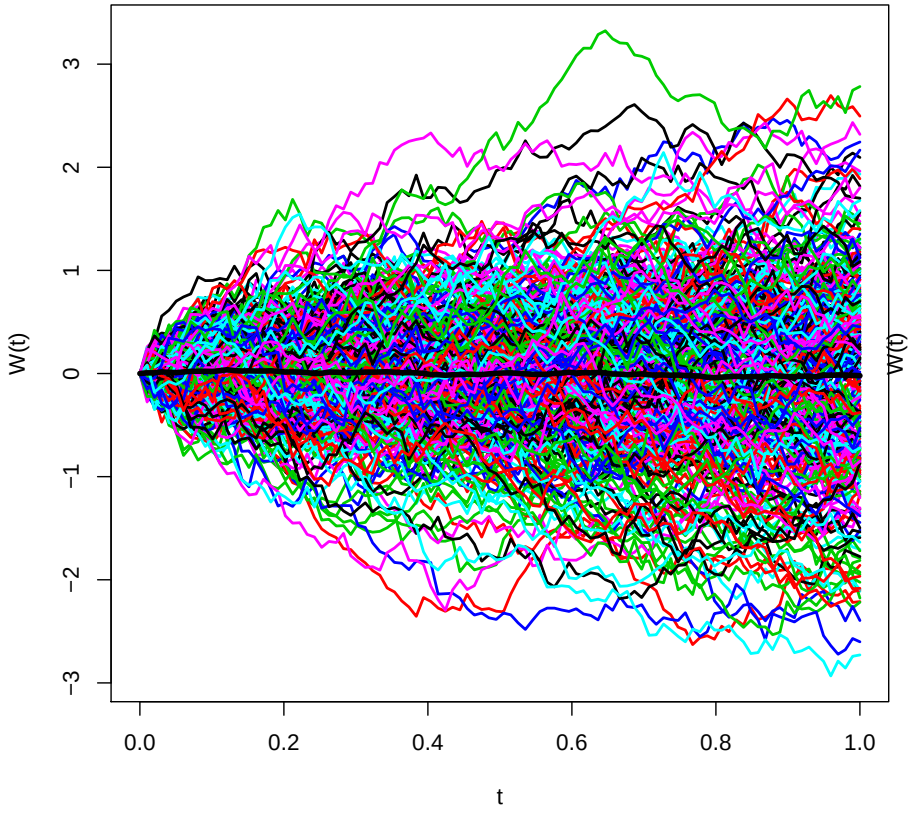
$$\widehat{\lambda}_{j,h} \xrightarrow{a.s.} \lambda_j$$

  y

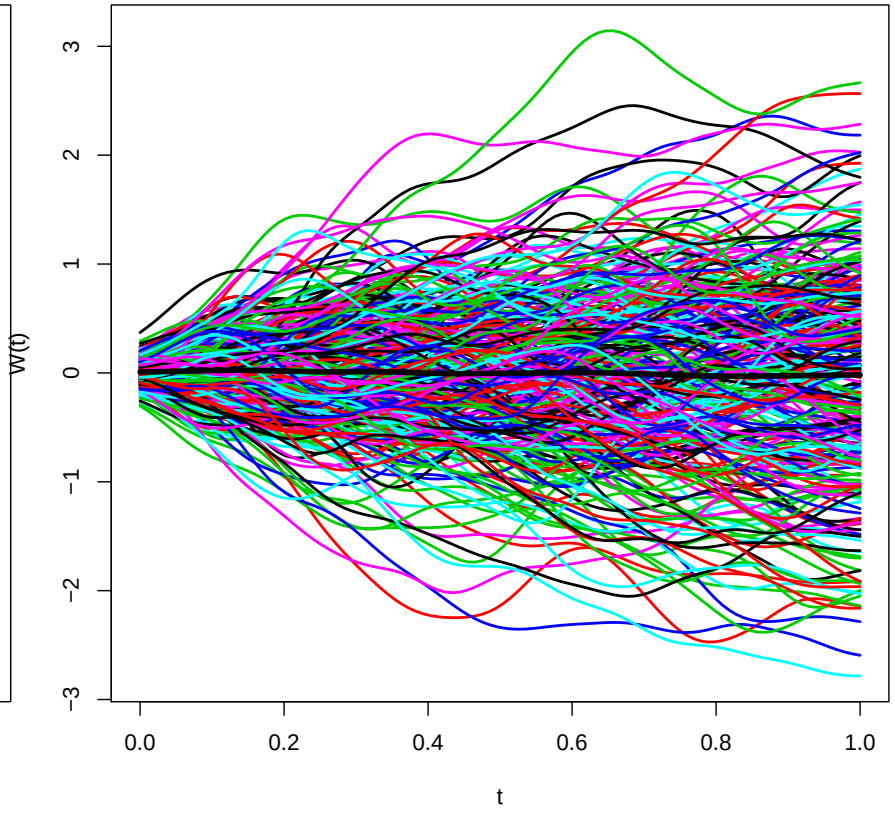
$$\text{signo}\left(\langle \widehat{\phi}_{j,h}, \phi_j \rangle\right) \widehat{\phi}_{j,h} \xrightarrow{a.s.} \phi_j \qquad \text{en } L^2[0,1]$$

# Ejemplo: Wiener

Wiener $W_i(t)$

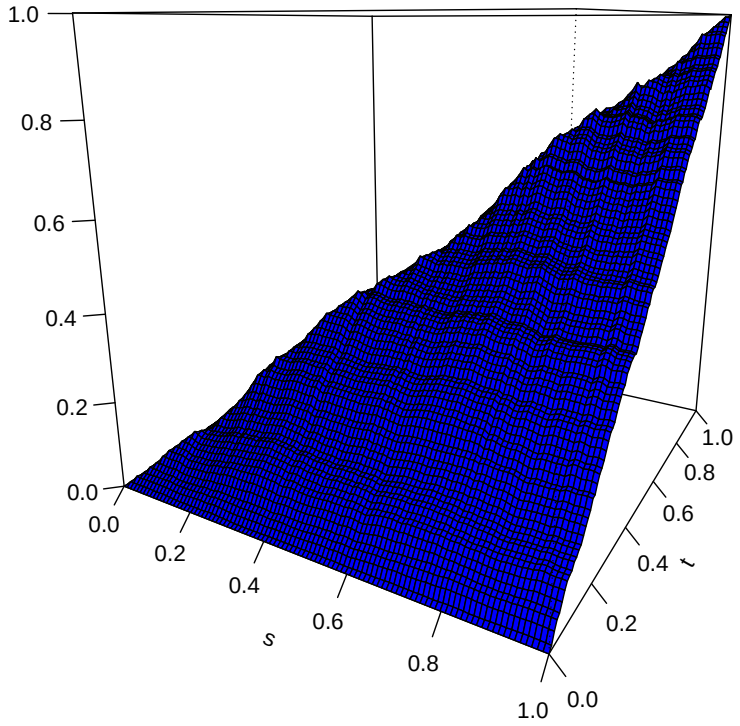



$\widehat{W}_{i,h}(t)$

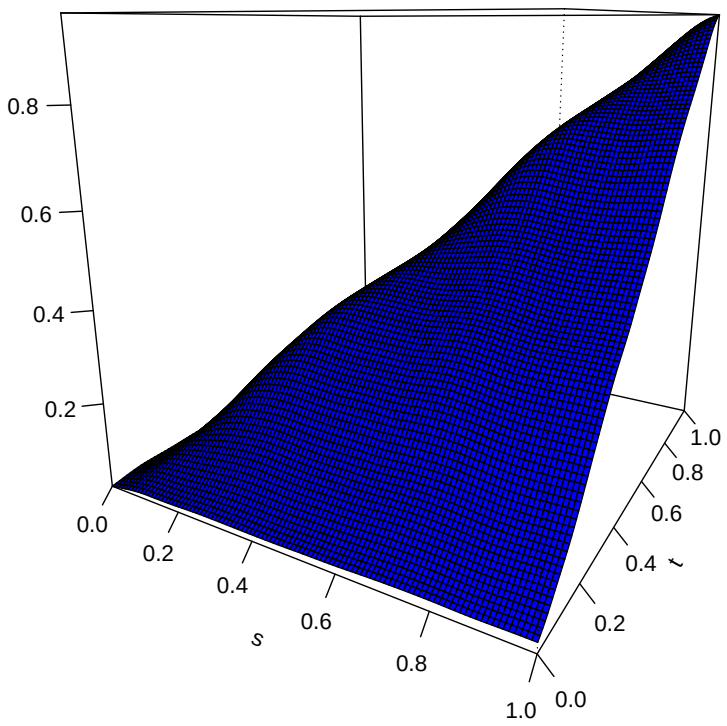


# Wiener: $\widehat{\Gamma}$ y $\widehat{\Gamma}_h$

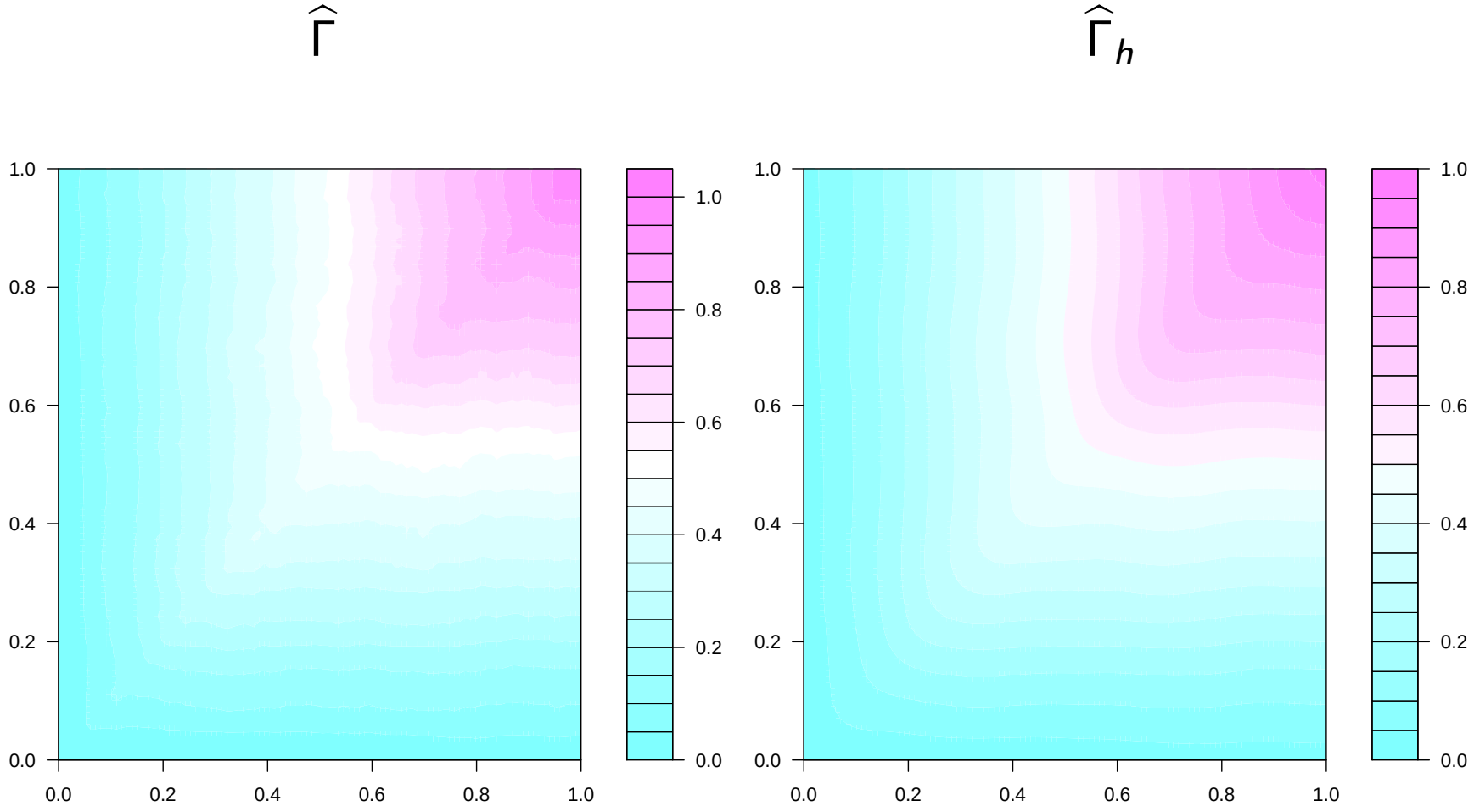
$\widehat{\Gamma}$




$\widehat{\Gamma}_h$

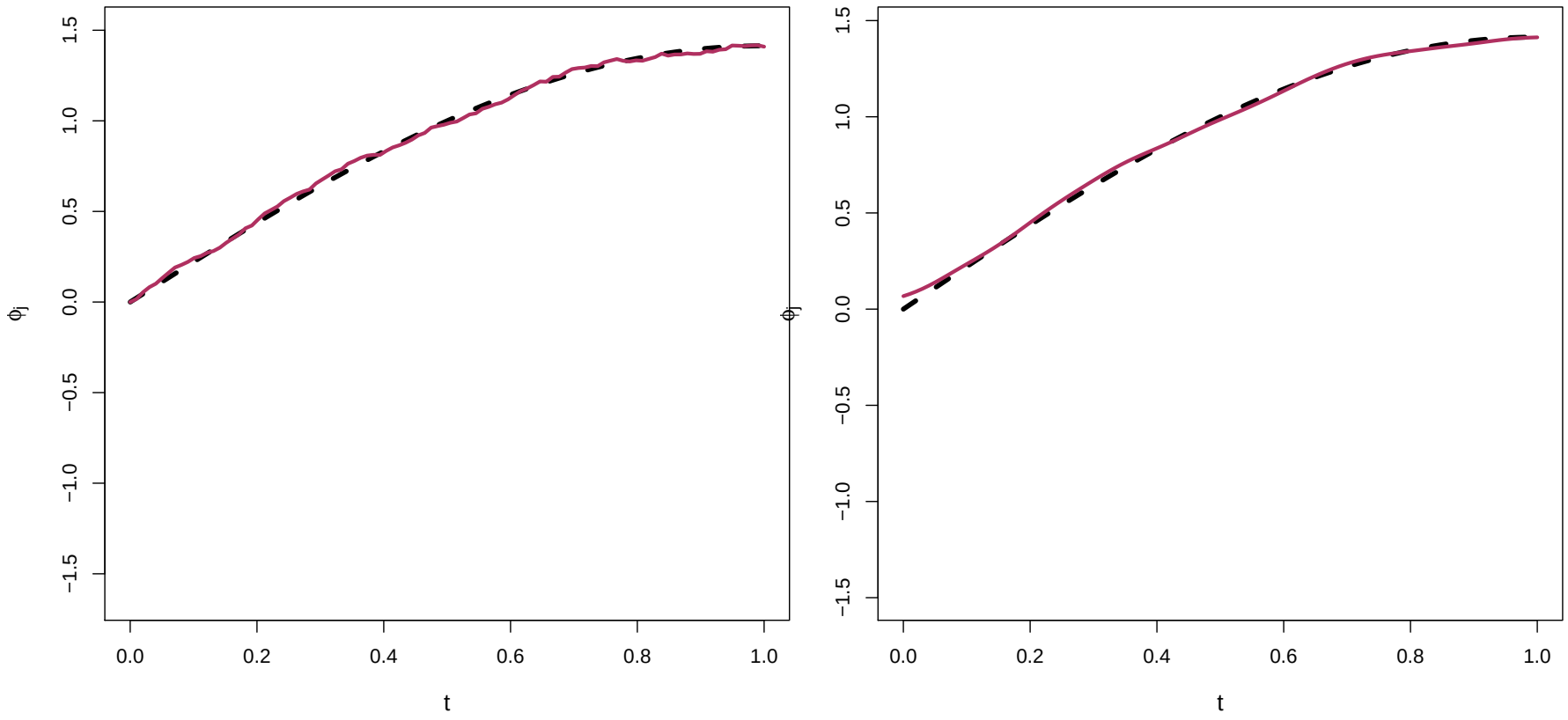


# $\widehat{\Gamma}$ y $\widehat{\Gamma}_h$



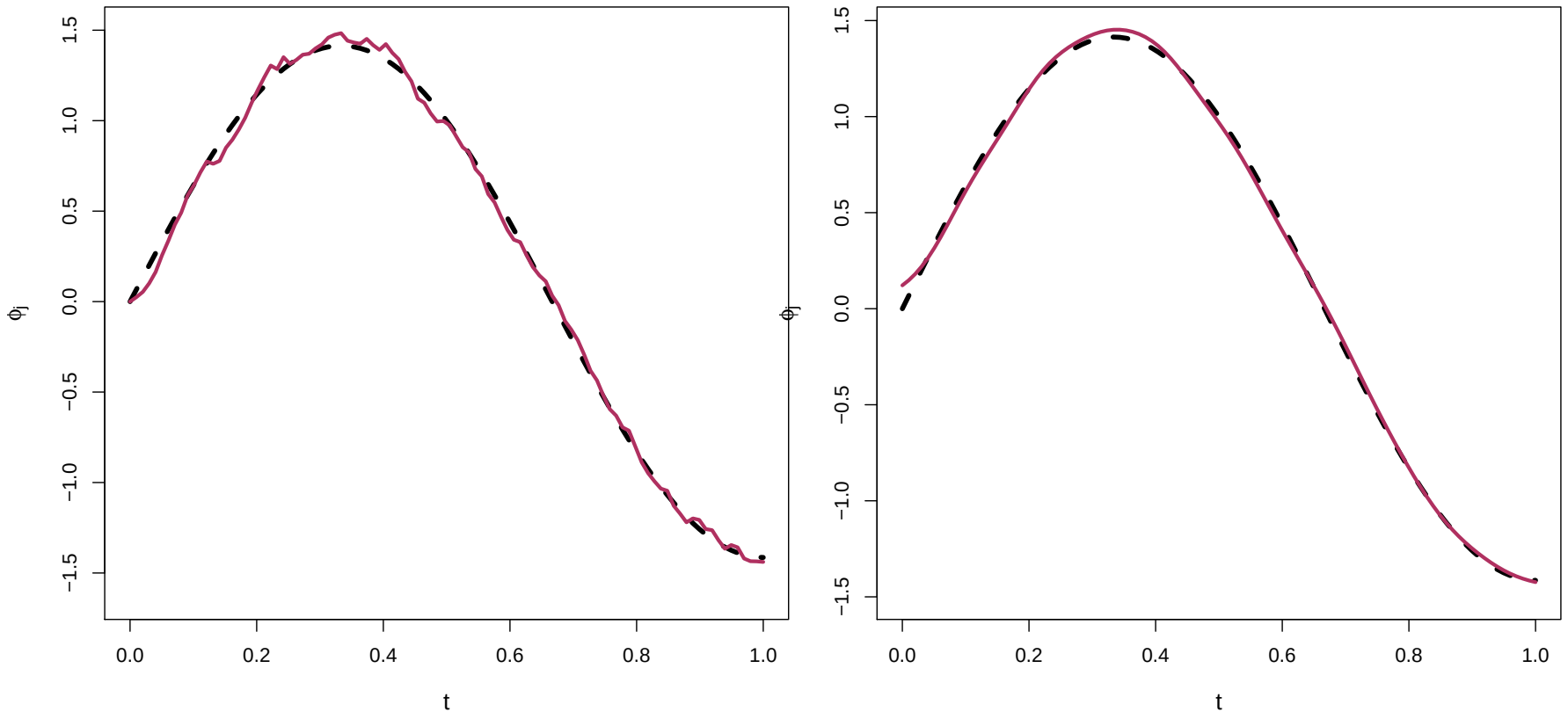
# $\widehat{\phi}_j$ y $\widehat{\phi}_{j,h}$: $j = 1$

$\widehat{\phi}_1$ $\widehat{\phi}_{1,h}$



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

$\widehat{\phi}_2$ $\widehat{\phi}_{2,h}$



# Mediana

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

$$\mu_{\mathrm{GM}} = \underset{u \in \mathcal{H}}{\operatorname{argmin}} \, \mathbb{E}\left(\|X - u\| - \|X\|\right) ,$$

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

$$\widehat{\mu}_{\mathrm{GM}} = \underset{u \in \mathcal{H}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(\|X_i - u\| - \|X_i\|\right) ,$$

# Mediana

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

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

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

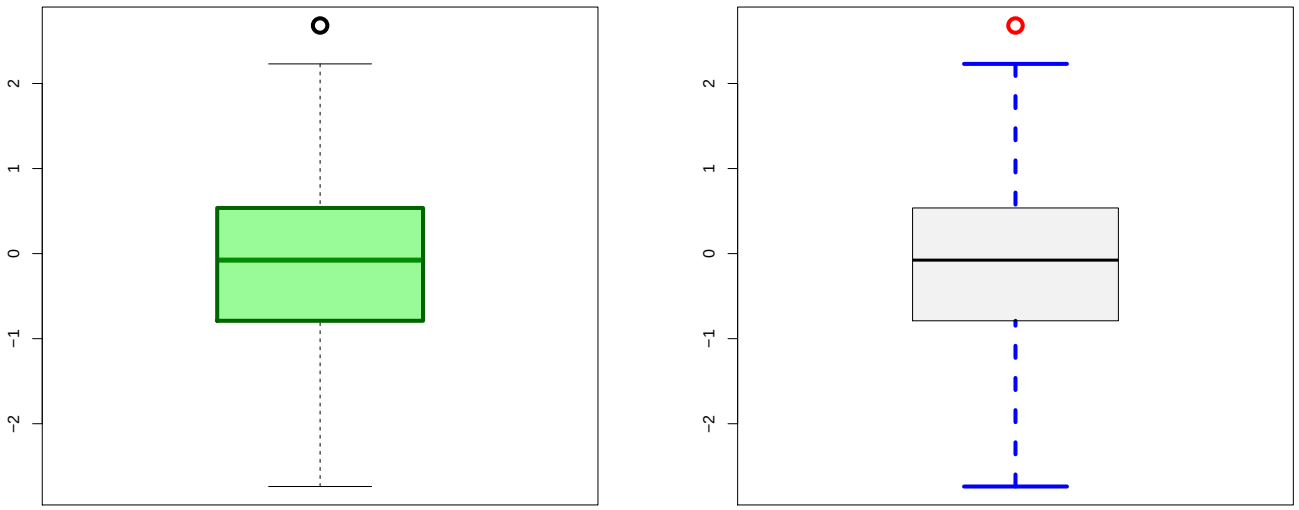
$$\|\widehat{\mu}_{\mathrm{GM}} - \mu_{\mathrm{GM}}\| \xrightarrow{a.s.} 0$$

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.

# Boxplot

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

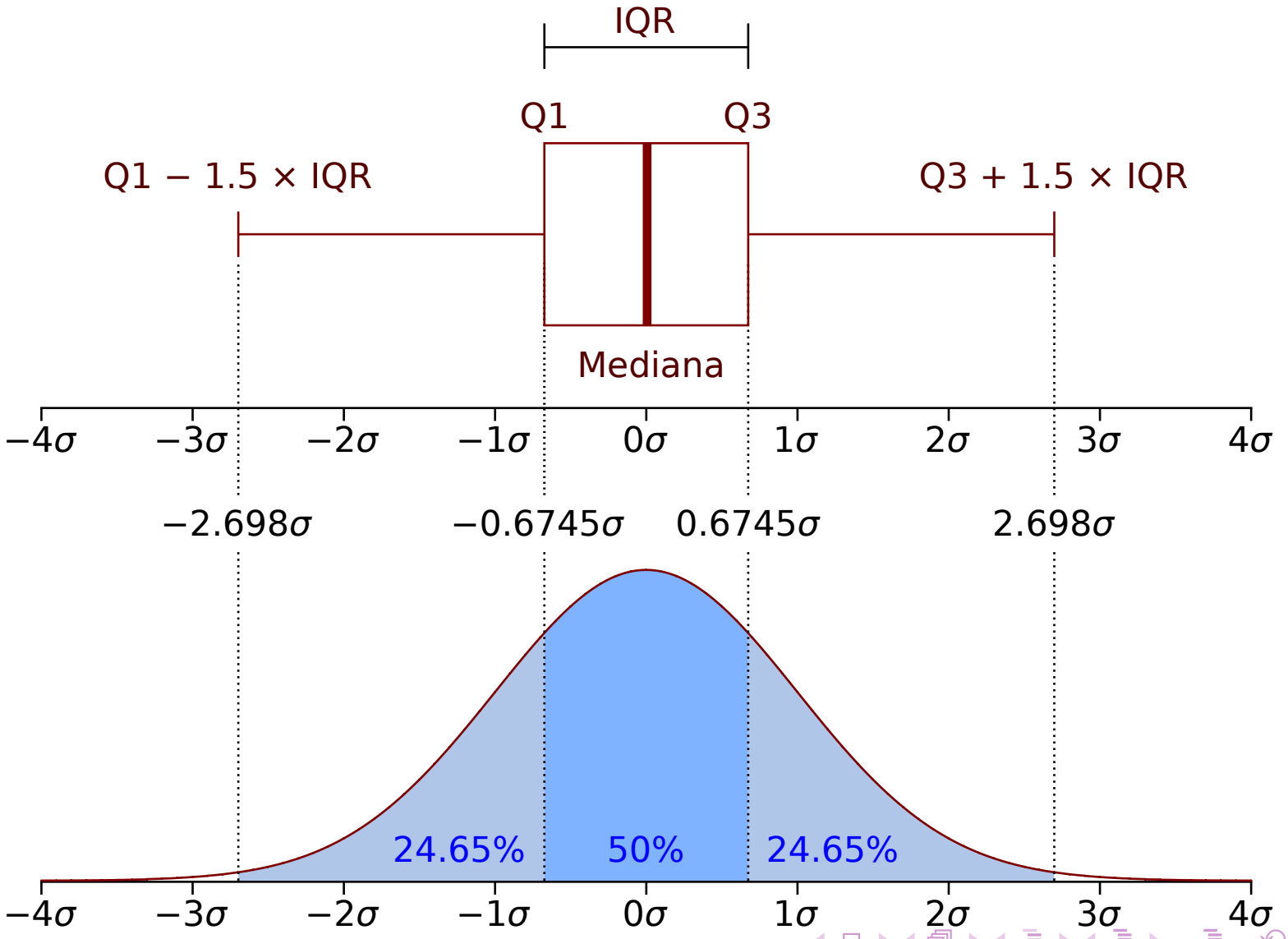



(a) Mediana y caja. (b) Bigotes y outliers

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

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

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



# 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 $X_1, \ldots, X_n \in \mathbb{R}$

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

- Si $x$ es la mediana, entonces
$$F_n(x) = \frac{1}{2} \qquad hsprof(x) = \frac{1}{2}$$
que es la mayor profundidad posible.
- Si $x = X^{(n)} = \max_i X_i$, entonces
$$F_n(x) = 1 \qquad hsprof(x) = 0$$
que es la menor profundidad posible.

# Profundidad

Otras medidas posibles de profundidad son la profundidad simplicial

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

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

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

$\mathcal{S}_{\mathbf{x}}$ es un semi–espacio cerrado cuya frontera pasa por $\mathbf{x}$.

- 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 poligonos convexos $\mathcal{D}_{k+1} \subset \mathcal{D}_k$

# Mediana: Donoho y Gasko (1992)

En el caso multivariado, la noción de mediana puede generalizarse

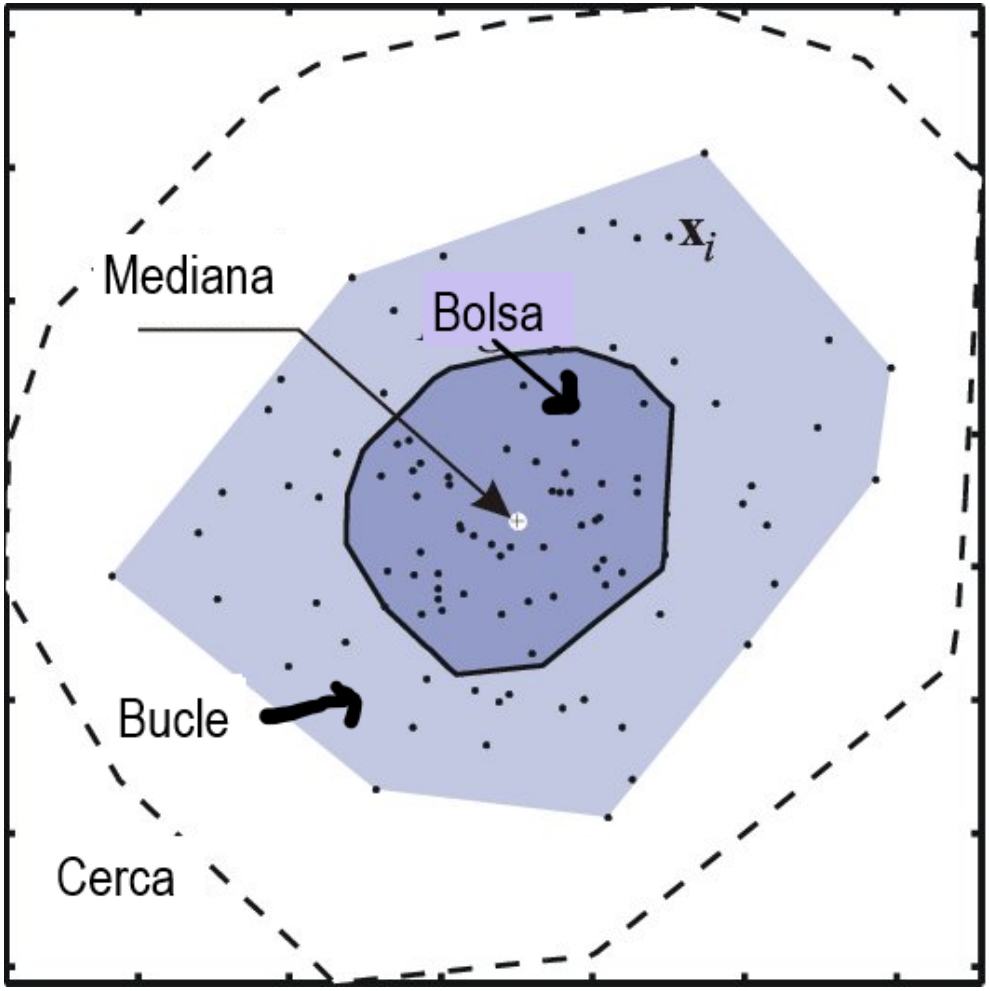
- La mediana de Tukey $\boldsymbol{\theta}^\star$ se define como el valor con máxima profundidad si es único

  En caso contrario, $\boldsymbol{\theta}^\star$ se define como el centro de gravedad de la región con mayor profundidad.

# Bagplot: Rousseeuw, Ruts and Tukey (1999)

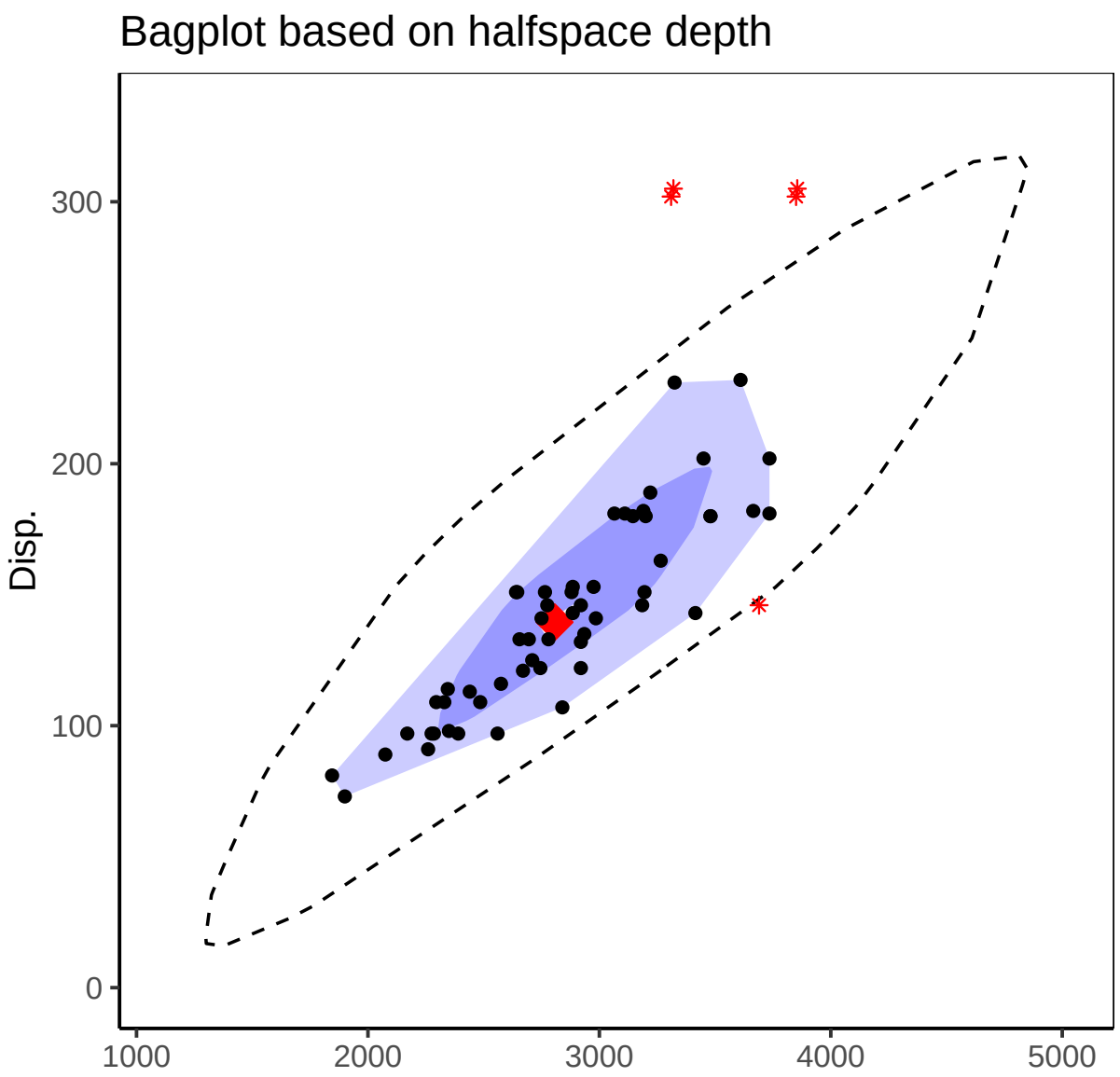
El bagplot es un gráfico bidimensional.

- 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.

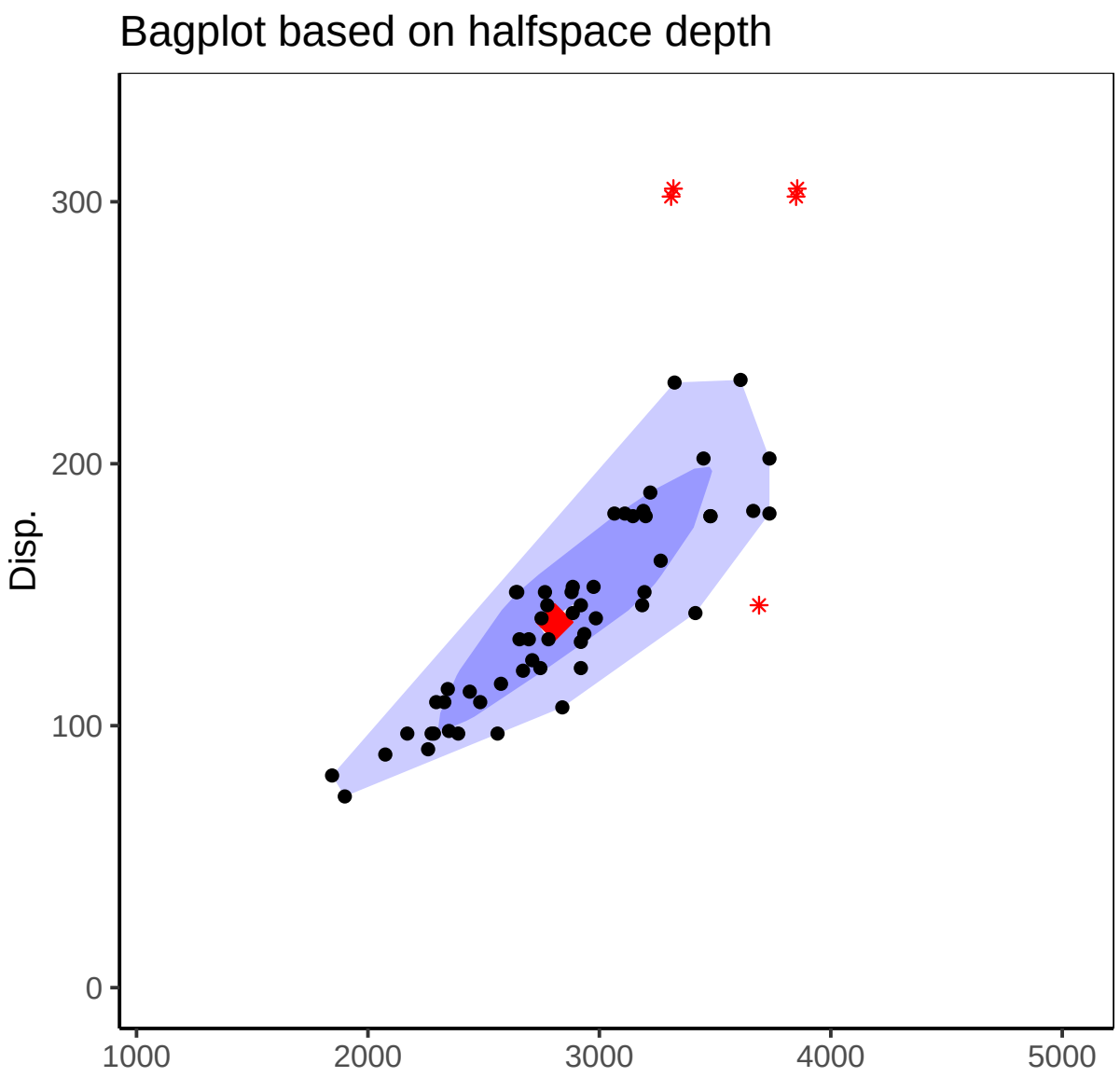

Mediana
Bolsa
$\mathbf{x}_i$
Bucle
Cerca

```
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(bcar, 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)
```

# Bagplot del Car Data



# Bagplot del Car Data



# 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. & Gonzáalez-Manteiga, W. (2008). Outlier detection in functional data by depth measures with application to identify abnormal NOx 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 nocion de profundidad funcional y estudian las propiedades de las ya deinidas**

# 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.

# Profundidad Integrada de Fraiman & Muñiz (2001)

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

Definamos la distribución empírica marginal como

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

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

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

$$D_n(x(t))$$

# 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.

# 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,

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

# 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

$$\|X_j - X_k\|$$

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

# 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.**

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

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

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

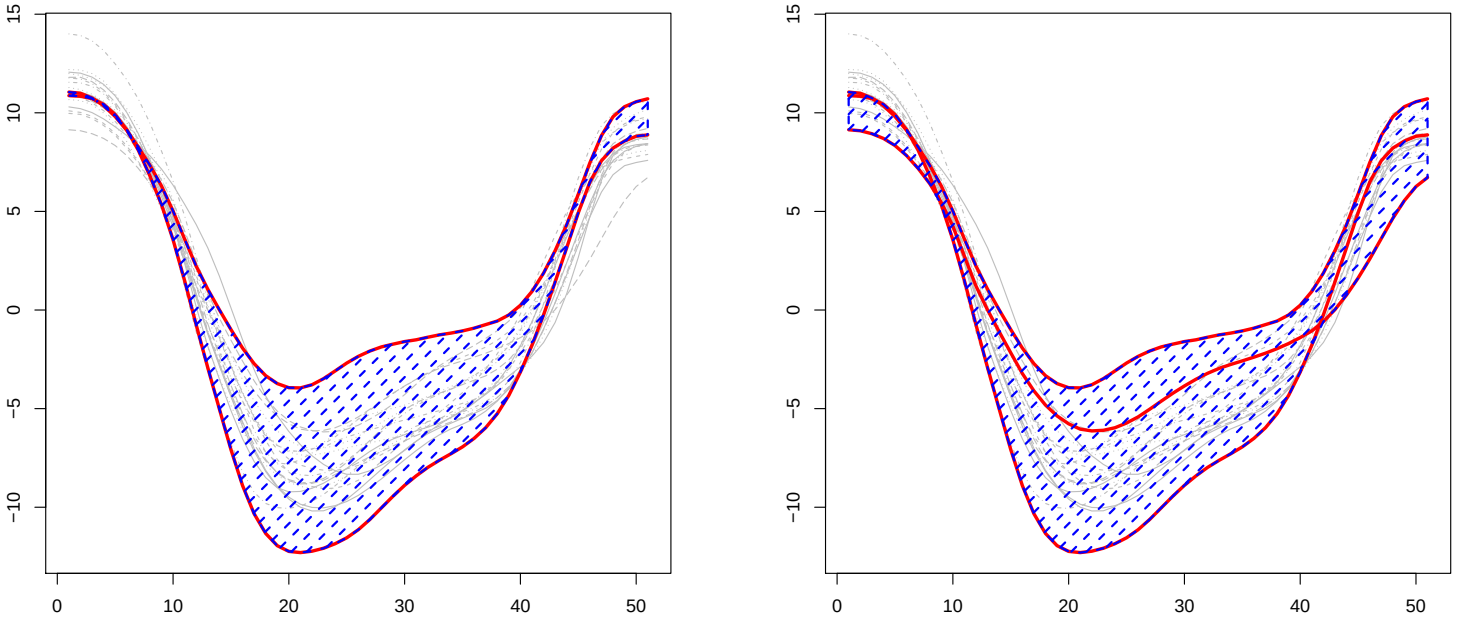
- La banda en $\mathbb{R}^2$ determinada por $k$ de curvas, $x_1, \ldots, x_k$, es

$$V(x_1, x_2, \ldots, x_k) = \{(t,y) : t \in \mathcal{I}, \min_{r=1,\ldots,k} x_r(t) \leq y \leq \max_{r=1,\ldots,k} x_r(t)\}$$

$$= \{(t,y) : t \in \mathcal{I}, y = \alpha_t \min_{r=1,\ldots,k} x_r(t) + (1-\alpha_t) \max_{r=1,\ldots,k} x_r(t),$$

$$\alpha_t \in [0,1]\}$$

# 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

# 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 $x \in \{X_1, \ldots, X_n\}$ y $j \geq 2$

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

- $BD_n^{(j)}(x)$ 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

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

# Profundidad de bandas: López Pintado & Romo (2009)

- Se puede ver que

$$BD_n^{(j)}(x) = \binom{n}{k}^{-1} \sum_{1 \le i_1 \le i_2 \le \cdots \le i_j \le n} \mathbb{I}\left\{x \in \mathcal{B}(X_{i_1}, X_{i_2}, \ldots, X_{i_j})\right\}$$

donde

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

$$\sum_{\ell=1}^{j} \alpha_\ell = 1 \qquad \alpha_\ell \ge 0\}$$

# Profundidad de bandas: Versión poblacional

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

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

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

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, $J = 3$.

# Mediana

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

$$\widehat{m}_{n,J} = \underset{x \in \{X_1,\ldots,X_n\}}{\operatorname{argmax}} \; BD_{n,J}(x) \; .$$

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

## 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 $j \geq 2$

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

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

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

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

es la "proporción de tiempo" que $x$ está en la banda.

# Profundidad de bandas modificada (MBD)

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

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

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

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

# Profundidad de banda modificada: Versión poblacional

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

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

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

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

# 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.

# *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.

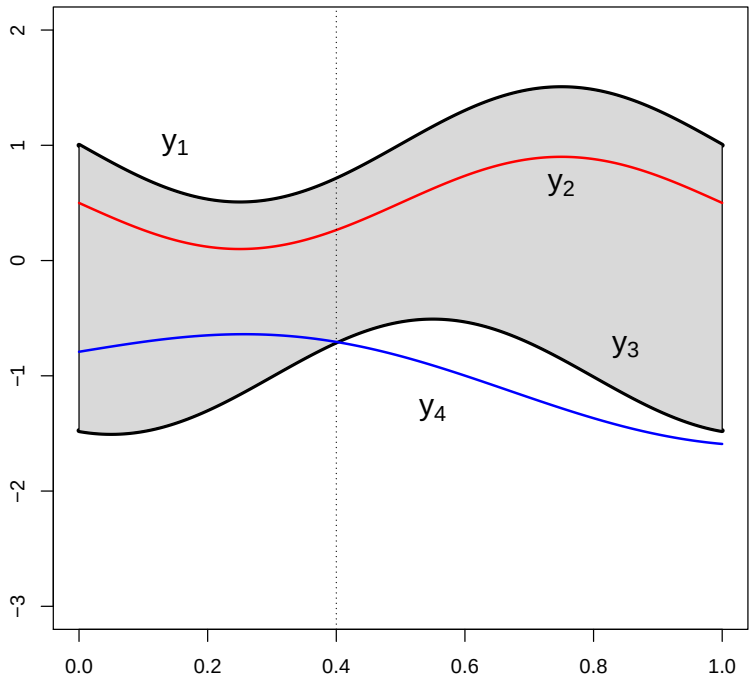
- La profundidad de banda depende más de la forma de las curvas dando más empates.

# *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.
- $X_i : \Omega \to L^2(\mathcal{I})$ o $\mathcal{C}(\mathcal{I})$, $1 \leq i \geq n$, $\mathcal{I} = [0, 1]$.
- Sea $X_{[i]}(t)$ la observación asociada al $i$-ésimo mayor valor de profundidad.
- $X_{[1]}(t), \ldots, X_{[n]}(t)$ 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.

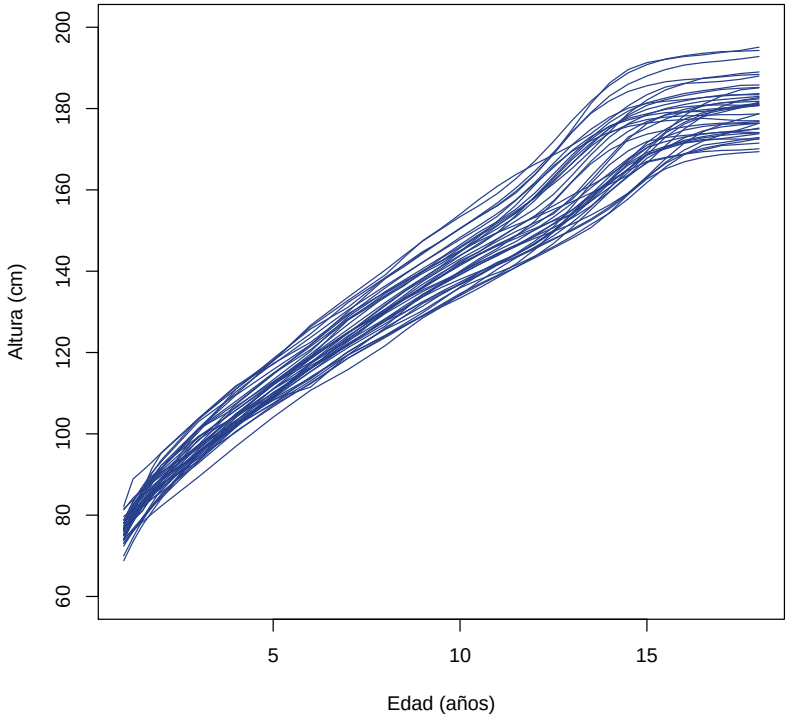
- La curva $X_{[1]}(t)$ es la mediana.

# Boxplot funcional

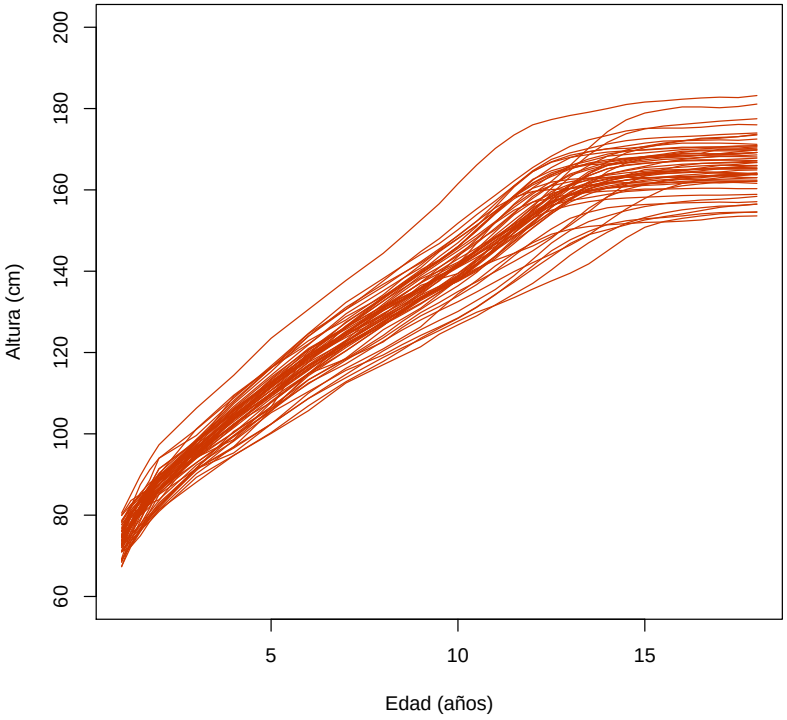
- Los bigotes del boxplot univariado son las líneas verticales del gráfico que se extienden desde la caja hasta la última observación que no es atípica.
- Definimos la *región exterior* inflando la región central 1.5 veces su tamaño.
- Cualquier curva fuera de estos límites se clasifica como un potencial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

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

# Altura de 39 niños y 54 niñas

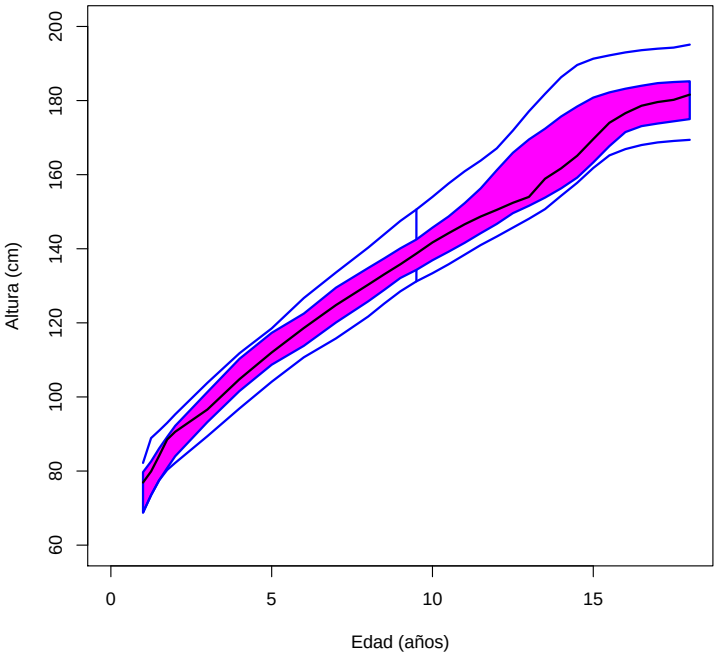



(c) Curvas para 39 varones.

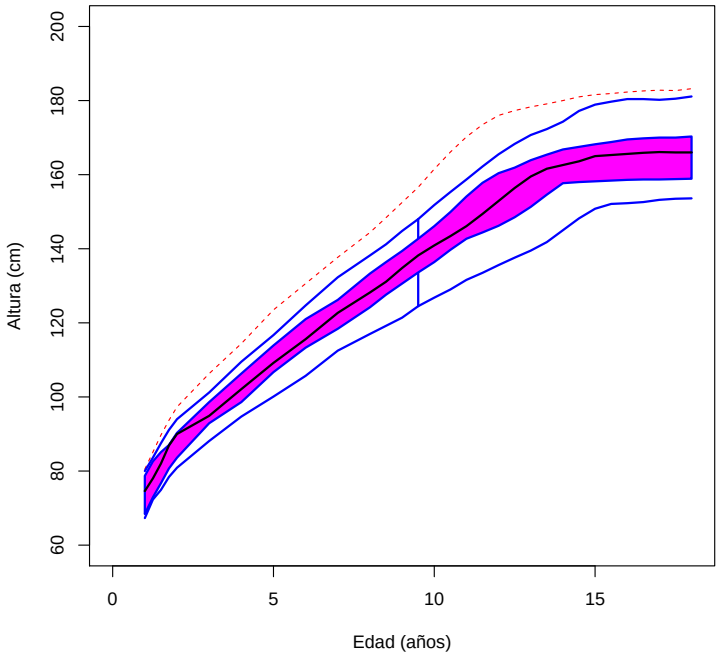



(d) Curvas para 54 mujeres.

# Boxplot funcional




(e) Boxplot funcional (varones).




(f) Boxplot funcional (mujeres) con un outlier (rojo).

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