Functional Data Analysis: Introducción

Graciela Boente

Introducción al FDA

- En muchos casos, observar y guardar funciones como resultado de un experimiento aleatorio es posible por el desarrollo de instrumentos de medición a tiempo real y de las capacidades de almacenamiento de datos.
- Ejemplo: Para pacientes incluídos en ensayos clínicos, la presión arterial se monitorea a tiempo continuo durante 24 o 48 hs.
- Evolución de la Estadística:
  univariada $\longrightarrow$ **multivariada** $\longrightarrow$ **funcional** (desde 1990, approx.).
- **Dato funcional**: en general, es una curva que procede de la realización de un proceso estocástico en tiempo continuo.

Introducción al FDA

- **Muestra**: conjunto de funciones observadas en instantes de tiempo.

- Dichas muestras en las que la función es observada en cada unidad muestral (individuo) se llaman **Datos Funcionales**

- El argumento de las curvas no siempre es el tiempo (Quimiometría)

- Las funciones aleatorias son los elementos estadísticos en estos casos. (Ramsay and Silverman 2005).

Introducción al FDA

- El Análisis de datos funcionales **FDA** se ocupa de la descripción estadística y el modelado de muestras de funciones aleatorias.
- Un amplio panorama sobre **FDA** puede verse en
  - Ramsay and Silverman (2002, 2005),
  - Ferraty and Vieu (2006),
  - Horváth and Kokoszka (2012), *Inference for functional data with applications*, Springer Science and Business Media.
  - Kokoszka and Reimherr (2017), *Introduction to functional data analysis*, Chapman and Hall CRC Press.
- Recursos disponibles en `R`
  - Librería `fda` (Ramsay, Wickham, Graves & Hooker)
  - Librería `fda.usc` (Febrero-Bande & Oviedo de la Fuente)
  - Librería `fds` (Hyndman & Shang)
  - `https://robjhyndman.com/software/`

Definición formal FDA

- Como en Ferraty y Vieu (2006) diremos que un elemento aleatorio $X : \Omega \to \mathcal{H}$ es una variable aleatoria funcional si $\mathcal{H}$, el espacio de posibles valores de X, tiene dimensión infinita. Una observación de X se dirá un dato funcional.

- Una observación x de X se llama un dato funcional.

- Un conjunto de datos funcionales $x_1, \ldots, x_n$ corresponde a la observación de n variables funcionales $X_1, \ldots, X_n$ con la misma distribución que X.

Definición formal FDA

- Sea $\mathcal{I} = [a; b] \subset \mathbb{R}$. Usualmente trabajamos con datos funcionales que son elementos de

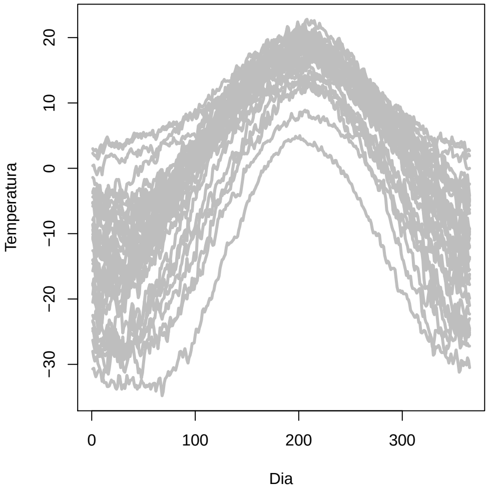
$$L^2(\mathcal{I}) = \{f : \mathcal{I} \to \mathbb{R};\ \text{tales que} \int_{\mathcal{I}} f^2(t)\, dt < \infty\}$$

- $L^2(\mathcal{I})$ con el producto interno usual

$$\langle f, g\rangle = \int_{\mathcal{I}} f(t)\, g(t)\, dt$$

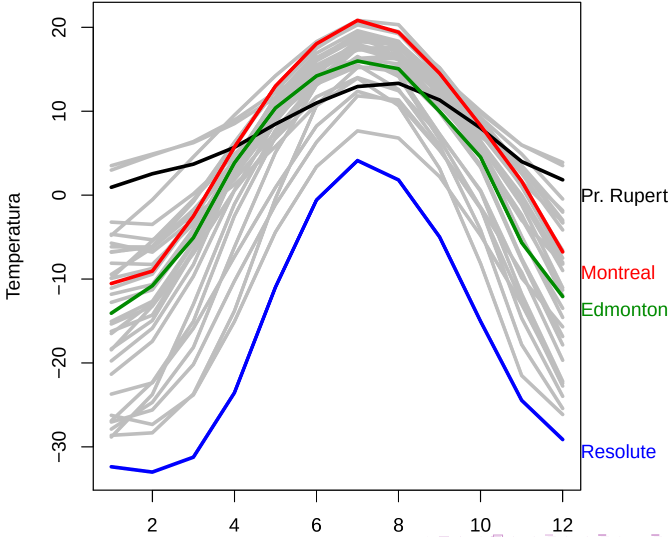
es un espacio de Hilbert separable.

Temperatura diarias de 35 estaciones climáticas en Canadá

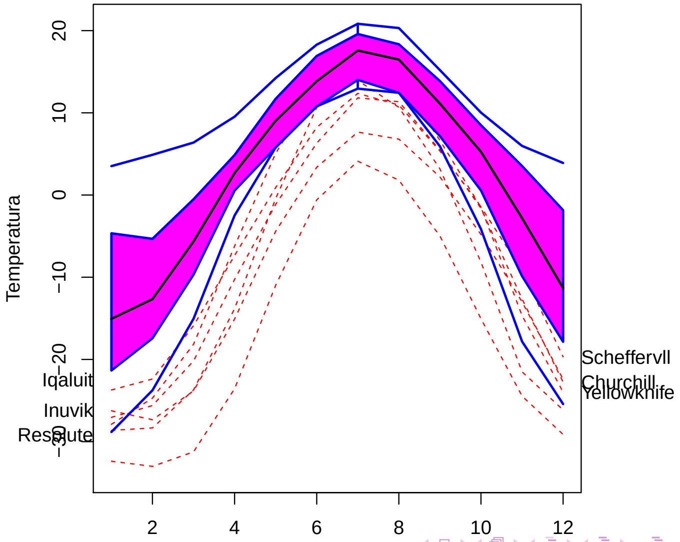


library(fda);data(daily) data(CanadianWeather)

Temperatura mensuales en Canadá: Promedios sobre 30 años



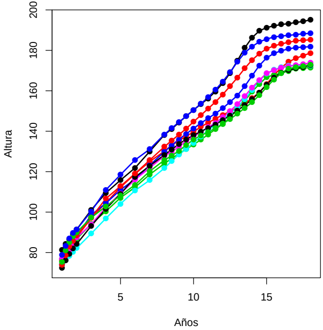
Temperatura mensuales en Canadá: Promedios sobre 30 años



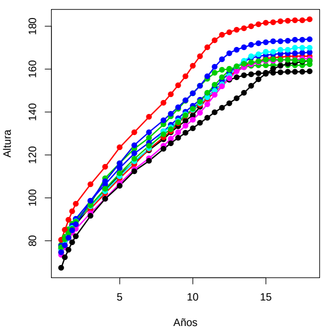
Datos de crecimiento

Altura de 39 niños y 54 niñas medidas sobre un conjunto de 31 edades (de 1 a 18 años) en el estudio *Berkeley Growth Study*

Varones

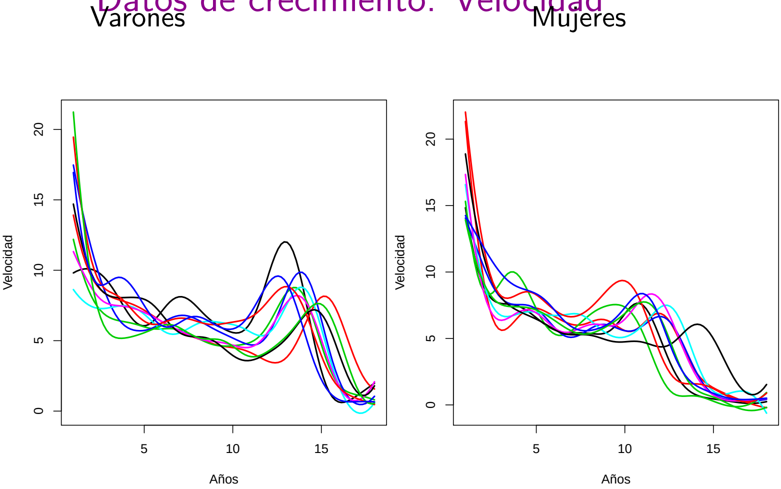


Mujeres




```
library(fda); data(growth); demo(growth)
```

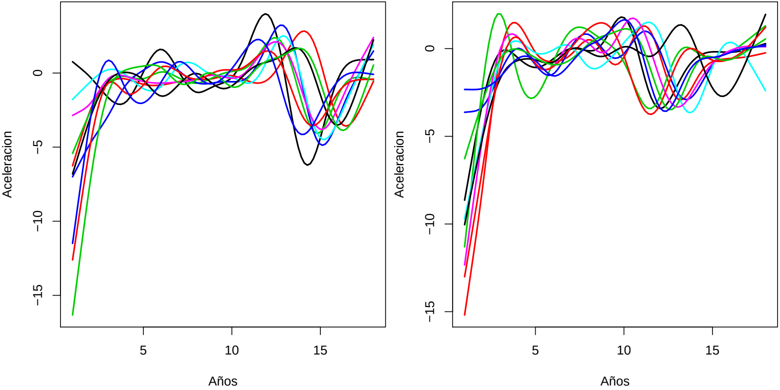
Datos de crecimiento: Velocidad



Las derivadas de datos funcionales son nuevos datos funcionales que pueden ser tan (o más) informativos que las funciones originales

Datos de crecimiento: Aceleración

Varones Mujeres



Reconstrucción

En la práctica se dispone de observaciones discretas, de modo que el primer paso en FDA es reconstruir la forma funcional.

- Información muestral: $X_1(t), X_2(t), \ldots, X_n(t)$
- Variable funcional: $X = \{X(t) : t \in \mathcal{I}\}$
- Funciones muestrales pertenecientes a $L^2(\mathcal{I})$

Problema: Observaciones discretas, x_{ik} , de $X_i(t)$ en tiempos

$$\{t_{i0}, t_{i1}, \ldots, t_{im_i}\} \quad t_{ij} \in \mathcal{I} \quad i = 1, \ldots, n$$

Reconstrucción

- En muchos casos, los datos se registran en la misma grilla de puntos

$$t_1, \ldots, t_p$$

para todos los individuos.

- Si el registro es hecho por un instrumento, como en el electroencefalograma o en la resonancia magnética, el tiempo es usualmente equiespaciado

$$t_j - t_{j-1} = t_{j+1} - t_j \qquad \text{para todo } j$$

- En el estudio de las propiedades asintótica, se supone que el espaciado $t_{j+1} - t_j$ converge a 0 cuando $n \to \infty$.

Por lo tanto, $p = p_n$ es una sucesión que crece a infinito.

Reconstrucción

- Aunque valores grandes de p llevan a lo que suele denominarse un problema de *alta dimensión*, también significa que se observan más datos en el intervalo $\mathcal{I}$, lo cual se convierte en una **bendición** más que en una maldición.

- Esta **bendición** se obtiene imponiendo suposiciones de **suavidad** sobre el proceso en L^2, de modo que mediciones en puntos de tiempo cercanas se puedan agrupar para superar la maldición de la dimensión.

- De esta forma, el suavizado sirve como herramienta de regularización.

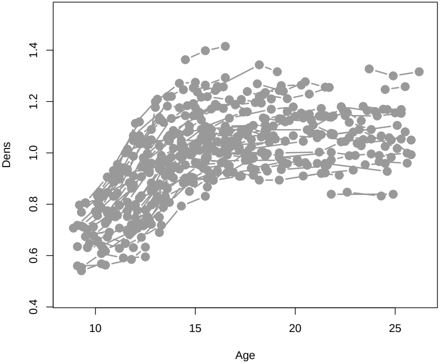
Reconstrucción

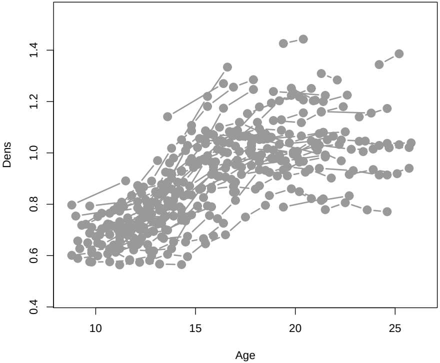
- Aunque no existe una definición formal de **datos funcionales densos**, la convención es declarar que los datos funcionales son densos (como opuesto a esparso) cuando $p_n \to \infty$ lo suficientemente rápido como para permitir que el estimador de $\mu(t) = \mathbb{E}X(t)$ tenga tasa de convergencia paramétrica $\sqrt{n}$ para distancia como la norma L^2.
- **Datos registrados en forma esparsa o rala** ocurren en muchos estudios longitudinales para los que los individuos se miden en diferentes tiempos con distintos números de mediciones por individuo.
- En este último caso, m_i el número de registros para el individuo i puede mantenerse acotado, o sea,

$$\sup_{1 \leq i \leq n} m_i \leq C < \infty$$

Datos ralos - Densidad ósea

Mujeres ($n = 153$) Hombres ($n = 127$)

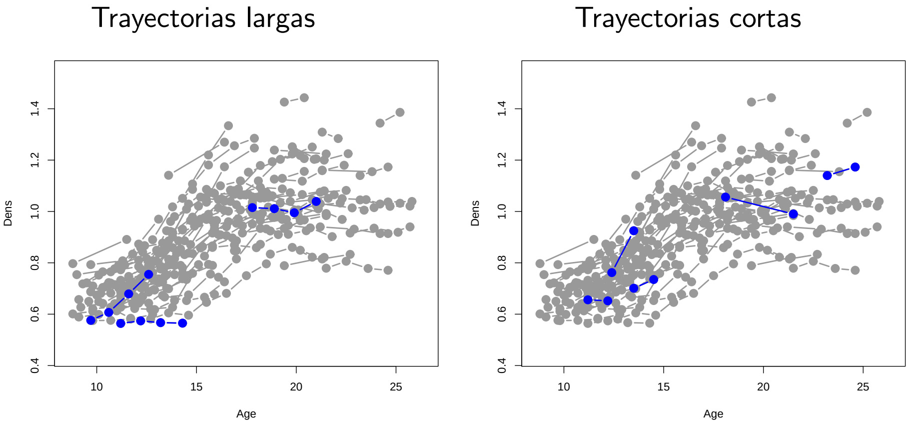




Densidad mineral espinal ósea (g/cm^2) tomada en diferentes edades. En forma agregada, hay 860 observaciones medidas sobre un período de casi 2 décadas, pero solo hay 2-4 mediciones para cada individuo.

Cortesía de Gareth James. James & Sugar (2003). ver también Yao, Wu & Zou (2016)

Datos ralos - Densidad ósea - Hombres



Solución en el caso denso

Supongamos que $X_i(t)$ se observa en tiempos

$$\{t_{i0}, t_{i1}, \ldots, t_{im_i}\} \quad t_{ij} \in \mathcal{I} \quad i = 1, \ldots, n$$

tales que $\{t_{i0}, t_{i1}, \ldots, t_{im_i}\}$ cubren $\mathcal{I}$

- Usar un suavizador como los basados en núcleos

$$\widehat{X}_i(t) = \frac{1}{nh} \sum_{j=1}^{m_i} K\left(\frac{t_{ij} - t}{h}\right) X_i(t_{ij})$$

  - **Inconveniente**: No permite evaluar fácilmente las derivadas, ya que las ventanas adecuadas para predecir $X_i(t)$ pueden no serlo para sus derivadas.

- **Usar Representación en bases**

Usar Representación en bases

Supongamos tener una base $\{\phi_j\}$ y sea p tal que

$$X_i(t) \approx \sum_{j=1}^{p} a_{ij}\phi_j(t)$$

- Usamos los valores $x_{ij} = X_i(t_{ij})$ donde $\{t_{i0}, t_{i1}, \ldots, t_{im_i}\}$ cubren $\mathcal{I}$ para estimar los coeficientes a_{ij} y obtener

$$\widehat{X}_i(t) = \sum_{j=1}^{p} \widehat{a}_{ij}\phi_j(t)$$

- **Ventaja**: Permite evaluar fácilmente las derivadas como

$$\widehat{X}_i^{(s)}(t) = \sum_{j=1}^{p} \widehat{a}_{ij}\phi_j^{(s)}(t)$$

Bases

Supongamos tener una base $\{\phi_j\}$ y sea p tal que

$$X_i(t) \approx \sum_{j=1}^{p} a_{ij}\phi_j(t)$$

- Bases fijas:
  - Base de Fourier
  - Base de Legendre
- Bases crecientes:
  - Base de Bernstein
  - Base de $B-$splines

Base de $B-$splines en $\mathcal{I} = [a, b]$

$$a = \tau_1 = \cdots = \tau_\ell < \tau_{\ell+1} < \cdots < \tau_{m+\ell+1} = \cdots = \tau_{m+2\ell} = b$$

$$\mathcal{T} = \{\tau_i\}_{i=1}^{m+2\ell}$$

$$\mathcal{I}_j = [\tau_{\ell+j}, \tau_{\ell+j+1}) \quad j = 0, \ldots, m-1$$

$$\mathcal{I}_m = [\tau_{m+\ell}, \tau_{m+\ell+1}]$$

- Dado un entero positivo ℓ, definamos

$$\mathcal{PP}_\ell = \{f : \text{ existen polinomios } p_0, p_1, \ldots, p_k \text{ de grado a lo sumo } \ell - \text{ tales que } f(t) = p_j(t) \text{ para todo } t \in \mathcal{I}_j\}.$$

Llamaremos a $\mathcal{PP}_m$ el espacio de los polinomios de a trozos de orden ℓ con nodos $\tau_\ell < \cdots < \tau_{m+\ell+1}$,

Base de $B-$splines en $\mathcal{I} = [a, b]$

- El espacio de los splines polinomiales de orden $\ell > 1$ con nodos $\mathcal{T}$, es

$$\mathcal{S}_\ell(\mathcal{T}) = \mathcal{PP}_\ell \cap \mathcal{C}^{\ell-2}([a, b])$$

- De acuerdo al Corolario 4.10 de Schumaker (1981), para cualquier $g \in \mathcal{S}_\ell(\mathcal{T})$ existe una clase de $B-$splines $\{B_j : 1 \leq j \leq k\}$, con $k = m + \ell$, tal que

$$g = \sum_{j=1}^{k} \lambda_j \, B_j \, .$$

Base de $B-$splines en $\mathcal{I} = [a, b]$

- La base cumple

$$B_j(t) = 0 \quad t \notin [y_j, y_{j+\ell}] \,, \qquad B_j(t) > 0 \quad t \in (y_j, y_{j+\ell})$$

$$\sum_{j=1}^{k} B_j(t) = 1 \qquad t \in [a, b]$$

- Más aún dada $f \in\in \mathcal{H}_r$, $\ell \geq r + 2$

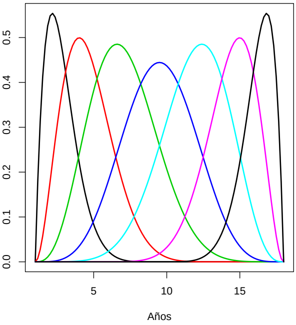
$$\mathcal{H}_r = \{g \in C^r[a, b] : \|g^{(j)}\|_\infty \leq C_1, \; 0 \leq j \leq r \;\; \text{y} \\ |g^{(r)}(z_1) - g^{(r)}(z_2)| \leq C_2 |z_1 - z_2|\} \,.$$

existe $g \in \mathcal{S}_\ell(\mathcal{T})$ tal que $\|g - f\|_\infty = O(k^{-r})$

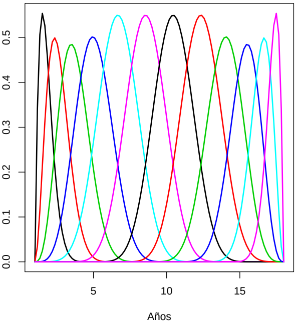
```
create.bspline.basis(rangeval = c(a, b), norder = ℓ,
breaks = 𝒯)
```

Datos de crecimiento: Bases de $B-$splines

5 nodos

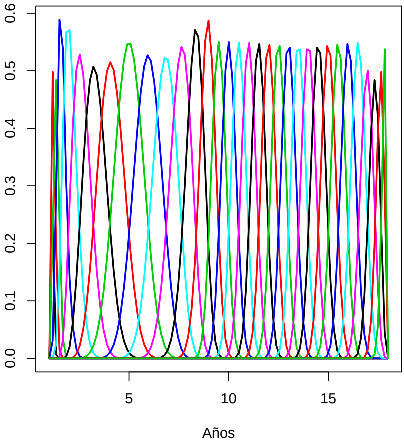


10 nodos



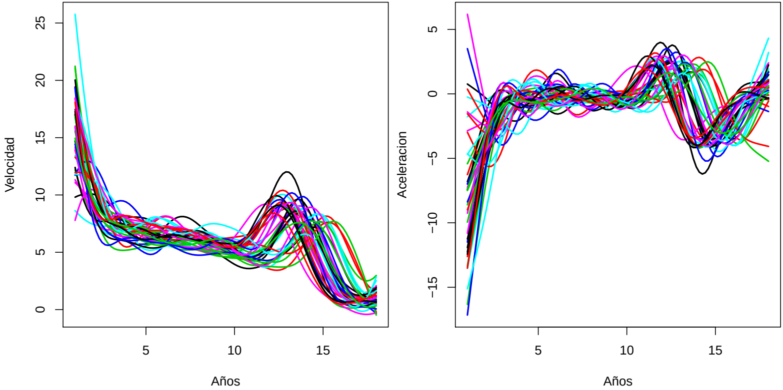
Datos de crecimiento: Bases de $B-$splines

Para los gráficos presentados de velocidad y aceleración se tomaron 31 nodos

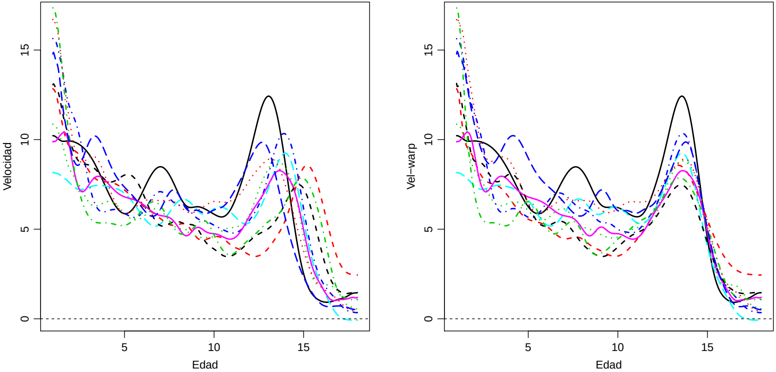


Datos de crecimiento: Alineamiento

Velocidad y aceleración de niños



Datos de crecimiento: Alineamiento velocidad



Esperanza

Sea $X : \Omega \to \mathcal{H}$ un elemento aleatorio.

Diremos que X es integrable si $\mathbb{E}\|X\| < \infty$.

- $\mu = \mathbb{E}(X)$ es el elemento de $\mathcal{H}$ que verifica que

$$\langle \mu, u \rangle = \mathbb{E}(\langle X, u \rangle) \qquad \forall u \in \mathcal{H}$$

- Si $\mathcal{H} = L^2(\mathcal{I})$

$$\mu(t) = \mathbb{E}(X(t)) \qquad \forall t \in \mathcal{I}$$

Covarianza

Sea $X : \Omega \to \mathcal{H}$ un elemento aleatorio.

Diremos que X es de cuadrado integrable si $\mathbb{E}\|X\|^2 < \infty$, en cuyo caso

- El operador de covarianza de X, Γ, se define como el operador lineal $\Gamma : \mathcal{H} \to \mathcal{H}$ tal que

$$\langle \Gamma u, v\rangle = \textsc{Cov}(\langle u, X\rangle, \langle v, X\rangle) \forall u, v \in \mathcal{H}$$

- En particular, $\langle \Gamma u, u\rangle = \textsc{Var}(\langle u, X\rangle^2)$

- $\Gamma u = \mathbb{E}\left(\langle u, X - \mu\rangle (X - \mu)\right)$

- $\Gamma = \mathbb{E}\left\{(X - \mu) \otimes (X - \mu)\right\}$
  donde $u \otimes v : \mathcal{H} \to \mathcal{H}$ se define como $(u \otimes v)w = \langle v, w\rangle\ u$.

Covarianza

- Si $\mathcal{H} = L^2(\mathcal{I})$, se llama núcleo de covarianza γ a

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

Se cumple

$$\begin{aligned}(\Gamma f)(s) &= \int_{\mathcal{I}} \mathrm{Cov}\left\{X(s), X(t)\right\} f(t) dt \\ &= \int_{\mathcal{I}} \gamma(s,t) f(t) dt\,,\end{aligned}$$

Covarianza

- Γ es un operador simétrico, definido no negativo. Por lo tanto, sus autovalores son no–negativos.

- Más aún, si $\lambda_j \geq 0$ son sus autovalores asociados a las autofunciones ϕ_j (ortonormales), por la igualdad de Parseval

$$\sum_{j\geq 1} \lambda_j = \sum_{j\geq 1} \mathbb{E}\left\{\langle X, \phi_j\rangle^2\right\} = \mathbb{E}\|X\|^2 < \infty$$

pues

$$\lambda_j = \langle \Gamma\phi_j, \phi_j\rangle = \mathbb{E}\left\{\langle X, \phi_j\rangle^2\right\}$$

Por lo tanto, Γ es un operador acotado de traza y esta condición es necesaria.

Covarianza

- **Contraejemplo** Sea $\mathcal{H} = L^2(0,1)$ y $\{\phi_j\}_{j\geq 1}$ una base ortonormal, definamos

$$\Upsilon = \sum_{j\geq 1} \frac{1}{j}\phi_j \otimes \phi_j$$

  - Υ es un operador acotado pues

$$\|\Upsilon u\|^2 = \left\| \sum_{j\geq 1} \frac{1}{j}\langle \phi_j, u\rangle \phi_j \right\| = \sum_{j\geq 1} \frac{1}{j^2}\langle \phi_j, u\rangle^2 \leq \sum_{j\geq 1}\langle \phi_j, u\rangle^2 = \|u\|^2$$

  - Υ es claramente simétrico y definido no–negativo. Los autovalores de Υ son $1/j$ y como

$$\sum_{j\geq 1} \frac{1}{j} = \infty$$

    Υ no es un operador de covarianza.

Ley de los grandes números

- $(\Omega, \mathcal{B}, \mathbb{P})$ un espacio de probabilidad.
- $\mathcal{H}$ un espacio de Hilbert separable.
- $X_n : \Omega \to \mathcal{H}$ una sucesión de elementos aleatorios i.i.d., $X_i \sim X$, tales que $\mathbb{E}\|X\|^2 < \infty$.
- Sea $\mu = \mathbb{E}X$, entonces

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu$$

$$\mathbb{P}\left(\|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\| \to 0\right) = 1$$

Ley de los grandes números

Chen, Y. & Zhu, W. (2011). Note on the strong law of large numbers in a Hilbert space. *General Mathematics*, **19**, 11-18.

- $(\Omega, \mathcal{B}, \mathbb{P})$ un espacio de probabilidad.
- $\mathcal{H}$ un espacio de Hilbert separable.
- $X_n : \Omega \to \mathcal{H}$ una sucesión de elementos aleatorios **independientes de a pares**, $X_i \sim X$, tales que $\mathbb{E}\|X\| < \infty$.
- Sea $\mu = \mathbb{E}X$, entonces

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mu$$

Elementos Gaussianos

- $(\Omega, \mathcal{B}, \mathbb{P})$ un espacio de probabilidad. $\mathcal{H}$ un espacio de Hilbert separable.

- Un elemento aleatorio $X : \Omega \to \mathcal{H}$ se dice Gaussiano si $\langle u, X \rangle$ es una variable aleatoria normal para todo $u \in \mathcal{H}$.

- X es Gaussiano si y sólo si $\mathbb{E}\|X\|^2 < \infty$ y para todo $u \in \mathcal{H}$

$$\varphi_X(u) = \mathbb{E}\left\{\exp\left(i\langle u, X\rangle\right)\right\} = \exp\left(i\langle u, \mu\rangle - \langle u, \Gamma u\rangle\right)$$

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

- Si X es Gaussiano entonces $\langle u, X\rangle \sim N(\langle u, \mu\rangle, \langle u, \Gamma u\rangle)$

Elementos Gaussianos: Ejemplos

- **Wiener** Un proceso de Wiener (o movimiento Browniano) $W(t) \in \mathcal{C}([0,1]) \subset L2([0,1])$ es un proceso Gaussiano tal que
  - $\mu = 0$
  - $\gamma(s,t) = \sigma^2 \min(s,t)$, con $\sigma^2 > 0$.
- W corresponde al desplazamiento de una partícula en un fluido homogeneo y sin movimiento.
- Un proceso de Wiener tiene incrementos independientes, o sea, si $0 \leq t_1 < t_2, \ldots, < t_k \leq 1$, entonces

$$W(t_2) - W(t_1), W(t_3) - W(t_2), \ldots, W(t_k) - W(t_{k-1})$$

son independientes.

Elementos Gaussianos: Ejemplos

- **Un puente Browniano** es el elemento aleatorio $X(t)$ con la misma distribución que

$$X(t) \sim W(t) - t\, W(1)$$

siendo $W(t)$ un movimiento Browniano.

  - $X(1) = X(0) = 0$ con probabilidad 1
  - $\mathbb{E}(X) = 0$
  - $\mathbb{E}X(s)X(t) = \sigma^2 s(1-t)$, si $s \leq t$.

Elementos Gaussianos: Ejemplos

- **Un proceso de Ornstein-Uhlenbeck** es el elemento aleatorio $Z(t)$ Gaussiano tal que

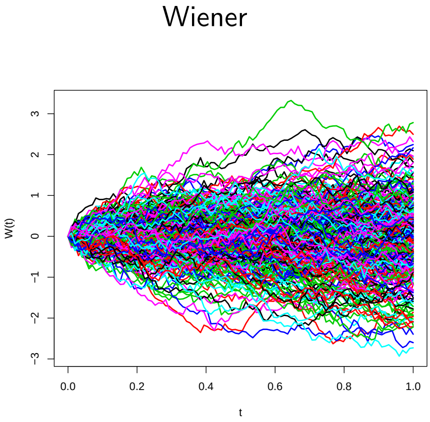
$$\mathbb{E}(Z) = 0 \qquad \mathrm{Cov}(Z(s), Z(t)) = \frac{\sigma^2}{2\,\theta} e^{-\theta(s+t)} \left(e^{2\theta(s+t)} - 1\right)$$

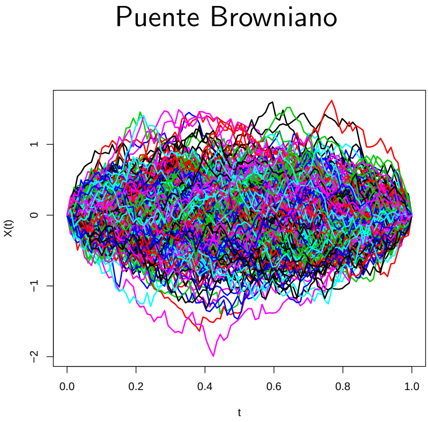
- **Proceso $Y(t)$** es el elemento aleatorio Gaussiano tal que

$$\mathbb{E}(Y) = 0 \qquad \mathrm{Cov}(Y(s), Y(t)) = \sigma^2 e^{-\frac{1}{\theta}|s+t|}$$

La función `rproc2fdata` del paquete `fda.usc` permite generar los procesos anteriores. Ver también la librería `somebm`

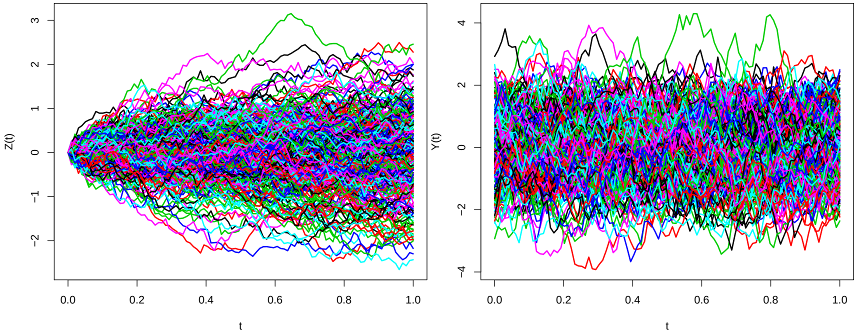
Elementos Gaussianos: Ejemplos



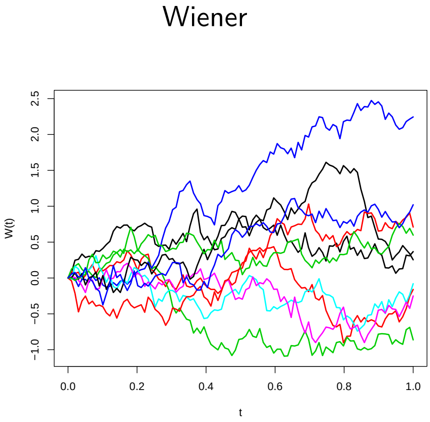


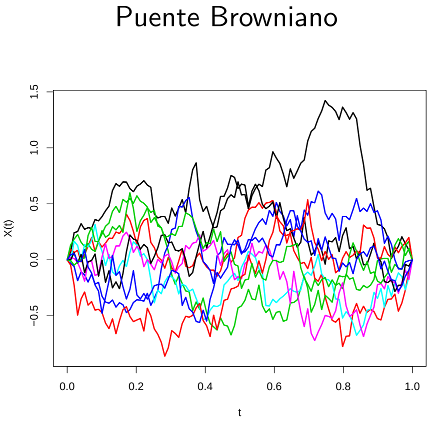
Elementos Gaussianos:Ejemplos

Ornstein-Uhlenbeck $Y(t)$



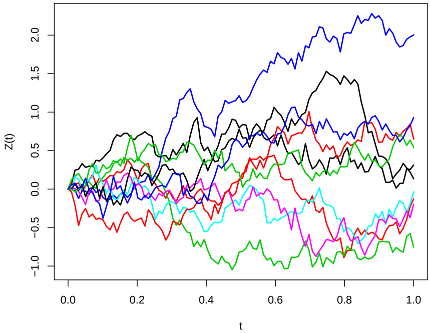
Elementos Gaussianos: Ejemplos



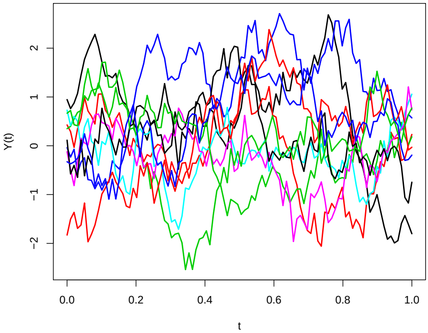


Elementos Gaussianos:Ejemplos

Ornstein-Uhlenbeck



$Y(t)$



Teorema central del límite

Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York

- $(\Omega, \mathcal{B}, \mathbb{P})$ un espacio de probabilidad.
- $\mathcal{H}$ un espacio de Hilbert separable.
- $X_n : \Omega \to \mathcal{H}$ una sucesión de elementos aleatorios **independientes de a pares**, $X_i \sim X$, tales que $\mathbb{E}\|X\|^2 < \infty$.

Teorema central del límite

- Sea $\mu = \mathbb{E}X$ y $\Gamma = \mathbb{E}\{(X-\mu)\otimes(X-\mu)\}$, entonces

$$Y_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i-\mu) \xrightarrow{D} N(0,\Gamma)$$

  - Sea $Y \sim N(0,\Gamma)$. Para todo $u \in \mathcal{H}$,

$$\langle u, Y_n\rangle \xrightarrow{D} \langle u, Y\rangle$$

  - Dado $\epsilon > 0$, existe $\mathcal{K}_\epsilon \subset \mathcal{H}$ compacto tal que

$$\mathbb{P}\left(Y_n \in \mathcal{K}_\epsilon\right) > 1-\epsilon \qquad \forall n$$