



Universidad de Valladolid

Facultad de Ciencias

Máster en Matemáticas

**Aspectos matemáticos y computacionales del
Análisis de Datos Funcionales**

Ismael Crehuet Lucas

Tutora: Luis Ángel García Escudero

Curso: 2023-2024

Resumen

Este Trabajo Fin de Máster se adentra en aspectos matemáticos relativos al Análisis de Datos Funcionales. Se hace uso de teoría matemática en espacios de Hilbert y se asume que las funciones aleatorias que aparecen en el Análisis de Datos Funcionales toman valores en un espacio de Hilbert separable. Se analizan conceptos esenciales en el Análisis de Datos Funcionales, estableciendo las bases teóricas para la posterior aplicación de técnicas estadísticas, detallando el estudio de las Componentes Principales Funcionales. Además, se proporcionan ejemplos prácticos para ilustrar la aplicación de estas técnicas a conjuntos de datos funcionales reales, mediante la aplicación del software estadístico R.

Palabras clave: Análisis de Datos Funcionales, Componentes Principales Funcionales, representación en bases funcionales, espacios de Hilbert.

Abstract

This Master's Final Project delves into mathematical aspects related to Functional Data Analysis. It makes use of mathematical theory in Hilbert spaces and assumes that the random functions appearing in Functional Data Analysis take values in a separable Hilbert space. Essential concepts in Functional Data Analysis are covered, establishing the theoretical foundations for the subsequent application of statistical techniques, detailing the study of Functional Principal Components. Additionally, practical examples are provided to illustrate the application of these techniques to real functional data sets, using statistical software R.

Key words: Functional Data Analysis, Functional Principal Components, representation in functional bases, Hilbert spaces.

Índice

1. Introducción	5
1.1. Introducción a los datos funcionales	5
2. Aspectos matemáticos	10
2.1. Espacios de Hilbert	10
2.2. Teoría espectral	18
2.3. Funciones aleatorias	21
2.4. Operador de covarianza y expansión de Karhunen-Loève	22
2.5. Representación en bases funcionales	28
3. Componentes principales funcionales empíricas	37
3.1. Obtención desde la discretización de las funciones	38
3.2. Obtención desde su representación en bases funcionales	41
4. Ejemplos	44
4.1. Ejemplo 1: Datos AEMET	44
4.2. Ejemplo 2: Movimiento Browniano	57
5. Conclusiones y líneas futuras	59

1. Introducción

1.1. Introducción a los datos funcionales

En términos generales, se puede pensar en *Análisis de Datos Funcionales* (FDA, Functional Data Analysis) como un enfoque estadístico que surge cuando los “datos” pueden ser naturalmente vistos como *funciones* o *curvas* en un espacio funcional. Así, una idea fundamental del FDA es que los objetos que deseamos estudiar son funciones:

$$\{x_i(t) : t \in [T_1, T_2], \quad i = 1, 2, \dots, n\}, \quad (1)$$

para las cuales los valores $x_i(t)$, en principio, existirían en cualquier punto t en el intervalo $[T_1, T_2]$.

Si denotamos por x_1, x_2, \dots, x_n a las funciones observadas, el conjunto de datos más simple suele aparecer por una discretización de estas funciones y el conocimiento de

$$x_i(t_{i,j}) \in \mathbb{R}, \quad t_{i,j} \in [T_1, T_2], \quad i = 1, 2, \dots, n, \quad j = 1, \dots, J_i. \quad (2)$$

Es decir, $\{t_{i,j}\}_{j=1}^{J_i}$ son los momentos en que se observa la función x_i , para $i = 1, \dots, n$, y pueden ser distintos en cada curva x_i .

Se suele contar con n funciones en un intervalo común $[T_1, T_2]$. Para simplificar la notación, comúnmente se asume que todas las curvas o funciones están definidas en el intervalo $[0, 1]$, mediante la transformación

$$t \mapsto \frac{t - T_1}{T_2 - T_1}.$$

El FDA es una técnica relativamente nueva, considerando la longevidad de la mayoría de las herramientas y ramas de las Matemáticas. Si bien el término de “Análisis de Datos Funcionales” fue acuñado por Ramsay (1982) [7] y por Ramsay y Dalzell (1991) [8], la historia de esta área es mucho más antigua y se remonta, por ejemplo, a Grenander (1950) [3] y Rao (1958) [12].

Los objetivos del FDA son esencialmente los mismos que los de cualquier otra rama de la Estadística. En esta memoria nos centraremos en buscar procedimientos que permitan representar los datos de manera adecuada para destacar sus características más importantes y estudiar fuentes importantes de variaciones entre los datos.

Como hemos comentado, el FDA se ocupa de conjuntos de datos donde las observaciones aleatorias son funciones. En contraste, el Análisis de Datos Multivariantes se centra en conjuntos de datos donde cada observación es un vector que almacena las múltiples variables involucradas. Aunque ambos enfoques comparten similitudes en términos de modelado estadístico, la naturaleza de los datos funcionales presenta desafíos y características distintivas que hacen que tratarlos como datos multivariantes en \mathbb{R}^p no sea lo más adecuado, aún cuando las funciones se encuentren discretizadas. En los datos multivariantes, las variables consideradas están organizados en un vector, y se trata de entender las relaciones y dependencias que existen entre tales variables. Por otro lado, en FDA, cada observación es una función que varía en una dimensión, como el tiempo o el espacio. Este enfoque implica que los valores en cualquier punto de la función estén de alguna forma relacionados con evaluaciones de la función en otros puntos o momentos cercanos, mientras que en datos multivariantes, el orden en que se proporcionan las variables aleatorias que constituyen el vector no son relevantes.

Además, la principal dificultad al tratar datos funcionales como datos multivariantes en \mathbb{R}^p radica en la alta dimensionalidad del espacio funcional, a la que nos puede llevar la discretización de las funciones. En datos multivariantes, p representa el número de variables, pero en datos funcionales, la dimensión sería infinita, ya que al menos a nivel teórico se considera la evaluación de la función a lo largo de la dimensión funcional, que sería esencialmente el objeto aleatorio tratado. Esto plantea problemas teóricos y computacionales significativos. Las funciones suelen exhibir patrones de suavidad y dependencias temporales o espaciales que no pueden ser capturados eficientemente por las técnicas multivariantes convencionales.

Otro aspecto importante es la variabilidad, que a diferencia de los datos multivariantes donde la variabilidad entre observaciones se puede representar mediante una matriz de covarianza finita, la variabilidad en datos funcionales se extiende a lo largo de una dimensión continua. Esto plantea desafíos en la definición y cálculo de dichas medidas. Para abordar estos problemas, se han desarrollado métodos especializados en el ámbito del FDA, que consideran la naturaleza continua de las funciones y aprovechan herramientas matemáticas no triviales. Así, por ejemplo, se utilizan técnicas de reducción de la dimensionalidad específicas, como el Análisis de Componentes Principales Funcionales, que trataremos con más detalle en esta memoria.

El interés y popularidad creciente en el FDA se debe a varios factores. Uno

de esos factores son los avances tecnológicos, ya que la recopilación de datos temporales se ha vuelto cada vez más accesible en muy diversas áreas. Las nuevas tecnologías permiten capturar datos a alta frecuencia y resolución, lo que lleva a conjuntos de datos muy ricos y detallados. Además, el avance en técnicas estadísticas y herramientas computacionales ha facilitado el análisis de este tipo de datos tan complejos.

En la Figura 1 se presentan algunos conjuntos de datos funcionales, entre los cuales se incluyen a modo de ejemplo las tasas de interés diarias de la curva de rendimiento del Banco Central Europeo (BCE) donde cada curva representa el rendimiento diario para bonos del tesoro; la temperatura media diaria registrada en diversas estaciones meteorológicas según datos proporcionados por la Agencia Estatal de Meteorología (AEMET) donde cada curva representa la temperatura diaria promedio registrada en cada estación durante un periodo concreto; los niveles de NO_x (óxidos de nitrógeno) en la localidad de Poblenou donde cada curva representa la evolución de los niveles en un día; el cambio porcentual del consumo mensual de electricidad en Estados Unidos donde se mide el consumo mensual de electricidad en los sectores residencial y comercial; las tasas de cáncer de mama para diferentes grupos de edad en mujeres australianas y las longitudes de onda recopiladas por el satélite TOPEX en el Amazonas. Nos centraremos en el ejemplo de la AEMET con más detalle en la Sección 4.

El objetivo de esta memoria es presentar algunos aspectos matemáticos del FDA y poner ejemplos de su aplicabilidad mediante software estadístico (en concreto software R [13]).

A nivel teórico, trabajaremos en el espacio de las funciones aleatorias cuadrado integrables, visto como un espacio de Hilbert, y, además, consideraremos que las funciones son realizaciones de una función aleatoria definida en un espacio probabilístico. Será importante tener en cuenta la distinción entre una función determinista cuadrado integrable (la integración se refiere al intervalo) y una función aleatoria cuadrado integrable (la integración se refiere al espacio de probabilidad). Daremos detalles más adelante de esta distinción.

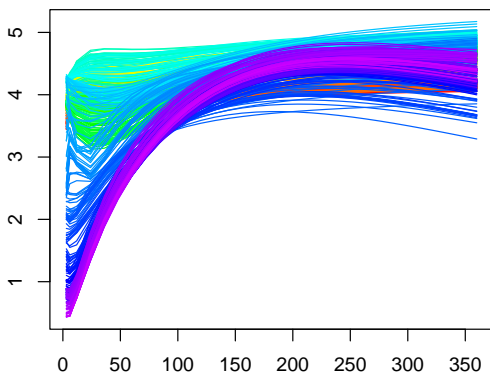
La Sección 2 aborda aspectos matemáticos esenciales en FDA relacionados con el estudio de funciones en espacios de Hilbert que resultan de interés, ya que proporcionan el sólido marco teórico para comprender la estructura y las propiedades de los datos funcionales. En esta sección se tratarán, pues, aspectos como las funciones aleatorias, el operador de covarianza, la teoría espectral o la representación en bases funcionales. Estos conceptos establecen unas bases teóricas

necesarias para comprender y aplicar técnicas estadísticas a nivel práctico en FDA.

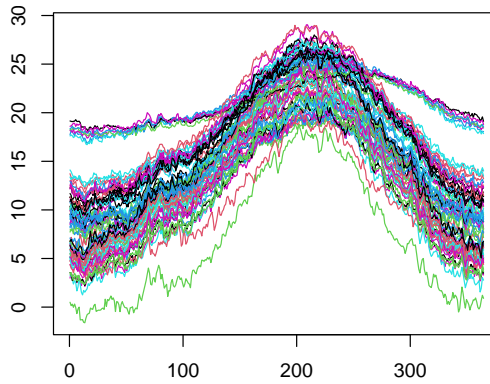
En la Sección 3 nos centraremos en el estudio detallado de las Componentes Principales Funcionales, explorando su obtención tanto a partir de la discretización de funciones como desde su representación en bases funcionales. Estas Componentes Principales Funcionales son cruciales para analizar la variabilidad y la estructura subyacente en conjuntos de datos funcionales.

Para ilustrar la aplicación práctica de los conceptos vistos en secciones anteriores, la Sección 4 presenta ejemplos concretos de aplicación de las técnicas vistas a nivel teórico al análisis de conjuntos de datos reales. Esto proporcionará una visión más completa y de carácter también más aplicado del FDA.

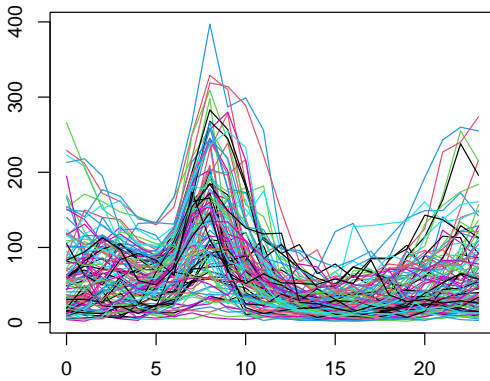
Finalmente, la Sección 5 presentará las conclusiones de esta memoria y describirá brevemente algunas posibles líneas de trabajo futuro.



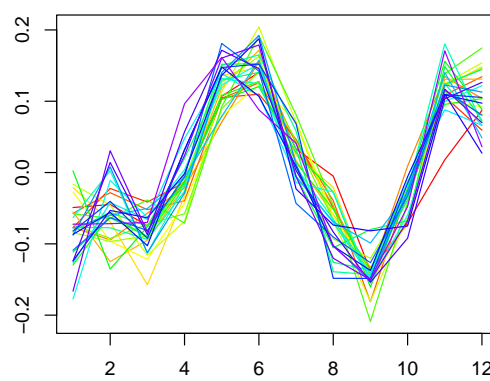
(a) Tasas de interés diarias.



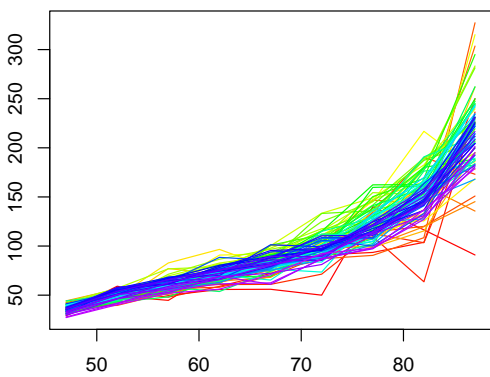
(b) Evolución de las temperaturas medias.



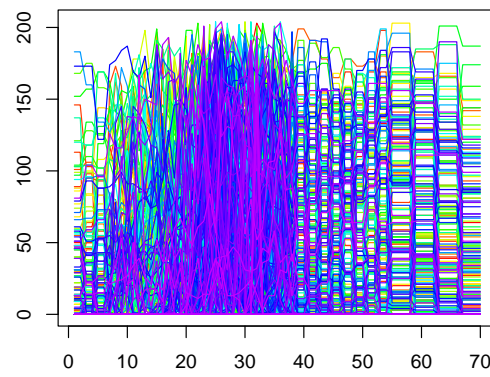
(c) Evolución de concentraciones de NOx.



(d) Consumo de electricidad.



(e) Tasas de cáncer de mama por edad.



(f) Longitudes de onda.

Figura 1: Ejemplos de datos funcionales.

2. Aspectos matemáticos

2.1. Espacios de Hilbert

La mayoría de las herramientas a nivel matemático del FDA hacen uso de los espacios de Hilbert. En este capítulo trataremos algunos aspectos matemáticos de los espacios de Hilbert que nos resultarán de interés en el correcto tratamiento de los datos funcionales.

Diremos que \mathcal{V} es un *espacio vectorial con producto interno* si para cada par de vectores x e y , hay un escalar $\langle x, y \rangle$, denominado producto interno de x e y , que satisface, para todo $(x, y) \in \mathcal{V} \times \mathcal{V}$:

- I) $\langle x, y \rangle = \overline{\langle y, x \rangle}$ (complejo conjugado);
- II) $\langle a_1x_1 + a_2x_2, y \rangle = a_1\langle x_1, y \rangle + a_2\langle x_2, y \rangle$;
- III) $\langle x, x \rangle \geq 0$ con igualdad si $x = 0$.

La norma del producto interno en el espacio vectorial se define por $\|x\| = \sqrt{\langle x, x \rangle}$, y cumple las siguientes propiedades:

- I) $\|ax\| = |a|\|x\|$;
- II) $\|\langle x, y \rangle\| \leq \|x\|\|y\|$;
- III) $\|x + y\| \leq \|x\| + \|y\|$;
- IV) $d(x, y) = \|x - y\|$ es una métrica.

La última propiedad IV) implica que todo espacio con producto interno es un *espacio métrico*, de hecho, es normado. Un *espacio normado* es un espacio vectorial con la norma que cumple las propiedades I) y III) anteriores; y $\|x\| = 0$ si y solo si $x = 0$. Todo espacio normado es un espacio métrico con la distancia definida por:

$$d(x, y) = \|x - y\|.$$

Recordemos también que una sucesión $\{x_i\}_i$ de puntos en un espacio métrico converge a un punto x si $d(x_i, x) \rightarrow 0$, cuando $i \rightarrow \infty$ y que una sucesión $\{x_i\}_i$ es una sucesión de Cauchy si $d(x_i, x_j) \rightarrow 0$, cuando $i, j \rightarrow \infty$. Dado que se tiene,

$$d(x_i, x_j) \leq d(x_i, x) + d(x, x_j),$$

cada sucesión convergente es una sucesión de Cauchy, pero en la otra dirección no es cierto, es decir, una sucesión de Cauchy no tiene por qué ser convergente.

Por otro lado, un espacio métrico en el cual cada sucesión de Cauchy tiene un límite es llamado *completo*, y un espacio vectorial con producto interno completo se dice que es un *espacio de Hilbert*.

Una pregunta que debemos abordar es en qué tipo de espacio viven los objetos funcionales, a los que simplemente llamamos funciones. Resulta que la configuración más favorable surge si las funciones se tratan como elementos de un espacio mucho más grande, el *espacio vectorial de funciones cuadrado integrables* L^2 , que es un espacio de Hilbert. Una función f es de cuadrado integrable si se cumple que,

$$\int f^2(t)dt < \infty.$$

Se define el producto interno de dos funciones, $f, g \in L^2$ como

$$\langle f, g \rangle = \int f(t)g(t)dt.$$

El producto interno también permite introducir la noción de distancia entre funciones, definiendo en primer lugar la norma

$$\|f\| = \sqrt{\langle f, f \rangle} = \left\{ \int f^2(t)dt \right\}^{1/2}$$

y, después, considerando $\|f - g\|$ como la distancia entre las funciones f y g . Existe una relación muy útil entre el producto interno de dos funciones y sus normas que se conoce como la *desigualdad de Cauchy-Schwarz*:

$$\left| \int f(t)g(t)dt \right| = |\langle f, g \rangle| \leq \|f\| \|g\| = \left\{ \int f^2(t)dt \right\}^{1/2} \left\{ \int g^2(t)dt \right\}^{1/2}.$$

En L^2 , se tiene que una sucesión $\{x_i\}_i$ converge a una función x si se tiene que:

$$\int [x_i(t) - x(t)]^2 dt \rightarrow 0.$$

Aunque los datos funcionales, típicamente una colección de funciones $x_i \in L^2$ definidas en un intervalo común, se ven como objetos de dimensión infinita, el producto interno en L^2 permite considerar la noción de ortogonalidad, lo que hace que la estructura geométrica del espacio sea muy similar a la del espacio euclidiano de dimensión finita, al menos a nivel intuitivo.

Una técnica común en FDA es proyectar las funciones en subespacios de dimensión finita para facilitar posteriores cálculos estadísticos.

Teorema 1. *Supongamos que \mathcal{G} es un subespacio cerrado de un espacio de Hilbert \mathcal{H} . Para cualquier $x \in \mathcal{H}$, definimos*

$$\delta = \inf_{z \in \mathcal{G}} \|x - z\|.$$

Entonces existe un único $y \in \mathcal{G}$ tal que $\|x - y\| = \delta$.

La prueba del Teorema 1 se puede encontrar en [4]. Un conjunto es cerrado si contiene todos sus puntos límite, es decir, si para cualquier sucesión $\{x_i\}_i$ convergente en \mathcal{G} , tenemos que $x := \lim x_i \in \mathcal{G}$. El vector y es la *proyección* de x en \mathcal{G} , el punto de \mathcal{G} más cercano a x .

Una función $L : \mathcal{H} \rightarrow \mathcal{H}$ se llama *operador lineal* o *transformación lineal* si para cualquier escalar a, b y cualquier $x, y \in \mathcal{H}$, se cumple la relación

$$L(ax + by) = aL(x) + bL(y). \quad (3)$$

Se dice que L es acotado si existe K constante tal que para cualquier $x \in \mathcal{H}$,

$$\|L(x)\| \leq K\|x\|.$$

La menor constante K es la llamada *norma del operador* L y se denota como $\|L\|_{\mathcal{L}}$.

Otra interpretación de la proyección se basa en el concepto de ortogonalidad. Se dice que x es *ortogonal* a y si $\langle x, y \rangle = 0$. Si \mathcal{G} es un subespacio, entonces el *complemento ortogonal* \mathcal{G}^\perp denota el conjunto de todos los $y \in \mathcal{H}$ que son ortogonales a cada vector en \mathcal{G} .

Teorema 2. *Supongamos que \mathcal{G} es un subespacio cerrado de un espacio de Hilbert \mathcal{H} . Entonces:*

- I) *Cada $x \in \mathcal{H}$ tiene una descomposición única $x = P(x) + Q(x)$, donde $P(x) \in \mathcal{G}$ y $Q(x) \in \mathcal{G}^\perp$. El operador lineal P se llama la *proyección en \mathcal{G}* .*
- II) *$P(x)$ y $Q(x)$ son los puntos más cercanos a x en \mathcal{G} y \mathcal{G}^\perp , respectivamente.*
- III) *Las aplicaciones $P : \mathcal{H} \rightarrow \mathcal{G}$ y $Q : \mathcal{H} \rightarrow \mathcal{G}^\perp$ son lineales.*
- IV) $\|x\|^2 = \|P(x)\|^2 + \|Q(x)\|^2$.

Una consecuencia de este teorema es que la proyección $P(x)$ es ortogonal a $x - P(x)$. El Teorema 2 conduce al siguiente resultado, conocido como el teorema de representación de Riesz. Es fácil ver que para cada $y \in \mathcal{H}$, la función $x \mapsto \langle x, y \rangle$ es lineal y continua en \mathcal{H} . Pero este teorema de representación de Riesz nos dice que, de hecho, no hay otras funciones lineales continuas:

Teorema 3. *Si $L : \mathcal{H} \rightarrow \mathbb{R}$ es lineal y continua, entonces existe un único $y \in \mathcal{H}$ tal que $L(x) = \langle x, y \rangle$.*

Demostración. Si $L(x) = 0$ para todo x , basta tomar $y = 0$. Para L no idénticamente 0, se define

$$\mathcal{G} = \{x : L(x) = 0\}.$$

Debido a la linealidad de L , \mathcal{G} es un subespacio, y debido a la continuidad de L es cerrado. Podemos usar entonces el Teorema 2, vemos que hay un $y \in \mathcal{G}^\perp$ tal que $L(x) = \langle x, y \rangle$. Sea $z \in \mathcal{G}^\perp$ con $\|z\| = 1$, observemos que

$$L(x) = L(x)\langle z, z \rangle = \langle L(x)z, z \rangle = \langle L(z)x + u(x), z \rangle, \text{ para } u(x) = L(x)z - L(z)x.$$

Observemos que $u(x) \in \mathcal{G}$. Dado que $z \in \mathcal{G}^\perp$, $\langle u(x), z \rangle = 0$. Por lo tanto,

$$L(x) = \langle L(z)x, z \rangle = \left\langle x, \overline{L(z)}z \right\rangle$$

Así, tomamos $y = \overline{L(z)}z$.

Para ver la unicidad, supongamos que $\langle x, y \rangle = \langle x, y' \rangle$ para cada x . Tomando $x = y - y'$, se tiene que $\|y - y'\| = 0$, por lo que $y = y'$. \square

A menudo, en FDA, se trabaja con proyecciones en subespacios de dimensión finita. Sea \mathcal{G} un subespacio de dimensión d en \mathcal{H} con una base e_1, \dots, e_d . Cada vector en \mathcal{G} es una combinación lineal única de e_1, \dots, e_d . En particular, para cada $x \in \mathcal{H}$, la aplicación P proyección que aparece en el Teorema 2 verifica:

$$P(x) = a_1 e_1 + a_2 e_2 + \dots + a_d e_d.$$

Según dicho teorema, se tiene que $x - P(x)$ es ortogonal a \mathcal{G} . Por tanto, para cada e_j , $\langle x - P(x), e_j \rangle = 0$, luego,

$$\langle e_1, e_j \rangle a_1 + \dots + \langle e_d, e_j \rangle a_d = \langle x, e_j \rangle, \quad j = 1, \dots, d.$$

Este es un sistema de d ecuaciones lineales en las incógnitas a_1, \dots, a_d , que puede determinarse de manera única si conocemos los productos internos $\langle e_i, e_j \rangle$ y $\langle x, e_j \rangle$. Si $\langle e_i, e_j \rangle = 0$, para $i \neq j$ y, además, $\langle e_j, e_j \rangle = 1$ entonces $a_j = \langle x, e_j \rangle$.

Si tomamos A un conjunto de índices arbitrario, decimos que $\{e_\alpha, \alpha \in A\}$ es un *sistema ortonormal* si

$$\langle e_\alpha, e_\beta \rangle = \begin{cases} 1 & \text{si } \alpha = \beta \\ 0 & \text{si } \alpha \neq \beta \end{cases}.$$

Espacios de Hilbert separables: Un espacio de Hilbert \mathcal{H} se dice que es *separable* si existe un sistema ortonormal contable $\{e_1, e_2, \dots\}$ tal que cada $x \in \mathcal{H}$ admite la expansión

$$x = \sum_{j=1}^{\infty} a_j e_j.$$

Ese sistema ortonormal contable se dice que es *completo* en \mathcal{H} . Más aún, un sistema ortonormal completo en un espacio de Hilbert separable suele llamarse *base ortonormal*.

Un conjunto $\{e_j, j \geq 1\}$ es una base ortonormal si no hay ningún elemento distinto de cero en \mathcal{H} que sea ortogonal a cada e_j . Se asume que la expansión converge en la norma de \mathcal{H} , es decir,

$$\lim_{J \rightarrow \infty} \left\| x - \sum_{j=1}^J a_j e_j \right\| = 0. \quad (4)$$

Por el Teorema 2, tenemos que

$$\|x\|^2 = \left\| \sum_{j=1}^J a_j e_j \right\|^2 + \left\| x - \sum_{j=1}^J a_j e_j \right\|^2.$$

La ortogonalidad de los e_j implica que

$$\left\| \sum_{j=1}^J a_j e_j \right\|^2 = \sum_{j=1}^J |a_j|^2,$$

y, junto con (4), implica que $\|x\|^2 = \sum_{j=1}^{\infty} |a_j|^2$. Nótese que $a_j = \langle x, e_j \rangle$.

Como resultado de todas estas afirmaciones, se tiene el siguiente teorema, donde la igualdad (5) se conoce como la igualdad de Parseval.

Teorema 4 (Teorema de Parseval). *Supongamos que \mathcal{H} es un espacio de Hilbert separable y $\{e_j\}$ es un sistema ortonormal completo. Entonces, para cualquier $x \in \mathcal{H}$,*

$$x = \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j$$

y

$$\|x\|^2 = \sum_{j=1}^{\infty} |\langle x, e_j \rangle|^2. \quad (5)$$

Retomando la definición de operador lineal, se tiene que los operadores acotados forman un espacio vectorial normado con las operaciones definidas puntualmente, es decir,

$$(L_1 + L_2)(x) = L_1(x) + L_2(x) \text{ y } (aL)(x) = aL(x).$$

Si suponemos ahora que \mathcal{H} es un espacio de Hilbert separable sobre el campo de los reales, la siguiente versión del teorema de Parseval en (5) es útil para definir los operadores de Hilbert-Schmidt, donde $\{e_j, j \geq 1\}$ sería cualquier base ortonormal en \mathcal{H} . Debido al Teorema 4 se tiene que

$$\begin{aligned} \langle x, y \rangle &= \left\langle \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j, \sum_{j'=1}^{\infty} \langle y, e_{j'} \rangle e_{j'} \right\rangle \\ &= \sum_{j, j'=1}^{\infty} \langle x, e_j \rangle \langle y, e_{j'} \rangle \langle e_j, e_{j'} \rangle = \sum_{j=1}^{\infty} \langle x, e_j \rangle \langle e_j, y \rangle, \end{aligned} \quad (6)$$

donde en la última igualdad se utiliza el hecho de que los e_j son ortonormales y que el producto interno es simétrico en el campo de los reales puesto que $\langle x, y \rangle = \langle y, x \rangle$.

Operador de Hilbert-Schmidt: Un operador lineal acotado Ψ se dice que es un *operador de Hilbert-Schmidt* si para una base ortonormal $\{e_j, j \geq 1\}$,

$$\|\Psi\|_{\mathcal{S}}^2 = \sum_{j=1}^{\infty} \|\Psi(e_j)\|^2 < \infty. \quad (7)$$

La relación anterior define la norma de Hilbert-Schmidt $\|\cdot\|_{\mathcal{S}}$, que no depende de la elección de la base. Denotamos el espacio vectorial de operadores de Hilbert-Schmidt por \mathcal{S} , que se puede ver que es un espacio de Hilbert separable con el

producto interno

$$\langle \Psi, \Phi \rangle_{\mathcal{S}} = \sum_{j=1}^{\infty} \langle \Psi(e_j), \Phi(e_j) \rangle. \quad (8)$$

Para ver que la norma $\| \cdot \|_{\mathcal{S}}$ está bien definida, supongamos que $\{e_j, j \geq 1\}$ y $\{f_i, i \geq 1\}$ son bases ortonormales, veamos que se tiene que

$$\sum_{i=1}^{\infty} \langle \Psi(f_i), \Phi(f_i) \rangle = \sum_{j=1}^{\infty} \langle \Psi(e_j), \Phi(e_j) \rangle.$$

Sabemos que $\{e_j, j \geq 1\}$ y $\{f_i, i \geq 1\}$ son bases ortonormales en \mathcal{H} y, por el Teorema 4 podemos escribir cualquier vector $x \in \mathcal{H}$ como:

$$x = \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j, \quad x = \sum_{i=1}^{\infty} \langle x, f_i \rangle f_i.$$

Así, podemos expresar Ψ y Φ como:

$$\Psi(f_i) = \sum_{j=1}^{\infty} \langle \Psi(f_i), e_j \rangle e_j, \quad \Phi(f_i) = \sum_{j=1}^{\infty} \langle \Phi(f_i), e_j \rangle e_j.$$

Ahora, usando (6), se obtiene:

$$\langle \Psi(f_i), \Phi(f_i) \rangle = \sum_{j=1}^{\infty} \langle \Psi(f_i), e_j \rangle \langle \Phi(f_i), e_j \rangle,$$

y, sumando sobre i :

$$\sum_{i=1}^{\infty} \langle \Psi(f_i), \Phi(f_i) \rangle = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \langle \Psi(f_i), e_j \rangle \langle \Phi(f_i), e_j \rangle.$$

Intercambiando el orden de la suma, que está justificado por (7) y, usando el hecho de que $\{f_i\}$ también es una base ortonormal, tenemos que

$$\langle \Psi(f_i), e_j \rangle = \langle \Psi(e_j), f_i \rangle \text{ y } \langle \Phi(f_i), e_j \rangle = \langle \Phi(e_j), f_i \rangle.$$

Sustituyendo estos valores en la expresión anterior se tiene que

$$\sum_{i=1}^{\infty} \langle \Psi(f_i), \Phi(f_i) \rangle = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} \langle \Psi(e_j), f_i \rangle \langle \Phi(e_j), f_i \rangle = \sum_{j=1}^{\infty} \langle \Psi(e_j), \Phi(e_j) \rangle,$$

donde la última igualdad se obtiene intercambiando nuevamente el orden de la suma y usando de nuevo (6).

Ejemplo (operador integral): Consideremos $\mathcal{H} = L^2$, y definamos un operador Ψ por

$$\Psi(x)(t) = \int \psi(t, s)x(s)ds,$$

donde $\psi(\cdot, \cdot)$ es una función bivalente llamada *núcleo* o *kernel*, que satisface

$$\int \int \psi^2(t, s)dtds < \infty.$$

Se puede ver que el operador Ψ es acotado. Esto puede ser probado introduciendo las funciones ψ_t definidas por $\psi_t(s) = \psi(t, s)$. Según el teorema de Fubini, $\psi_t \in L^2$ para casi todo t y para tales t , la desigualdad de Cauchy-Schwarz implica que para cualquier $x \in L^2$ se tenga

$$\left\{ \int \psi_t(s)x(s)ds \right\}^2 \leq \int \psi_t^2(s)ds \int x^2(s)ds.$$

Por lo tanto,

$$\begin{aligned} \|\Psi(x)\|^2 &= \int \left\{ \int \psi(t, s)x(s)ds \right\}^2 dt \leq \\ &\int \left\{ \int \psi^2(t, s)ds \int x^2(s)ds \right\} dt = \left(\int \int \psi^2(t, s)dtds \right) \|x\|^2. \end{aligned}$$

Esto implica que

$$\|\Psi\|_{\mathcal{L}} \leq \left(\int \int \psi^2(t, s)dtds \right)^{1/2}.$$

Para probar que $\Psi \in \mathcal{S}$, nos podemos fijar en que para cualquier base ortonormal $\{e_j, j \geq 1\}$, se tiene que

$$\begin{aligned} \sum_{j=1}^{\infty} \|\Psi(e_j)\|^2 &= \int \sum_{j=1}^{\infty} \left\{ \int \psi(t, s)e_j(s)ds \right\}^2 dt \\ &= \int \sum_{j=1}^{\infty} \langle \psi_t, e_j \rangle^2 dt = \int \|\psi_t\|^2 dt = \int \int \psi^2(t, s)dtds \end{aligned}$$

sin más que aplicar el Teorema 4 (Parseval). El cálculo anterior muestra además que

$$\|\Psi\|_{\mathcal{S}} = \left(\int \int \psi^2(t, s)dtds \right)^{1/2}. \quad (9)$$

2.2. Teoría espectral

Hablaremos ahora sobre algunos resultados de “teoría espectral de operadores” que resultan útiles para el FDA.

Supongamos que L es un operador en un espacio vectorial, decimos que un número λ es un *valor propio* o *autovalor* de L si existe un vector no nulo x en dicho espacio vectorial tal que

$$L(x) = \lambda x.$$

Cualquier vector x de este tipo se denomina *vector propio* o *autovector* de L correspondiente a λ . Si el operador actúa en un espacio de funciones, a menudo x es llamado *función propia* o *autofunción*. El conjunto de todos los autovectores correspondientes a λ forma un subespacio lineal cuya dimensión se conoce como la multiplicidad de λ .

Un operador L sobre un espacio vectorial de producto interno se dice que es *definido no negativo* si para cada vector x , se tiene que $\langle L(x), x \rangle \geq 0$, y *definido positivo* si la igualdad solo puede cumplirse para $x = 0$.

Teorema 5. *Todos los autovalores de un operador definido no negativo son no negativos y todos los autovalores de un operador definido positivo son positivos.*

Demostración. Si que L es un operador definido no negativo, λ su autovalor y x un autovector correspondiente, entonces,

$$\lambda \|x\|^2 = \lambda \langle x, x \rangle = \langle \lambda x, x \rangle = \langle L(x), x \rangle \geq 0$$

y por lo tanto, $\lambda \geq 0$. Como x es un autovector distinto de cero, entonces $\|x\| > 0$. Si L es definido positivo, entonces $\lambda \|x\|^2 > 0$, por lo que $\lambda > 0$. \square

Un operador L en un espacio con producto interno es *simétrico* o *autoadjunto* si, para cualquier x e y , se tiene que

$$\langle L(x), y \rangle = \langle x, L(y) \rangle.$$

Teorema 6. *Los autovectores correspondientes a autovalores distintos de un operador simétrico son ortogonales.*

Demostración. Supongamos que $\lambda_1 \neq \lambda_2$ son autovalores con autovectores x_1 y x_2 , respectivamente. Observemos que

$$\lambda_1 \langle x_1, x_2 \rangle = \langle L(x_1), x_2 \rangle = \langle x_1, L(x_2) \rangle = \lambda_2 \langle x_1, x_2 \rangle.$$

De esto se deduce que $\langle x_1, x_2 \rangle = 0$ ya que $\lambda_1 \neq \lambda_2$. \square

Teorema 7 (Teorema de Hilbert-Schmidt). *Supongamos que $\Psi \in \mathcal{S}$ es un operador simétrico. Entonces, existe un sistema ortonormal $\{e_j, j \geq 1\}$ formado por autovectores de Ψ , de tal manera que para todo $x \in \mathcal{H}$, se tiene la siguiente representación única:*

$$x = \sum_{j=1}^{\infty} a_j e_j + v, \text{ para } \Psi(v) = 0.$$

La demostración del Teorema 7 no es trivial y se puede encontrar en [1].

Corolario 1 (Teorema espectral). *Supongamos que $\Psi \in \mathcal{S}$ es un operador simétrico. Entonces, existe una base ortonormal $\{v_j, j \geq 1\}$ tal que para todo $x \in \mathcal{H}$,*

$$\Psi(x) = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j,$$

donde λ_j es un autovalor correspondiente a la autofunción v_j (de Ψ).

Demostración. Consideremos el conjunto

$$\mathcal{V} = \{x \in H : \Psi(x) = 0\}.$$

El subespacio \mathcal{V} es un espacio de Hilbert separable y sea $\{v_k^*, k \geq 1\}$ una base ortonormal en \mathcal{V} . La base requerida $\{v_j, j \geq 1\}$ es la unión de $\{e_j, j \geq 1\}$ y $\{v_k^*, k \geq 1\}$. El resultado se da porque $a_j = \langle x, e_j \rangle$ y cada v_k^* correspondería al autovalor cero. \square

Supongamos $\Psi \in \mathcal{S}$ un operador simétrico y definido no negativo y que queremos encontrar un vector $x \in \mathcal{H}$ con $\|x\| = 1$ tal que $\langle \Psi(x), x \rangle$ sea máximo. Dado que Ψ es definido no negativo, todos los autovalores λ_j son no negativos y los ordenamos en orden no creciente, $\lambda_1 \geq \lambda_2 \geq \dots$. Para garantizar una solución única, asumimos que $\lambda_1 > \lambda_2$ y debemos maximizar

$$\langle \Psi(x), x \rangle = \left\langle \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle v_j, x \right\rangle = \sum_{j=1}^{\infty} \lambda_j \langle x, v_j \rangle^2$$

sujeeto a

$$\sum_{j=1}^{\infty} \langle x, v_j \rangle^2 = 1.$$

El máximo claramente se alcanza si tomamos $x = v_1$ (o $x = -v_1$), luego, por el Teorema 6, v_1 es ortogonal a todos los otros autovectores. Hemos demostrado así que $\langle \Psi(x), x \rangle$ es máximo si $x = v_1$, el autovector correspondiente al autovalor más grande, y el máximo es igual a λ_1 , el autovalor más grande.

En FDA, es útil también resolver un segundo problema que se puede formular de la siguiente manera: “encontrar x con norma unitaria que maximice $\langle \Psi(x), x \rangle$, pero que sea ortogonal a v_1 ”. Veremos que $x = v_2$. Luego queremos maximizar $\langle \Psi(x), x \rangle$ en un subespacio ortogonal a v_1 y v_2 , etc. De forma más general:

Teorema 8. *Supongamos Ψ un operador de Hilbert-Schmidt simétrico y definido no negativo con autofunciones v_j y autovalores λ_j que satisfacen*

$$\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}.$$

Entonces, para $i \leq p$, se tiene que

$$\sup\{\langle \Psi(x), x \rangle : \|x\| = 1, \langle x, v_j \rangle = 0, 1 \leq j \leq i - 1\} = \lambda_i,$$

y el supremo se alcanza si $x = v_i$, con v_1, \dots, v_p las autofunciones del operador Ψ asociadas a los autovalores mayores $\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}$. La función x maximizadora es única salvo el signo.

Demostración. La prueba viene dada por el hecho de que:

$$\langle \Psi(x), x \rangle = \sum_{j=i}^{\infty} \lambda_j \langle x, v_j \rangle^2,$$

junto a que

$$\sum_{j=i}^{\infty} \langle x, v_j \rangle^2 = 1.$$

□

Todo anterior implica que, para un operador lineal simétrico definido no negativo con autovalores ordenados como $\lambda_1 \geq \lambda_2 \geq \dots$, tengamos las relaciones

$$\|\Psi\|_S^2 = \sum_{i=1}^{\infty} \lambda_i^2 \quad y \quad \|\Psi\|_L = \lambda_1.$$

Esto permite comprender un poco mejor la diferencia entre un operador acotado y uno Hilbert-Schmidt. Un operador acotado tiene autovalores acotados, mientras que un operador de Hilbert-Schmidt tiene autovalores que tienden lo suficientemente rápido a cero como para que la suma de autovalores sea finita.

2.3. Funciones aleatorias

Una colección \mathfrak{G} de subconjuntos del espacio métrico \mathcal{S} se denominan σ -álgebra si se cumple que:

- I) $\emptyset \in \mathfrak{G}$.
- II) $A \in \mathfrak{G} \Rightarrow A^c \in \mathfrak{G}$ (complemento).
- III) $A_j \in \mathfrak{G}, j \geq 1 \Rightarrow \bigcup_{j=1}^{\infty} A_j \in \mathfrak{G}$.

La σ -álgebra más pequeña que contiene conjuntos abiertos se llama la σ -álgebra de Borel, y sus miembros son los *conjuntos de Borel* o boherianos. Una función μ definida en \mathfrak{G} se llama *medida de probabilidad* si:

- I) Es no negativa.
- II) $\mu\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j)$ si ningún par de conjuntos A_j tiene una intersección no vacía (los A_j son mutuamente disjuntos).
- III) $\mu(\mathcal{S}) = 1$.

Supongamos que $(\Omega, \mathfrak{D}, P)$ es un espacio de probabilidad, donde Ω es el espacio muestral, \mathfrak{D} una σ -álgebra y P una medida de probabilidad. Sea \mathcal{S} un espacio métrico. Decimos que una asignación $X : \Omega \rightarrow \mathcal{S}$ es un *elemento aleatorio* en \mathcal{S} si $X^{-1}(A) \in \mathfrak{D}$ para cada conjunto de Borel A . Si $\mathcal{S} = \mathbb{R}$, X es una *variable aleatoria* y si $\mathcal{S} = \mathbb{R}^k$ entonces X es un *vector aleatorio*. Nuestro interés se centrará en el caso en el que \mathcal{S} sea un espacio de funciones, en cuyo caso X se denomina *función aleatoria*. Cada elemento aleatorio X induce una medida de probabilidad en \mathcal{S} , llamada *distribución de X* , y definida por

$$\mu(A) = P(X^{-1}(A)) = P(\omega \in \Omega : X(\omega) \in A) = P(X \in A).$$

Por simplificar, supondremos que todas las realizaciones $X(\omega)$ son elementos del espacio $L^2[0, 1]$, lo que significa que para cada $\omega \in \Omega$,

$$\|X(\omega)\|^2 = \int_0^1 \{X(\omega)(t)\}^2 dt < \infty.$$

Si nos fijamos en $\|X\| : \Omega \rightarrow \mathbb{R}$ tal que $\omega \mapsto \|X(\omega)\|$, tenemos que $\|X\|$ es una variable aleatoria tomando valores no negativos. Si $E\|X\|^2 < \infty$ diremos que nuestra función aleatoria X es de cuadrado integrable.

Supongamos ahora que las funciones aleatorias están definidas en un espacio de probabilidad común Ω y toman valores en un espacio de Hilbert separable \mathcal{H} con la distancia definida por una norma $\|\cdot\|$. Una función aleatoria X es *integrable* si $E\|X\| < \infty$ y débilmente integrable si, para cualquier $y \in H$, $E|\langle X, y \rangle| < \infty$ y si existe $e \in \mathcal{H}$ tal que $E[\langle X, y \rangle] = \langle e, y \rangle$. La función e es única y se denota por EX , el *valor esperado de X* . Por definición,

$$E[\langle X, y \rangle] = \langle EX, y \rangle, \quad \text{para todo } y \in \mathcal{H}. \quad (10)$$

Teorema 9. *Supongamos que X es integrable.*

I) $\|EX\| \leq E\|X\|$ (*propiedad contractiva*).

II) *Para cualquier operador acotado L , $L(X)$ es integrable y $E[L(X)] = L(EX)$.*

Demostración. Sea $e = EX$. Entonces,

$$|\langle e, y \rangle| = |E\langle X, y \rangle| \leq E|\langle X, y \rangle| \leq \|y\| E\|X\|.$$

Tomando $y = e$, obtenemos $\|e\|^2 \leq \|e\|E\|X\|$, por lo que se cumple la primera afirmación. Para demostrar la segunda, observemos que $L(X)$ es integrable porque

$$E\|L(X)\| \leq E\|L\|_{\mathcal{L}}\|X\| = \|L\|_{\mathcal{L}}E\|X\| < \infty.$$

Definimos $e' = L(e)$. La igualdad se sigue si para cada $y \in H$, $\langle e', y \rangle = \langle L(e), y \rangle$. Se cumple, por la definición de operador autoadjunto L^* , $\langle L(x), y \rangle = \langle x, L^*(y) \rangle$, que

$$\langle L(e), y \rangle = \langle e, L^*(y) \rangle = \mathbb{E}[\langle X, L^*(y) \rangle] = \mathbb{E}[\langle L(X), y \rangle] = \langle e', y \rangle.$$

□

En el FDA contaremos de partida con n funciones observadas $\{x_i\}_{i=1}^n$. Supondremos que $x_i \in L^2[0, 1]$, el espacio de las funciones de cuadrado integrable en el intervalo $[0, 1]$, y que las curvas $\{x_i\}_{i=1}^n$ son realizaciones independientes e igualmente distribuidas de la función aleatoria X .

2.4. Operador de covarianza y expansión de Karhunen-Loève

Trabajando con las funciones aleatorias con valores en un espacio de Hilbert, el *operador de covarianza* se define como $C : \mathcal{H} \rightarrow \mathcal{H}$ tal que

$$y \mapsto C(y) = E[\langle X - EX, y \rangle (X - EX)].$$

A menudo se asume $EX = 0$, de modo que se tiene $C(y) = E[\langle X, y \rangle X]$.

Si $\mathcal{H} = \mathbb{R}$, entonces $C(y) = E[XyX] = \text{Var}[X]y$, por lo que la acción de C se reduce a la multiplicación por $\text{Var}[X]$, y C se puede identificar con la varianza. Si $\mathcal{H} = \mathbb{R}^d$, C se puede identificar de manera similar con la matriz de covarianza.

Si $\mathcal{H} = L^2[0, 1]$, entonces $C(y)$ es una función en $L^2[0, 1]$ cuyo valor en $t \in [0, 1]$ es

$$C(y)(t) = E \left[\left(\int_0^1 X(s)y(s)ds \right) X(t) \right] = \int_0^1 c(t, s)y(s)ds, \quad (11)$$

donde

$$c(t, s) = E[X(t)X(s)], \quad \text{para } s, t \in [0, 1]. \quad (12)$$

Vemos que el operador de covarianza C en (11) es un operador con núcleo $c(t, s)$, llamado *función de covarianza*. La ecuación (11) muestra que el operador de covarianza se puede identificar con la función de covarianza correspondiente (12), que en realidad se usa con más frecuencia en el FDA.

Teorema 10. *Supongamos que $C : \mathcal{H} \rightarrow \mathcal{H}$ es un operador lineal. El operador C es el operador de covarianza de alguna función aleatoria cuadrado integrable X si y solo si*

I) C es simétrico.

II) C es no negativo definido.

III) Los autovalores de C satisfacen $\sum_{j=1}^{\infty} \lambda_j < \infty$.

Demostración. Veamos que si C es un operador de covarianza entonces se verifican I), II), y III). Por (10), se tiene que

$$\langle C(y), z \rangle = \langle E[\langle X, y \rangle X], z \rangle = E[\langle \langle X, y \rangle X, z \rangle] = E[\langle X, y \rangle \langle X, z \rangle]. \quad (13)$$

Esto muestra que $\langle C(y), z \rangle = \langle y, C(z) \rangle$. Por (13), se tendría que $\langle C(y), y \rangle = E[\langle X, y \rangle^2] \geq 0$.

Supongamos ahora que v_j , $j \geq 1$, son las autofunciones de C . De nuevo por (13), se llega a

$$E[\langle X, v_j \rangle^2] = \langle C(v_j), v_j \rangle = \langle \lambda_j v_j, v_j \rangle = \lambda_j.$$

Por el Teorema 4, se tiene que

$$\sum_{j=1}^{\infty} \lambda_j = \sum_{j=1}^{\infty} E[\langle X, v_j \rangle^2] = E \sum_{j=1}^{\infty} \langle X, v_j \rangle^2 = E\|X\|^2 < \infty.$$

El recíproco también es cierto en el sentido de que si C es un operador verificando las tres propiedades I), II) y III) en el enunciado entonces es operador de covarianza para alguna función aleatoria X de cuadrado integrable, la prueba de esto es más compleja y puede encontrarse en [6]. \square

Fijemos ahora $p \geq 1$ y supongamos que u_1, u_2, \dots, u_p es un sistema ortonormal en $L^2[0, 1]$. La proyección de una función aleatoria X en el subespacio generado por este sistema será

$$\sum_{k=1}^p \langle X, u_k \rangle u_k.$$

El sistema u_1, u_2, \dots, u_p es óptimo en el sentido de minimizar la distancia esperada entre X y su aproximación en el subespacio generado por u_1, \dots, u_p si minimiza

$$S(u_1, u_2, \dots, u_p) = E \left\| X - \sum_{k=1}^p \langle X, u_k \rangle u_k \right\|^2 \quad (14)$$

sobre todos los sistemas ortonormales que consistan en p funciones. Para encontrar dicho sistema, observemos que, debido a la ortonormalidad de los u_k , podemos escribir

$$\begin{aligned} S(u_1, u_2, \dots, u_p) &= E \left\langle X - \sum_{k=1}^p \langle X, u_k \rangle u_k, X - \sum_{k=1}^p \langle X, u_k \rangle u_k \right\rangle \\ &= E \|X\|^2 - 2E \sum_{k=1}^p \langle X, u_k \rangle^2 + E \sum_{k=1}^p \langle X, u_k \rangle^2 \\ &= E \|X\|^2 - \sum_{k=1}^p E \langle X, u_k \rangle^2. \end{aligned}$$

Dado que, por la ecuación (10),

$$E \langle X, u_k \rangle^2 = E \langle \langle X, u_k \rangle X, u_k \rangle = \langle C(u_k), u_k \rangle,$$

minimizar $S(u_1, u_2, \dots, u_p)$ es equivalente a maximizar

$$\sum_{k=1}^p \langle C(u_k), u_k \rangle.$$

Ahora supongamos que se cumple $\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}$. Según el Teorema 8, el valor máximo posible de $\langle C(u_1), u_1 \rangle$ es λ_1 y se alcanza si $u_1 = v_1$.

Posteriormente, como u_2 debe ser ortogonal a $u_1 = v_1$, el valor máximo posible de $\langle C(u_2), u_2 \rangle$ es λ_2 y se alcanza en $u_2 = v_2$. Continuando de esta manera, establecemos el siguiente resultado:

Teorema 11. *Supongamos que X es cuadrado integrable y que los autovalores de su operador de covarianza C cumplen $\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}$. Para cualquier p fijo, el error esperado (14) se minimiza si $u_j = v_j$, donde los v_j son las autofunciones de C con norma unitaria asociadas a estos autovalores $\lambda_1, \dots, \lambda_p$.*

Estableciendo $\mu = EX$, se observa que

$$X - \mu = \sum_{j=1}^{\infty} \xi_j v_j, \quad \text{para } \xi_j = \langle X - \mu, v_j \rangle. \quad (15)$$

Esta descomposición es denominada *expansión de Karhunen-Loève* y las funciones v_j serán las *Componentes Principales Funcionales*. Las Componentes Principales Funcionales son, por lo tanto, las autofunciones del operador de covarianza normalizadas para tener norma unitaria. Las funciones v_j forman una base ortonormal y las variables aleatorias ξ_j suelen denominarse *scores*. Se puede demostrar que

$$E[\xi_j] = 0, \quad E[\xi_j^2] = \lambda_j \quad \text{y} \quad E[\xi_j \xi_i] = 0, \quad \text{si } i \neq j. \quad (16)$$

Por lo tanto, mediante la identidad de Parseval del Teorema 4,

$$E\|X - \mu\|^2 = \sum_{j=1}^{\infty} \lambda_j. \quad (17)$$

La identidad (17) tiene la siguiente interpretación: “ λ_j es la varianza de la función aleatoria X en la dirección principal v_j ” y “la varianza de X es igual a la suma de las varianzas de las proyecciones de X en las Componentes Principales Funcionales”. La relación (17) es una *descomposición ANOVA* (Analysis of Variance) que se utiliza con mucha frecuencia en FDA.

En el espacio $L^2[0, 1]$, las descomposiciones (15) y (17) toman la forma

$$X(t) = \mu(t) + \sum_{j=1}^{\infty} \xi_j v_j(t), \quad E \int_0^1 (X(t) - \mu(t))^2 dt = \sum_{j=1}^{\infty} \lambda_j. \quad (18)$$

Vemos que cada función aleatoria cuadrado integrable es la suma de la función de media $\mu(\cdot)$ y una función de error aleatoria cuyo valor esperado es cero y que

es suma de las componentes principales funcionales $v_j(\cdot)$ con pesos aleatorios ξ_j . La importancia de cada componente principal funcional v_j se cuantifica por la desviación estándar de su peso ξ_j , que es igual a $\sqrt{\lambda_j}$. El valor esperado del error cuadrático integrado entre una función aleatoria y su función de media es igual a la suma de los autovalores del operador de covarianza.

Ejemplo (Movimiento Browniano): Una función aleatoria

$$W = \{W(t), t \in [0, 1]\}$$

se denomina *proceso de Wiener* o *movimiento Browniano* si se cumplen las siguientes condiciones:

- I) $W(0) = 0$.
- II) Si $0 \leq s < t \leq 1$, entonces $W(t) - W(s)$ es una variable normal con media cero y varianza $t - s$.
- III) Para cualquier $0 \leq t_0 < t_1 < \dots < t_k \leq 1$, las variables aleatorias

$$W(t_j) - W(t_{j-1}),$$

para $1 \leq j \leq k$, son independientes.

La forma habitual de aproximar un movimiento Browniano se basa en la observación de que para $0 \leq k \leq K$

$$W\left(\frac{k}{K}\right) - W\left(\frac{k-1}{K}\right) = N\left(0, \frac{1}{K}\right) = \frac{1}{\sqrt{K}}N_k,$$

así que, con variables normales estándar independientes N_k ,

$$W\left(\frac{k}{K}\right) = \frac{1}{\sqrt{K}} \sum_{i=1}^k N_k, \quad N_k \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1), \quad \text{para } 0 \leq k \leq K.$$

Se tiene que

$$\mu(t) = \mathbb{E}[W(t)] = 0, \text{ y } c(t, s) = \text{Cov}(W(t), W(s)) = \min(t, s).$$

Por tanto,

$$\lambda_j = \frac{1}{(j - \frac{1}{2})^2 \pi^2}, \text{ y } v_j(t) = \sqrt{2} \sin\left(\left(j - \frac{1}{2}\right) \pi t\right), \quad j = 1, 2, \dots$$

Para comprobar esto, debemos resolver la ecuación integral

$$\int_0^1 \min(s, t)v(s) ds = \lambda v(t),$$

que puede escribirse como

$$\int_0^t sv(s) ds + t \int_t^1 v(s) ds = \lambda v(t).$$

Derivando respecto a t , se obtiene que

$$\int_0^t v(s) ds = \lambda v'(t).$$

Derivando nuevamente, obtenemos la ecuación diferencial

$$-v(t) = \lambda v''(t).$$

Una solución general para esta ecuación tiene la forma

$$v(t) = A \sin\left(\frac{t}{\sqrt{\lambda}}\right) + B \cos\left(\frac{t}{\sqrt{\lambda}}\right).$$

Las constantes A y B se determinan mediante condiciones de contorno. Dado que:

$$v(0) = 0, \text{ entonces } v(0) = B = 0.$$

Al establecer $t = 1$, obtenemos $v'(1) = 0$, lo que implica que

$$\cos\left(\frac{1}{\sqrt{\lambda}}\right) = 0.$$

Esto significa que $1/\sqrt{\lambda} = -\pi/2 + j\pi$, para $j \geq 1$, por lo que todos los posibles autovalores son como se especifica en la afirmación, y las correspondientes autofunciones son

$$A \sin(\pi(j - 1/2)t).$$

La condición $\int v_j^2(t) dt = 1$ implica $A = \sqrt{2}$.

La expansión (15) se convierte en

$$W(t) = \sum_{j=1}^{\infty} \frac{N_j}{(j - \frac{1}{2})\pi} \sqrt{2} \sin\left(\left(j - \frac{1}{2}\right) \pi t\right) \quad (19)$$

donde los N_j son las variables aleatorias estándar normales ($\mathbb{E}(N_j) = 1$ y $\text{Var}(N_j) = 1$) e independientes. Los scores ξ_j se dan por

$$\xi_j = \frac{N_j}{(j - \frac{1}{2})\pi}$$

La descomposición de la varianza (18) se convierte, por tanto, en

$$\mathbb{E} \int_0^1 W^2(t) dt = \int_0^1 t dt = \frac{1}{2} = \sum_{j=1}^{\infty} \frac{1}{(j - \frac{1}{2})^2 \pi^2}.$$

La expansión (19) ofrece una forma de generar aproximaciones suaves al movimiento Browniano de la forma:

$$\sum_{j=1}^J \frac{N_j}{(j - \frac{1}{2})\pi} \sqrt{2} \sin \left(\left(j - \frac{1}{2} \right) \pi t \right).$$

con $J \in \mathbb{N}$.

La expansión muestra que el movimiento Browniano es una combinación lineal de funciones ortonormales v_j sinusoidales con scores que son variables aleatorias normales.

En general, se dice que una función aleatoria X es gaussiana si los scores ξ_j en (15) son variables aleatorias gaussianas y admiten en $L^2[0, 1]$ una descomposición de la forma

$$X(t) = \mu(t) + \sum_{j=1}^{\infty} \sqrt{\lambda_j} N_j v_j(t), \quad \text{y} \quad \lambda_j = E \left\{ \int (X(t) - \mu(t)) v_j(t) dt \right\}^2.$$

2.5. Representación en bases funcionales

Un primer paso bastante habitual al trabajar con datos funcionales de la forma (2) es expresarlos mediante su aproximación usando bases funcionales:

$$x(t) \simeq \sum_{m=1}^M b_m B_m(t) = b' B(t), \quad \text{para} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix}, \quad B(t) = \begin{bmatrix} B_1(t) \\ B_2(t) \\ \vdots \\ B_M(t) \end{bmatrix}, \quad (20)$$

donde B_1, B_2, \dots, B_M son una base funcional.

La representación (20) refleja la intuición de que los datos (2) son observaciones de funciones que comparten algunas propiedades de forma, y por lo tanto, pueden aproximarse como combinaciones lineales de M formas básicas $\{B_m\}_{m=1}^M$, siendo M típicamente menor que el número de puntos de tiempo observados, J_n .

Si el número de puntos t_{ij} es muy grande, la expansión (20) sirve para reemplazar prácticamente los datos escalares originales por una colección más pequeña en los coeficientes b_m . Si los puntos de tiempo t_{ij} difieren entre sujetos, la expansión coloca las curvas en un dominio común (M -dimensional), para que sean más fácilmente comparables.

Una vez seleccionada una base, cada observación puede estar sujeta a errores aleatorios, y es necesaria la formulación de un problema de mínimos cuadrados para obtener un vector $\mathbf{b}_i = (b_{i1}, \dots, b_{iM})'$ que represente a la curva x_i . Es decir,

$$x_i(t_{ij}) = \sum_{m=1}^M b_{im} B_m(t_{ij}) + \varepsilon_{ij} \quad i = 1, \dots, n, \quad j = 1, \dots, J_i.$$

donde ε_{ij} son errores aleatorios cuyo valor esperado es cero. El objetivo es minimizar la suma de los cuadrados de las diferencias entre valores observados y valores predichos. Por lo tanto, la función de error cuadrático para este caso sería:

$$\text{SSE} = \sum_{i=1}^n \sum_{j=1}^{J_i} [x_i(t_{ij}) - \sum_{m=1}^M b_{im} B_m(t_{ij})]^2.$$

Si lo reescribimos de forma matricial, tenemos que:

$$\mathbf{x}_i = \mathbf{B}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad \mathbf{x}_i \in \mathcal{M}_{J_i \times 1}, \quad \mathbf{B}_i \in \mathcal{M}_{J_i \times M}, \quad \mathbf{b}_i \in \mathcal{M}_{M \times 1}, \quad \boldsymbol{\varepsilon}_i \in \mathcal{M}_{J_i \times 1} \quad (21)$$

para expresar

$$\begin{bmatrix} x_i(t_{i1}) \\ x_i(t_{i2}) \\ \vdots \\ x_i(t_{iJ_i}) \end{bmatrix} = \begin{bmatrix} B_1(t_{i1}) & B_2(t_{i1}) & \dots & B_M(t_{i1}) \\ B_1(t_{i2}) & B_2(t_{i2}) & \dots & B_M(t_{i2}) \\ \vdots & \vdots & \ddots & \vdots \\ B_1(t_{iJ_i}) & B_2(t_{iJ_i}) & \dots & B_M(t_{iJ_i}) \end{bmatrix} \begin{bmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{Mi} \end{bmatrix} + \begin{bmatrix} \xi_{i1} \\ \xi_{i2} \\ \vdots \\ \xi_{iJ_i} \end{bmatrix}.$$

Nótese que usamos la notación $\mathcal{M}_{a \times b}$ para referirnos a matrices con a filas y b columnas. Entonces, la formulación matricial del problema de mínimos cuadrados sería:

$$\text{SSE} = \sum_{i=1}^n (\mathbf{x}_i - \mathbf{B}_i \mathbf{b}_i)' (\mathbf{x}_i - \mathbf{B}_i \mathbf{b}_i) = \sum_{i=1}^n \|\mathbf{x}_i - \mathbf{B}_i \mathbf{b}_i\|^2. \quad (22)$$

Derivando SSE con respecto a b_i se obtiene la ecuación

$$2\mathbf{B}'_i\mathbf{B}_i\mathbf{b}_i - 2\mathbf{B}'_i\mathbf{x}_i = 0,$$

es decir,

$$\mathbf{B}'_i\mathbf{B}_i\mathbf{b}_i = \mathbf{B}'_i\mathbf{x}_i,$$

y resolver esto para \mathbf{b}_i proporciona la estimación que minimiza la solución de mínimos cuadrados:

$$\mathbf{b}_i = (\mathbf{B}'_i\mathbf{B}_i)^{-1}\mathbf{B}'_i\mathbf{x}_i \in \mathbb{R}^M.$$

La aproximación simple de mínimos cuadrados es apropiada en situaciones donde asumimos que los residuos ε_{ij} alrededor de la verdadera curva son independientes e idénticamente distribuidos con media cero y varianza constante σ^2 . Obviamente, este proceso se repite para cada función x_1, \dots, x_n , obteniendo los vectores de coeficientes \mathbf{b}_i que minimizan la suma de los cuadrados de los errores para cada función.

Deducir la elección adecuada de M en (20) y el tipo de bases es esencial, y no existe una forma de hacerlo universal que se ajuste razonablemente a todos los conjuntos de datos. Diferentes tipos de bases se adaptan mejor a conjuntos de datos con comportamientos específicos.

Entre las bases comúnmente utilizadas se encuentran las bases de Fourier, ideales para datos periódicos, y los sistemas de bases B-spline, eficaces para datos no periódicos exhibiendo cierta suavidad. Mostraremos estas bases, aunque las wavelets, que son igualmente populares, no se detallarán en esta memoria.

Bases de Fourier: La *base de Fourier* es la más utilizada cuando las funciones presentan cierta periodicidad. Se fundamenta en que las series de Fourier están basadas en funciones “seno” y “coseno” de la siguiente forma:

$$B_0(t) = 1 \quad B_{2k-1} = \sin\left(\frac{k}{\omega}t\right) \quad B_{2k} = \cos\left(\frac{k}{\omega}t\right) \quad (23)$$

para

$$k = 1, \dots, \frac{M-1}{2},$$

donde el parámetro ω determina el periodo. Nótese que en el caso de la base de Forier, se tiene que $\mathbf{B}'_i\mathbf{B}_i = \mathbf{I}_M$. Concretamente, dicha propiedad se tiene para

cualquier base ortonormal y, por tanto, en este caso se obtiene

$$\mathbf{b}_i = \begin{bmatrix} \langle B_1, x_i \rangle \\ \langle B_2, x_i \rangle \\ \vdots \\ \langle B_M, x_i \rangle \end{bmatrix}.$$

En R, la librería `fda` [10] (Functional Data Analysis) tiene acceso a diversas funciones y herramientas que facilitan el análisis de los datos funcionales. La librería `fda.usc` [2] (Functional Data Analysis and Utilities for Statistical Computing), desarrollada por investigadores esencialmente de la Universidad de Santiago de Compostela, realiza análisis exploratorio y descriptivo, ayuda a explicar y modelar la relación entre una variable dependiente e independiente, realiza predicciones, incluye métodos para la clasificación supervisada o no supervisada y lleva a cabo análisis de modelos de varianza para datos funcionales.

```
install.packages("fda")
install.packages("fda.usc")
library(fda)
library(fda.usc)
```

La función `create.fourier.basis` sirve para crear una base de Fourier, como puede verse en la Figura 2. El argumento `rangeval` es un vector que contiene los valores inicial y final del intervalo sobre el cual el objeto de datos funcionales puede evaluarse y `nbasis` es un número de funciones de base, que en este caso, es igual a 5.

```
baseFourier <- create.fourier.basis(rangeval=c(0,1), nbasis=5)
plot(baseFourier)
```

Bases de B-splines: Se ha utilizado la referencia [14] como base para este apartado. Supongamos que contamos con puntos $\{(t_j, y_j)\}_{j=1}^J$, donde los valores y_j observados en el caso de considerar una función x serán

$$y_j = x(t_j) + \varepsilon_j,$$

donde ε_j es un error aleatorio cuyo valor esperado es cero y varianza constante σ^2 . Nuestra idea será aplicar esto mismo a las discretizaciones de nuestras funciones x_i , para $i = 1, \dots, n$, en nuestro conjunto de datos $\{(t_{ij}, y_{ij})\}_{j=1}^{J_i}$ para $i = 1, \dots, n$.

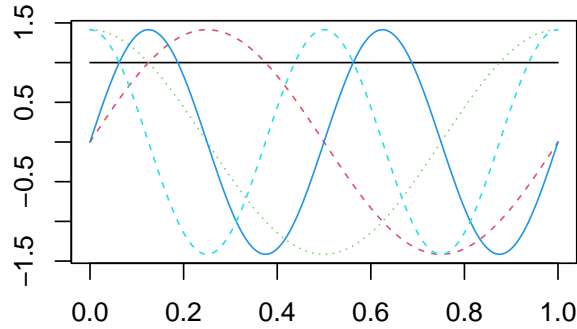


Figura 2: Sistema de 5 bases de Fourier en el intervalo $[0,1]$.

Supongamos que estimamos x eligiendo \hat{x} igual a la función f que minimiza la suma de cuadrados

$$\sum_{j=1}^J (y_j - f(t_j))^2,$$

es decir, conseguir el máximo “ajuste” a los datos. Minimizar sobre todas las funciones lineales del tipo $f(t) = a+bt$ da como resultado el estimador de regresión lineal mínimo cuadrático y minimizar respecto a todas las funciones posibles sin ninguna restricción da como resultado una función f que “interpola” los datos. Decimos que una función interpola exactamente los datos si $y_j = f(t_j)$ para todo $j = 1, \dots, J$.

Una forma de obtener soluciones intermedias es minimizar sumas de cuadrados penalizadas del tipo

$$f \mapsto M_\lambda(f) = \sum_{j=1}^J (y_j - f(t_j))^2 + \lambda J(f), \quad (24)$$

donde $J(f)$ es una penalización por “rugosidad” tal que $J(f)$ toma un mayor valor si la función f es menos “suave”. Nos centramos en el caso especial de

$$J(f) = \int (f''(t))^2 dt, \quad (25)$$

donde f'' denota la derivada segunda de la función f . El parámetro λ controla el equilibrio entre “ajuste” y “suavidad” en (24).

Sea $f = \hat{x}$ la función que minimiza $f \mapsto M_\lambda(f)$. Cuando $\lambda = 0$, la solución \hat{x} es la función interpolante y \hat{x} converge a la recta de mínimos cuadrados cuando $\lambda \rightarrow \infty$. El parámetro λ controla el nivel de suavizado y la cuestión es cómo

obtener \hat{x} para $0 < \lambda < \infty$. Para ello, conviene definir primero los splines, que serán polinomios definidos a trozos tal como se explicará a continuación.

Sea $\xi_1 < \xi_2 < \dots < \xi_k$ un conjunto de puntos ordenados, llamados *nodos*, contenidos en un intervalo (a, b) . Un spline cúbico es una función continua formada por polinomios cúbicos a trozos sobre los intervalos $(\xi_1, \xi_2), \dots, (\xi_{k-1}, \xi_k)$ y que tengan derivadas primera y segunda continuas en los nodos. De forma general, un spline de orden L es un polinomio a trozos de grado $L - 1$ con $L - 2$ derivadas continuas en los nodos y un spline que es lineal más allá de los nodos de frontera se llama spline natural.

Los splines cúbicos (con $L = 4$) son los splines más comúnmente utilizados en la práctica. La función \hat{x} que minimiza $M_\lambda(f)$ con la penalización sobre la derivada segunda en (25) es un spline cúbico natural con nodos en $\{t_j\}_{j=1}^J$. El estimador \hat{x} asociado a una penalización λ fija se llama “spline de suavizado”.

Para proporcionar una representación explícita para \hat{x} , construiremos una base para el conjunto de splines. Dados los nodos $\xi_1 < \xi_2 < \dots < \xi_k$ en el intervalo (a, b) se definen

$$H_1(t) = 1, \quad H_2(t) = t, \quad H_3(t) = t^2, \quad H_4(t) = t^3 \quad \text{y} \quad H_m(t) = (t - \xi_{m-4})_+^3,$$

para $m = 5, \dots, k + 4$. Estas funciones $\{H_1, \dots, H_{k+4}\}$ forman una base para el conjunto de splines cúbicos en estos nodos, llamada la base de potencias truncadas. Así, cualquier spline cúbico con estos nodos puede escribirse como

$$f(t) = \sum_{m=1}^{k+4} b_m H_m(t).$$

Ahora introducimos una base diferente para el conjunto de splines naturales llamada *base B-splines*, que es particularmente adecuada para cálculos computacionales. Sea $\xi_0 = a$ y $\xi_{k+1} = b$, introducimos nuevos nodos τ_1, \dots, τ_L de tal manera que

$$\begin{aligned} \tau_1 &\leq \tau_2 \leq \dots \leq \tau_L \leq \xi_0, \\ \tau_{r+L} &= \xi_j \quad \text{para } r = 1, \dots, k, \end{aligned}$$

y $\tau_{K+L+1}, \dots, \tau_{K+2L}$ tales que

$$\xi_{k+1} \leq \tau_{k+L+1} \leq \dots \leq \tau_{k+2L}.$$

La elección de estos nodos adicionales es arbitraria aunque generalmente, se toma $\tau_1 = \dots = \tau_L = \xi_0$ y $\xi_{k+1} = \tau_{k+L+1} = \dots = \tau_{k+2L}$.

Así, de forma recursiva, definimos

$$B_{m,1}(t) = \begin{cases} 1 & \text{si } \tau_m \leq t < \tau_{m+1} \\ 0 & \text{en otro caso} \end{cases}$$

para $m = 1, \dots, k + 2L - 1$ y, posteriormente, para $l \leq L$, definimos por recursividad

$$B_{m,l}(t) = \frac{t - \tau_m}{\tau_{m+l-1} - \tau_m} B_{m,l-1}(t) + \frac{\tau_{m+l} - t}{\tau_{m+l} - \tau_{m+1}} B_{m+1,l-1}(t),$$

para $m = 1, \dots, k + 2L - 1$.

Se entiende que si el denominador es 0, entonces la función se define como 0. Las funciones $\{B_{m,4}, m = 1, \dots, k + 4\}$ forman una base para el conjunto de splines cúbicos, denominadas *funciones de base B-splines*. La ventaja de las funciones de base B-splines es que tienen soporte compacto en el sentido de que son nulos cuando nos alejamos de intervalos asociados a nodos contiguos a t , lo que facilita su interpretación y reduce tiempos de cálculo.

Una vez creadas estas bases podemos definir los estimadores tipo spline para $\{(t_{ij}, x(t_{ij}))\}_{j=1}^{J_i}$ cuando $i = 1, \dots, n$. Este estimador será escrito de la forma

$$\hat{x}_i(t) = \sum_{m=1}^M b_{mi} B_m(t),$$

donde B_1, B_2, \dots, B_M son una base de splines naturales (por ejemplo, B-splines con $M = J+4$ si $J_i = J$ para todo $i = 1, \dots, n$). Unos $\{b_{mi}\}_{m=1}^M$ razonables pueden obtenerse mediante minimización cuadrática con una penalización del tipo a la que aparece en (24) donde interviene una matriz con términos

$$\Omega_{mm'} = \int B_m''(t) B_{m'}''(t) dt,$$

para $\{m, m'\} \subset \{1, \dots, M\}$. Hay una fórmula cerrada para obtener dichos coeficientes

$$\mathbf{b}_i = \begin{bmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{Mi} \end{bmatrix},$$

de la forma

$$\mathbf{b}_i = (\mathbf{B}_i' \mathbf{B}_i + \lambda \Omega)^{-1} \mathbf{B}_i' \mathbf{x}_i,$$

donde \mathbf{B}_i y \mathbf{x}_i son matrices y vectores que coinciden con los usados anteriormente. El λ óptimo se determina habitualmente por técnicas de validación cruzada.

No obstante, es incluso más común trabajar directamente sin usar penalizaciones, y simplemente fijar M como el número de elementos de la base de B-splines a considerar y aplicar mínimos cuadrados tal como se hacía en (22).

La función `create.bspline.basis` crea una base B-spline usando el paquete `fda` en R, como puede verse en la Figura 3. El argumento `nbasis` especifica el número de bases de B-splines, que se ignora si se suministra `breaks`, un vector que especifica los nodos que definen la base de B-splines, en cuyo caso `nbasis = nbreaks + norder - 2`, donde ($k =$) `nbreaks = length(breaks)` y `norder` es un entero que especifica el orden de los B-splines, que es uno más alto que su grado, y el valor predeterminado de 4 ($L = 4$) proporciona los splines cúbicos.

```

nodos <- seq(0, 1, length=11)
baseBspline <- create.bspline.basis(rangeval=c(0,1), nbasis=13,
                                   breaks=nodos)
plot(baseBspline, lty=1)

```

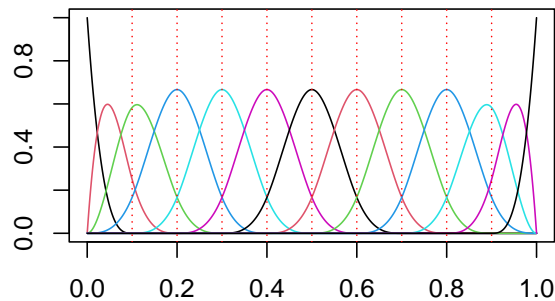


Figura 3: Sistema de 13 bases B-spline en el intervalo $[0,1]$.

Si retomamos el ejemplo del conjunto de datos AEMET mostrado en la Figura 1(b), el analizar los datos brutos (tomados sin ninguna transformación o limpieza de los mismos) puede dificultar considerablemente la realización de cualquier estudio estadístico por su “rugosidad”. Por lo tanto, un primer paso a llevar a cabo puede ser un suavizado de los datos para trabajar con “tendencias” que describan la evolución anual promedio de las temperaturas y con el objetivo de eliminar, en la medida de lo posible, alteraciones puntuales sin comprometer la preservación de características fundamentales de esas evoluciones.

En la Figura 4, se muestran las curvas suavizadas de los promedios de las temperaturas usando la función `create.bspline.basis`, donde, en cada caso, se utiliza un número diferente de elementos de la base B-spline.

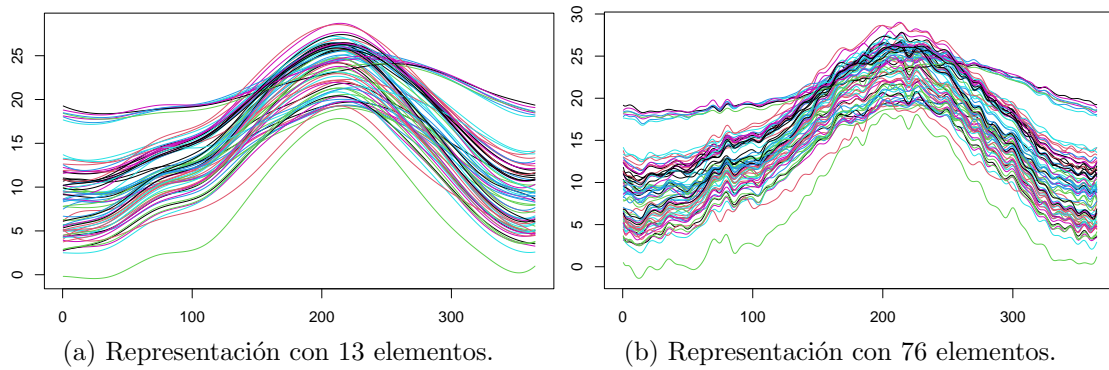


Figura 4: Representación de los datos AEMET usando bases de B-splines con diferentes números de elementos.

3. Componentes principales funcionales empíricas

Si observamos curvas x_1, x_2, \dots, x_n en $L^2[0, 1]$ como realizaciones de una función aleatoria X de cuadrado integrable, entonces la *función de media muestral* $\hat{\mu}(t)$ y la *función de covarianza muestral* $\hat{c}(t, s)$ se definen como

$$\hat{\mu}(t) = \frac{1}{n} \sum_{i=1}^n x_i(t)$$

y

$$\hat{c}(t, s) = \frac{1}{n} \sum_{i=1}^n (x_i(t) - \hat{\mu}(t))(x_i(s) - \hat{\mu}(s)),$$

respectivamente.

Las funciones muestrales $\hat{\mu}$ y \hat{c} se pueden ver como estimadores de las funciones poblacionales μ y c que surgen de una función aleatoria X cuyo muestreo ha dado lugar a dichas curvas x_1, \dots, x_n de tal forma que

$$\mu(t) = EX(t),$$

y

$$c(t, s) = E[(X(t) - \mu(t))(X(s) - \mu(s))].$$

Cuando n es grande, $\hat{\mu}(\cdot)$ y $\hat{c}(\cdot, \cdot)$ son buenos estimadores de $\mu(\cdot)$ y $c(\cdot, \cdot)$ debido a que, al aumentar el tamaño de la muestra, se obtiene una mejor aproximación de las características promedio y de variabilidad de las curvas observadas. Existen resultados que prueban esta última afirmación que no serán tratados en esta memoria.

Partiendo de estas estimaciones $\hat{\mu}(\cdot)$ y $\hat{c}(\cdot, \cdot)$ podemos pensar en buscar estimadores de las componentes principales funcionales vistas en la Sección 2.4. Con esto en mente, definimos las *Componentes Principales Funcionales Estimadas o Empíricas* por

$$\int \hat{c}(t, s) \hat{v}_j(s) ds = \hat{\lambda}_j \hat{v}_j(t), \quad \text{para } j = 1, 2, \dots, n. \quad (26)$$

Estos $\hat{\lambda}_j$ y $\hat{v}_j(t)$ son los estimadores empíricos de los v_j y λ_j teóricos presentados en (15) y con una interpretación análoga.

Las Componentes Principales Funcionales teóricas, que fueron definidas como las autofunciones ortogonales del operador de covarianza C en la Sección 2.4, deben abordarse con precaución. Las autofunciones v_j se definen por

$$C(v_j) = \lambda_j v_j,$$

por lo que si v_j es una autofunción, entonces también lo es av_j , para cualquier escalar no nulo. Las Componentes Principales Funcionales v_j se normalizan, por lo que $\|v_j\| = 1$, pero esto no determina el signo de v_j . Así que si \hat{v}_j es una estimación calculada a partir de los datos, solo podemos esperar que $\hat{c}_j \hat{v}_j$ esté cerca de v_j , donde

$$\hat{c}_j = \text{signo}(\langle \hat{v}_j, v_j \rangle).$$

Al discutir las propiedades de los \hat{v}_j , siempre asumiremos implícitamente que este signo ha sido tenido en cuenta, sabiendo que \hat{c}_j no puede calcularse a partir de los datos, por lo que debe uno asegurarse de que los procedimientos para los que usemos estos \hat{v}_j no dependan de los \hat{c}_j .

El problema de estimar las componentes principales $\hat{v}_j(t)$ y las puntuaciones $\hat{\lambda}_j$ no es trivial debido a la limitación inherente de que no tenemos acceso a las funciones completas x_i para cualquier t , sino que solo las conocemos en un número finito de momentos concretos del tiempo. Además, en muchas situaciones, estos momentos de tiempo t_{ij} están prefijados y no pueden ser elegidos libremente. Esta limitación plantea desafíos significativos para la estimación precisa de las componentes principales y los scores.

Además, si los momentos de tiempo t_{ij} están prefijados y no coinciden con los momentos de cambio clave en las funciones $x_i(t)$, podríamos perder información crucial sobre la variación temporal de las funciones. Esto puede conducir a una subestimación de la variabilidad y una distorsión de las relaciones temporales entre las observaciones.

3.1. Obtención desde la discretización de las funciones

Un enfoque simple para resolver la dificultad anteriormente apuntada es discretizar las funciones observadas x_i en una rejilla equiespaciada t_1, \dots, t_J de valores en el intervalo $[0, 1]$ donde se tiene que

$$t_{j+1} - t_j = \frac{1}{J},$$

para $j = 1, \dots, J - 1$. Esto permite representar las funciones de manera discreta y reducir la dimensión del problema. Sin embargo, hay que tener en cuenta que también introduce ciertas limitaciones, como la pérdida de información en la representación de las funciones.

Suponemos que los datos están centrados, es decir, se reemplaza $x_i(t_j)$ por

$$x_i(t_j) - \frac{1}{n} \sum_{i=1}^n x_i(t_j).$$

Esto da como resultado una matriz X

$$X = \begin{bmatrix} x_1(t_1) & x_1(t_2) & \dots & x_1(t_j) & \dots & x_1(t_J) \\ x_2(t_1) & x_2(t_2) & \dots & x_2(t_j) & \dots & x_2(t_J) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_i(t_1) & x_i(t_2) & \dots & x_i(t_j) & \dots & x_i(t_J) \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ x_n(t_1) & x_n(t_2) & \dots & x_n(t_j) & \dots & x_n(t_J) \end{bmatrix} \in \mathcal{M}_{n \times J},$$

cuya media de sus columnas es siempre cero. Como es habitual en el Análisis de Componentes Principales (PCA) multivariante, se buscan autovectores u de la matriz de varianza-covarianza muestral (recuérdese que los datos están centrados)

$$\frac{1}{n} X'X$$

con norma unitaria. Es decir, vectores que cumplan

$$\frac{1}{n} X'X u = \lambda u,$$

para $\|u\| = 1$.

Es importante notar que $J \gg n$ y que, por tanto, la matriz

$$X'X \in \mathcal{M}_{J \times J}$$

para, típicamente, un J grande. En este caso, un enfoque posible para encontrar las soluciones de la ecuación de autovalores

$$\frac{1}{n} X'X u = \lambda u$$

es trabajar en términos de la descomposición en valores singulares de la matriz X . Una descomposición en valores singulares de X es una factorización del tipo

$$X = U \Lambda V'$$

con $U \in \mathcal{M}_{n \times n}$, $V \in \mathcal{M}_{J \times n}$ ortogonales y $\Lambda \in \mathcal{M}_{n \times n}$ una matriz diagonal formada con los valores singulares de X en su diagonal principal ordenados de mayor a menor.

Si λ_i es el i -ésimo autovalor de la matriz $X'X$, entonces $\sqrt{\lambda_i}$ es el i -ésimo valor singular de la matriz X y sus autovectores corresponden a las columnas de U .

Los valores singulares en Λ en nuestro problema proporcionan información sobre la importancia relativa de los modos de variabilidad en los datos y las matrices U y V' describen cómo se relacionan las filas y columnas originales de la matriz X . De esta forma, nos quedamos con los vectores por columnas de la matriz U .

La primera autofunción $\hat{v}_1(t)$ se puede obtener de la primera columna de la matriz U de tal forma que

$$\hat{v}_1(t_j) = u_{1j}$$

para $j = 1, \dots, J$, con $(u_{11}, \dots, u_{1j}, \dots, u_{1J})'$ siendo la primera columna de la matriz U .

Usando cuadratura numérica y viendo \hat{v}_1 como una función se tiene que:

$$\int_0^1 c(t_j, s) \hat{v}_1(s) ds \simeq \sum_{j=1}^J c(t_j, t_s) \hat{v}_1(t_s) \frac{1}{J} \simeq \lambda \frac{1}{J} \hat{v}_1(t_s).$$

Luego \hat{v}_1 es aproximadamente autofunción (con autovalor distinto de λ) del operador de covarianzas. Necesitamos que la función $\hat{v}_1(\cdot)$ tenga norma en $L^2[0, 1]$ igual a 1, es decir:

$$\|\hat{v}_1\|^2 = \int \hat{v}_1^2(s) ds = 1,$$

de esta forma, necesitamos que

$$\int \hat{v}_1^2(s) ds \simeq \sum_{j=1}^J \frac{1}{J} \hat{v}_1^2(t_j) = 1.$$

Como sabemos que lo que se cumple es que

$$\sum_{j=1}^J \hat{v}_1^2(t_j) = \sum_{j=1}^J u_{1j}^2 = 1$$

(la primera columna de U tiene norma 1), entonces se necesita un reescalado usando $\sqrt{J} \cdot \hat{v}_1(t_j)$ como autofunción de norma 1. En realidad, lo que obtenemos

es la autofunción evaluada en t_1, \dots, t_J , pero se puede realizar una interpolación para obtener la autofunción $\hat{v}_1(t)$ definida para todo $t \in [0, 1]$.

El proceso sería similar para $\hat{v}_2(t), \dots, \hat{v}_p(t)$, usando las otras autofunciones de las Componentes Principales Empíricas.

3.2. Obtención desde su representación en bases funcionales

Aplicando lo visto en la Sección 2.5, una forma de reducir el problema de autofunciones (26) a una forma discreta o matricial es expresar cada función $x_i(t)$ como una combinación lineal de funciones de base prefijadas $B_1(t), \dots, B_M(t)$, como por ejemplo, bases de Fourier o de B-splines. El número M de funciones de base utilizadas depende de diversas consideraciones: cuántos puntos de muestreo discretos había en los datos originales, la eficiencia de las funciones de base para reproducir el comportamiento de las funciones originales, etc.

Recordemos que una vez esté seleccionada la base, cada observación $x_i(t_{ij})$ puede estar sujeta a errores aleatorios

$$x_i(t_{ij}) = \sum_{m=1}^M b_{im} B_m(t_{ij}) + \varepsilon_{ij}$$

con $i = 1, \dots, n$, y $j = 1, \dots, J_i$. donde ε_{ij} son unos errores aleatorios. Reescrita de forma matricial, tal como vimos en la Sección 2.5, se tiene

$$\mathbf{x}_i = \mathbf{B}_i \mathbf{b}_i + \varepsilon_i.$$

Este enfoque supone que cada función tiene una representación

$$x_i(t) = \sum_{m=1}^M b_{im} B_m(t).$$

Así, las curvas se reducen a una representación de la forma:

$$x_i(t) \simeq \begin{bmatrix} b_{i1} & b_{i2} & \dots & b_{iM} \end{bmatrix} \begin{bmatrix} B_1(t) \\ B_2(t) \\ \vdots \\ B_M(t) \end{bmatrix} = \mathbf{b}'_i \mathbf{B}(t).$$

Creamos ahora una matriz β cuyas filas son los vectores \mathbf{b}_i , $i = 1, \dots, n$. Es decir,

$$\beta = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \vdots \\ \mathbf{b}_n \end{bmatrix} \in \mathcal{M}_{n \times n}.$$

Podemos, de nuevo, trabajar con β centrada, es decir, las medias por columnas son cero, que es equivalente a centrar las funciones con $x_i(t) - \hat{\mu}(t)$ para

$$\hat{\mu}(t) = \frac{(\mathbf{b}'_1 \mathbf{B}(t) + \dots + \mathbf{b}'_n \mathbf{B}(t))}{n}.$$

De esta forma, en términos matriciales, la función de covarianza sería

$$\hat{c}(t, s) = \frac{1}{n} \mathbf{B}(t)' \beta' \beta \mathbf{B}(s).$$

Ahora, reescribimos también la autofunción del operador de covarianza mediante una expansión en base $\{B_m\}_{m=1}^M$. Es decir,

$$v(t) = \sum_{m=1}^M u_m B_m(t) = u' \mathbf{B}(t) = \mathbf{B}'(t) u$$

para un $u \in \mathbb{R}^M$. Esto da como resultado para una matriz $W \in \mathcal{M}_{M \times M}$ cuyos elementos son

$$w_{m_1 m_2} = \int B_{m_1}(t) B_{m_2}(t) dt,$$

para $m_1, m_2 = 1, \dots, M$.

Por tanto,

$$\int \hat{c}(t, s) v(s) ds = \int \frac{1}{n} \mathbf{B}(t)' \beta' \beta \mathbf{B}(s)' u ds = \mathbf{B}(t)' \frac{1}{n} \beta' \beta W u = \lambda \mathbf{B}(t)' u. \quad (27)$$

El problema de autofunciones (26), que se debe verificar para todos los valores de $t \in [0, 1]$, se puede expresar como

$$\frac{1}{n} \beta' \beta W u = \lambda u.$$

Queremos además que $v(t) = \mathbf{B}'(t) u$ verifique

$$\|v\|^2 = \int v(t)^2 dt = 1,$$

lo que es equivalente a pedir que $u'Wu = 1$.

De manera similar, dos funciones v_{l_1} y v_{l_2} serán ortogonales si y solo si los vectores correspondientes de coeficientes satisfacen $u'_{l_1}Wu_{l_2} = 0$ para $l_1 \neq l_2$ donde los u_l son autovectores de $\beta'\beta W$ asociados a autovalores $\lambda_1 > \lambda_2 > \dots$. Luego, para obtener los componentes principales requeridos, definimos $\tilde{u} = W^{\frac{1}{2}}u$ y resolvemos el problema de autovalores simétricos con $\|\tilde{u}\| = 1$ siguiente:

$$\frac{1}{n}W^{\frac{1}{2}}\beta'\beta W^{1/2}\tilde{u} = \rho\tilde{u}$$

y calculamos $u = W^{-\frac{1}{2}}\tilde{u}$ para cada autovector. La autofunción será, por tanto,

$$v(t) = \mathbf{B}'(t)W^{-\frac{1}{2}}\tilde{u}.$$

Nótese que se tiene

$$\int v(t)^2 dt = \tilde{u}'W^{-\frac{1}{2}}WW^{-\frac{1}{2}}\tilde{u} = \tilde{u}'\tilde{u} = 1.$$

Un caso particular se da si la base es ortonormal, ya que se cumple que $\|B_m\| = 1$ y $\langle B_{m_1}, B_{m_2} \rangle = 0$ para $m_1 \neq m_2$, lo que significa que $W = \mathbf{I}_M$. Esto sucede por ejemplo, en el caso en que la base sea la de Fourier. Así, el problema se reduce a hacer PCA con la matriz de coeficientes β , y solo necesitamos buscar los autovectores asociados a los M autovalores mayores de la matriz

$$\frac{1}{n}\beta'\beta,$$

ya que

$$W^{-\frac{1}{2}} = \mathbf{I}_M.$$

4. Ejemplos

4.1. Ejemplo 1: Datos AEMET

Nuevamente, nos enfocamos en los datos proporcionados por AEMET. El conjunto de datos consiste en un resumen diario recopilado de 73 estaciones meteorológicas, abarcando el periodo comprendido entre los años 1980 y 2009. Cada estación meteorológica está caracterizada por información geográfica detallada, que incluye su nombre, provincia, altitud, longitud y latitud. Además, se presenta el promedio diario de la temperatura, la precipitación y la velocidad del viento durante dicho intervalo temporal con una frecuencia diaria. Es importante destacar que la temperatura se registra en grados centígrados, la velocidad del viento en metros por segundo, y (el logaritmo de) la precipitación de lluvia en milímetros. La diferencia logarítmica es una medida que proporciona una forma de cuantificar el cambio porcentual entre dos periodos consecutivos de tiempo.

Expondremos el código utilizado para el caso de la temperatura media diaria, siendo similar para la precipitación y la velocidad del viento diarias. Previo al inicio, se ha realizado una manipulación de los datos, incluyendo una categoría que señala la distribución de las distintas estaciones meteorológicas en cinco zonas climáticas, predefinidas para la subdivisión del territorio español: Islas Canarias, Norte, Sur, Este y Centro. Asimismo, se han asignado colores específicos a cada categoría.

La codificación asignada a cada una de las estaciones analizadas se resume en la siguiente tabla:

ACR1	A Coruña (N)	ACR2	Alvedro (N)
ACR3	Santiago de Compostela (N)	ALA	Vitoria (N)
ALB	Albacete (C)	ALI1	Alicante (E)
ALI2	El Altet (E)	ALM	Almería (S)
AST1	Avilés (N)	AST2	Oviedo (N)
BAD	Badajoz (C)	BAL1	Ibiza (E)
BAL2	Menorca (E)	BAL3	Palma de Mallorca (E)
BAL4	Son Sant Joan (E)	BAR1	Observatorio Fabra (E)
BAR2	Barcelona (E)	BUR	Burgos (C)
CAD1	Jerez de la Frontera (S)	CAD2	Tarifa (S)
CAN	Santander (N)	CAS	Castellón (E)

CRE	Ciudad Real (C)	COR	Córdoba (S)
CUE	Cuenca (C)	GIR	Girona (E)
GRA1	Aeropuerto de Granada (S)	GRA2	Base aérea (S)
GUA	Molina de Aragón (C)	GUI1	Igueldo (N)
GUI2	Fuenterrabia (N)	HUE	Huesca (C)
LRI	Logroño (C)	LPA1	Fuerteventura (IC)
LPA2	Lanzarote (IC)	LPA3	Las Palmas (IC)
LEO1	León (C)	LEO2	Ponferrada (C)
MAD1	Colmenar Viejo (C)	MAD2	El Retiro (C)
MAD3	Barajas (C)	MAD4	Cuatro vientos (C)
MAD5	Getafe (C)	MAD6	Torrejón de Ardoz (C)
MAD7	Puerto de Navacerrada (C)	MAL	Málaga (S)
MEL	Melilla (S)	MUR1	Alcantarilla (E)
MUR2	SAN Javier (E)	NAV	Pamplona (C)
OUR	Ourense (N)	PON	Vigo (N)
SAL1	Salamanca (C)	SAL2	Matacan (C)
SCT1	Hierro (IC)	SCT2	Izaña (IC)
SCT3	La Palma (IC)	SCT4	Santa Cruz de Tenerife (IC)
SCT5	Los Rodeos (IC)	SCT6	Tenerife Sur Aeropuerto (IC)
SEV1	Morón de la Frontera (S)	SEV2	Sevilla (S)
SOR	Soria (C)	TAR1	Reus (E)
TAR2	Tortosa (E)	VLC1	Valencia (E)
VLC2	Aeropuerto de Valencia (E)	VLL1	Valladolid (C)
VLL2	Villanubla (C)	VIZ	Bilbao (N)
ZAM	Zamora (C)	ZAR1	Daroca (C)
ZAR2	Zaragoza (C)		

En la Figura 5(a), se presentan los datos sin procesar del promedio de las temperaturas ($^{\circ}\text{C}$) al promediar diariamente en las temperaturas medias los 30 años correspondientes al período comprendido entre 1980 y 2009. Estos datos corresponden a diversas estaciones meteorológicas distribuidas a lo largo del país.

Usando el paquete `fda.usc` de R, se incluyen en la Figura 5(b) las curvas suavizadas obtenidas mediante una representación basada en B-splines para facilitar la realización de los estudios previstos, como el Análisis de Componentes Principales. La selección del número de bases (en este caso 15 elementos de la base de

B-splines) se ha realizado de manera subjetiva, tomando en cuenta visualizaciones de los datos suavizados.

La función `fdata2fd` calcula los coeficientes funcionales a partir de datos funcionales representados en una base indicando el número de bases.

```
plot(aemet$temp)
aemet.temp.fd <-fdata2fd(aemet$temp,type.basis="bspline",
                        nbasis=15)
plot(aemet.temp.fd)
```

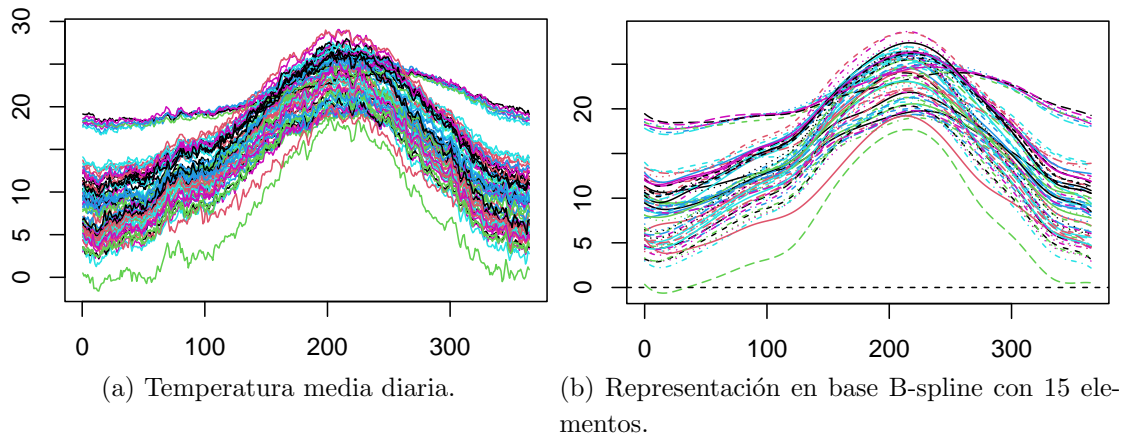


Figura 5: Evolución de la temperatura media diaria.

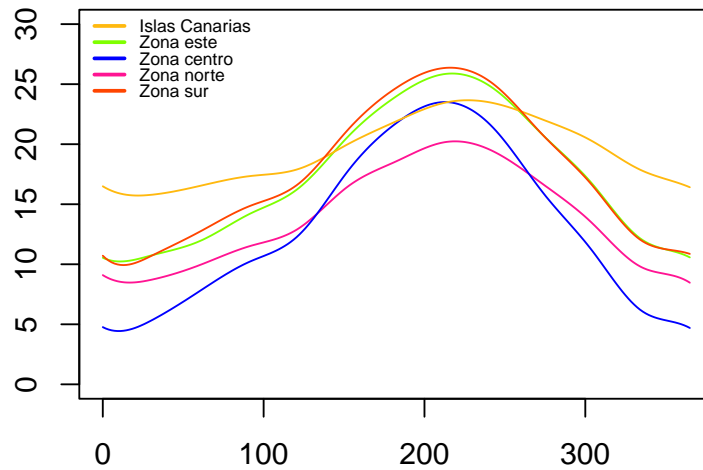


Figura 6: Evolución de la temperatura media en las distintas regiones.

Se presenta en la Figura 6 la representación del comportamiento diario del promedio de temperatura de las diversas estaciones meteorológicas distribuidas en las cinco zonas climáticas. Las estaciones en las Islas Canarias exhiben una variación térmica anual mínima, manteniendo temperaturas templadas y agradables de manera constante a lo largo de todo el año, fenómeno atribuible a su ubicación geográfica y al clima subtropical que prevalece en dicha región.

En contraste, las estaciones ubicadas en el interior de la Península presentan características propias de un clima más continental, con una marcada variación térmica que se manifiesta en veranos cálidos e inviernos extremadamente fríos, esto debido a la ausencia de la moderación térmica proporcionada por la influencia marina. Por otro lado, las estaciones al norte de la península, con su clima oceánico, exhiben una amplitud térmica más moderada, caracterizada por veranos frescos e inviernos templados, gracias a la influencia atenuante del océano.

Por último, las regiones mediterráneas y del sur de España tienden a registrar temperaturas más elevadas a lo largo del año en comparación con las del interior, manifestando notables diferencias entre los meses de invierno y verano y evidenciando así una considerable variación térmica en estas áreas.

El código utilizado para generar la representación de la Figura 6 se presenta a continuación. La función `mean.fd` calcula la media de un conjunto de funciones en un objeto de datos funcionales, y `legend` añade una leyenda.

```
norte.temp.fd <- aemet.temp.fd[norte]
norte.temp.mean <- mean.fd(norte.temp.fd)
centro.temp.fd <- aemet.temp.fd[centro]
centro.temp.mean <- mean.fd(centro.temp.fd)
este.temp.fd <- aemet.temp.fd[este]
este.temp.mean <- mean.fd(este.temp.fd)
islas.temp.fd <- aemet.temp.fd[islas]
islas.temp.mean <- mean.fd(islas.temp.fd)
sur.temp.fd <- aemet.temp.fd[sur]
sur.temp.mean <- mean.fd(sur.temp.fd)
plot(norte.temp.mean, ylim=c(0,30), col=colores[4])
lines(centro.temp.mean, col=colores[3])
lines(este.temp.mean, col=colores[2])
lines(islas.temp.mean, col=colores[1])
lines(sur.temp.mean, col=colores[5])
legend("topleft", col=colores, legend=c("Islas Canarias",
    "Zona este", "Zona centro", "Zona norte", "Zona sur"),
    bty = "n", lwd=2, cex=0.6)
```

Después de aplicar técnicas de suavizado a los datos, se procede a realizar el Análisis de Componentes Principales Funcionales. Como ya se ha comentado, este análisis tiene como finalidad examinar los modos de variación más significativos presentes en los datos y, simultáneamente, proporcionar una visión sobre la complejidad de las curvas, permitiendo evaluar cuántos comportamientos de variabilidad esencial se pueden identificar en ellas.

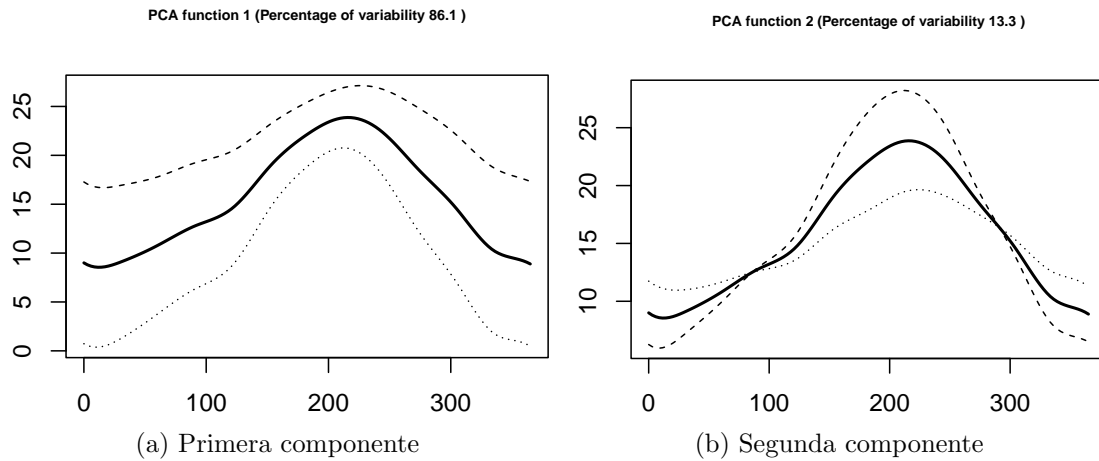


Figura 7: Componentes principales funcionales.

La Figura 7 representa la perturbación resultante de sumar o restar a la función media del conjunto de datos un múltiplo apropiado de cada una de las dos componentes principales \hat{v}_1 y \hat{v}_2 . Esto facilita la comprensión de la variabilidad explicada por cada componente. La función `pca.fd` realiza un Análisis de Componentes Principales Funcionales también con el objetivo mostrar los tipos de variación en una muestra de funciones. Estos modos de variación también son a veces conocidos como “armónicos”. Calcula además, los autovalores que indican la importancia de cada modo de variación, y los scores, que representan cómo de fuerte es en cada función observada la variación característica correspondiente a cada componente.

La función `plot.pca.fd` muestra esos tipos de variación y los autovalores indican la importancia relativa de cada modo de variación.

```
aemet.temp.pca=pca.fd(aemet.temp.fd, nharm=2)
plot.pca.fd(aemet.temp.pca, harm =1, cex.main=0.6)
plot.pca.fd(aemet.temp.pca, harm =2, cex.main=0.6)
```

En los títulos de estas figuras podemos ver la variabilidad explicada por cada componente. Dado que ambas componentes explican conjuntamente un 99,4 % de la variabilidad total, no parece ser necesario tratar con una tercera componente principal. Los porcentajes de variabilidad explicada de cada componente principal se derivan de los autovalores del problema de búsqueda de autofunciones.

La primera componente funcional, por sí sola, explica el 86,1 % de la variabilidad. A medida que el valor asociado en esta componente aumenta en una de las funciones analizadas, se registra una temperatura más elevada a lo largo de todo el año. Por el contrario, un valor menor indica una estación ubicada en un lugar más frío en general durante el año. La segunda componente, explica el 13,3 % y permite clasificar las funciones de temperatura observadas según la estabilidad de las temperaturas a lo largo del año. Una curva con un valor más alto en esta componente presentará temperaturas más altas en el centro y más bajas en los extremos, indicando una región con veranos más cálidos e inviernos más fríos. Por otro lado, un valor menor en esta componente indica una mayor suavidad o estabilidad en la curva. Así, los comportamientos de las curvas pueden clasificarse en función de la mayor o menor influencia de estas dos componentes.

Ahora procedemos a analizar el gráfico de scores, el cual representa las coordenadas de cada una de las estaciones meteorológicas en un espacio bidimensional definido por las dos componentes principales. Esta representación facilita la relación entre las estaciones según su proximidad en el gráfico. El código empleado para la Figura 8 es el siguiente.

```
plot(aemet.temp.pca$scores[, 1], aemet.temp.pca$scores[, 2],
     col=aemet$df$colors)
text(aemet.temp.pca$scores[, 1], aemet.temp.pca$scores[, 2],
     labels = names, pos = 3, cex = 0.7)
```

Si dividimos el gráfico en cuatro cuadrantes iguales utilizando los ejes horizontal y vertical, podríamos interpretar que las estaciones en el cuadrante, con valores negativos para el primer score y positivos para el segundo, corresponden a lugares fríos, donde hay una gran diferencia entre los meses más cálidos y los más fríos. En la parte superior derecha se encuentran las estaciones con temperaturas más altas, que también presentan una variación anual considerable entre los meses cálidos y los meses fríos. Los puntos por debajo de cero en la segunda componente representan temperaturas más regulares a lo largo del año, siendo los puntos a la izquierda más fríos y a la derecha correspondientes a zonas más cálidas.

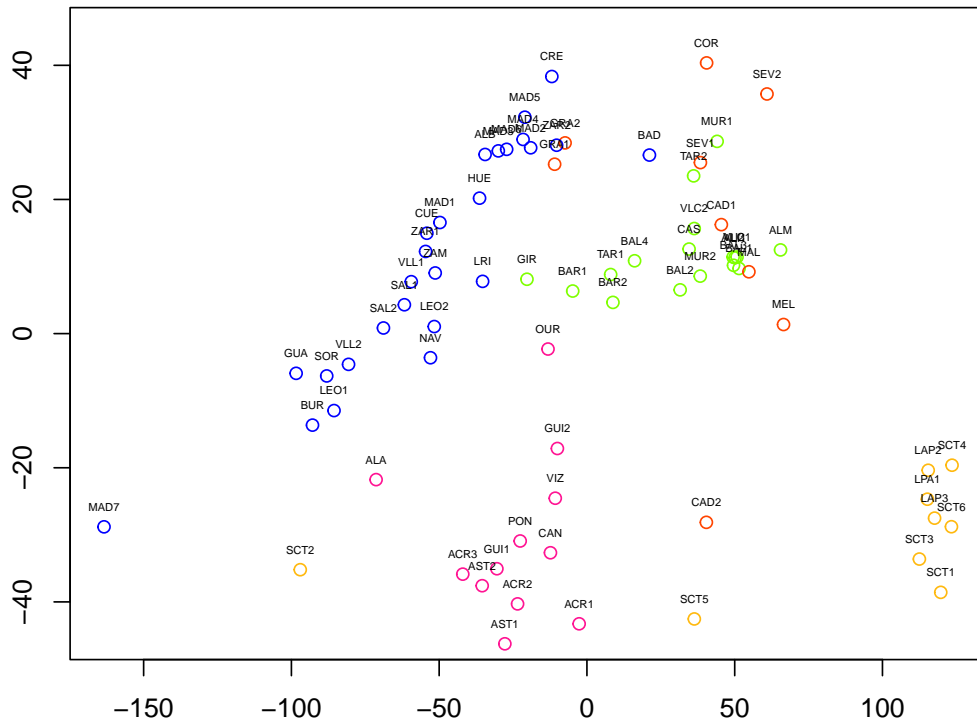


Figura 8: Gráfico de scores de las curvas de temperatura.

Además, las estaciones meteorológicas de las Islas Canarias se encuentran muy juntas entre sí y aisladas de otros puntos. Esto se debe a que, aunque destacan muy positivamente en la primera componente principal debido a las altas temperaturas generales, presentan una variación muy pequeña a lo largo del año, lo que se refleja en scores altamente negativos en la segunda componente principal. Esta peculiaridad climática se debe, en gran medida, a su posición geográfica única. En la nube principal de puntos, se observa que en la parte inferior se concentran las estaciones correspondientes a la zona norte. Mayoritariamente con valores positivos para ambas componentes se encuentran las estaciones de la zona este y sur, caracterizadas por temperaturas más altas a lo largo del año y una amplitud térmica significativa. Otro grupo destacado abarca las estaciones céntricas en la primera componente con scores altos en la segunda. Aquí se incluyen muchas estaciones en la submeseta sur, así como las ubicadas en el noreste de la península. Aunque su comportamiento medio anual no destaca, están ubicadas en zonas con notables diferencias entre los meses cálidos y fríos. Por último, hay un grupo de estaciones con temperaturas algo más frías y una menor amplitud térmica, principalmente

ubicadas en la submeseta norte.

Destaca visualmente la presencia de dos puntos correspondientes a la estación ubicada en el Puerto de Navacerrada (MAD7), en Madrid, y la ubicada en Izaña (SCT2), en la isla canaria de Tenerife, muy cerca del pico del Teide. Dadas sus altitudes de casi 1900 metros y 2400 metros, respectivamente, registran temperaturas mucho más bajas a lo largo del año que cualquiera de las otras estaciones de la zona.

En la Figura 9 se presentan los datos correspondientes al promedio de la velocidad del viento diario (m/s) a lo largo del período citado en las diversas estaciones meteorológicas españolas. Al igual que en el caso de las temperaturas, se incluyen curvas suavizadas obtenidas mediante una representación basada en B-splines, de nuevo con 15 elementos, proporcionando así una visualización más comprensible de la tendencia de la velocidad del viento a lo largo del tiempo.

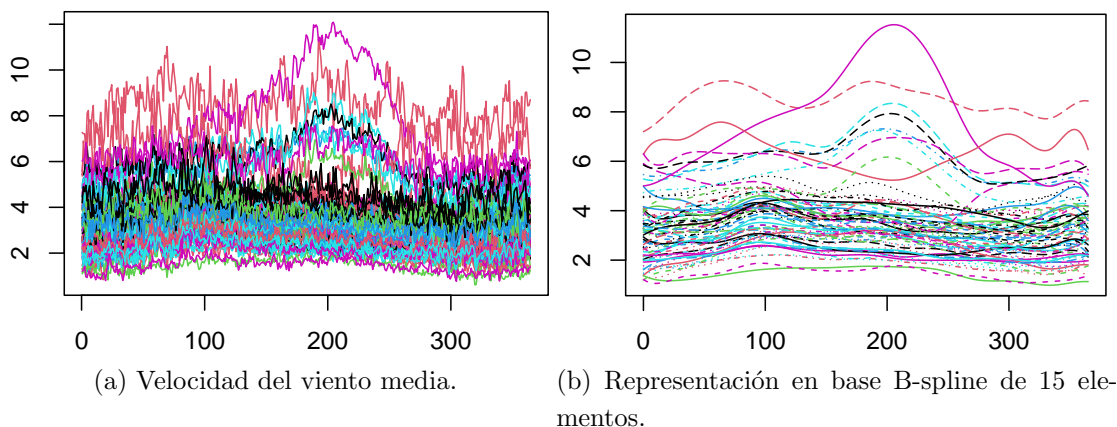


Figura 9: Evolución de la velocidad del viento diaria.

En la Figura 10, se observa que la velocidad del viento en las Islas Canarias es significativamente superior a la de toda la península a lo largo de todo el año, este fenómeno puede ser atribuido a la influencia constante de los vientos alisios, con intensificaciones notables durante los meses de verano. En cuanto a las diversas regiones que componen la península, se destaca que la velocidad del viento tiende a ser ligeramente superior en los meses de primavera. En particular, el sur de España exhibe velocidades superiores a las de otras zonas, mientras que en el centro de la península se registra una menor velocidad del viento durante prácticamente todo el año.

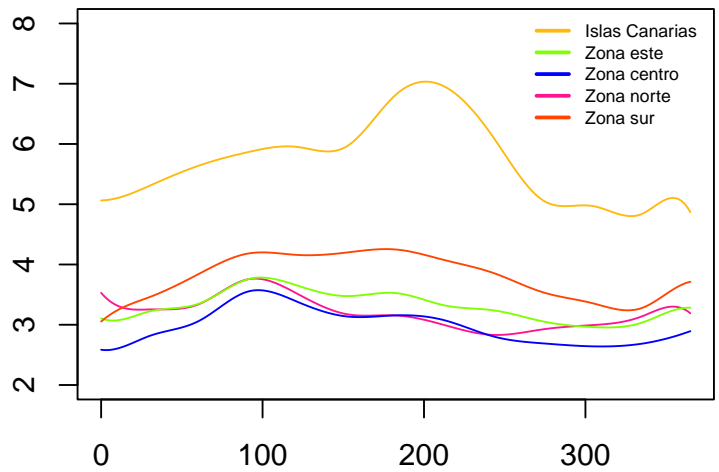


Figura 10: Evolución del viento media en las distintas regiones.

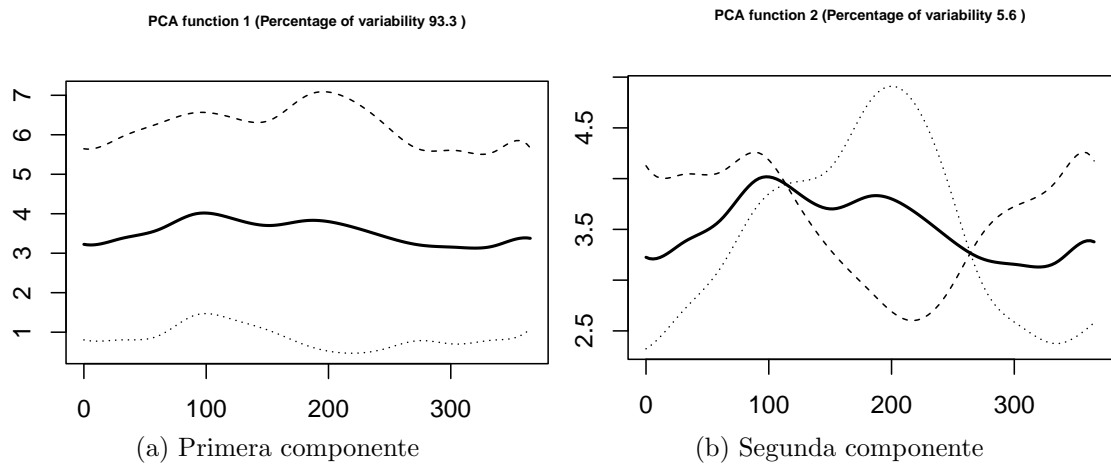


Figura 11: Componentes principales funcionales.

La primera componente principal asociada con la velocidad del viento en las diversas estaciones meteorológicas del país explica el 93,3% de la variabilidad, mientras que la segunda componente explica casi el 5,6%. De manera conjunta, ambas componentes explican el 98,9% de la variabilidad total, como se ilustra en la Figura 11.

En relación con la primera componente, se puede establecer que a medida que el valor asociado a esta componente aumenta en una de las curvas analizadas, se espera una mayor velocidad del viento global a lo largo de todo el año para esa estación específica. De manera inversa, valores más bajos indicarán estaciones

ubicadas en lugares donde se registra una menor velocidad del viento. En lo que respecta a la segunda componente, su correspondencia con el comportamiento de la velocidad del viento anual es de naturaleza algo más compleja. Un valor más bajo en esta componente indica que en la región representada, la velocidad del viento es mayor durante los meses cálidos en comparación con los meses fríos. Por otro lado, un valor más alto sugiere que la región experimentará ráfagas de viento más fuertes o, alternativamente, que la intensidad del viento variará de manera más intermitente a lo largo del año.

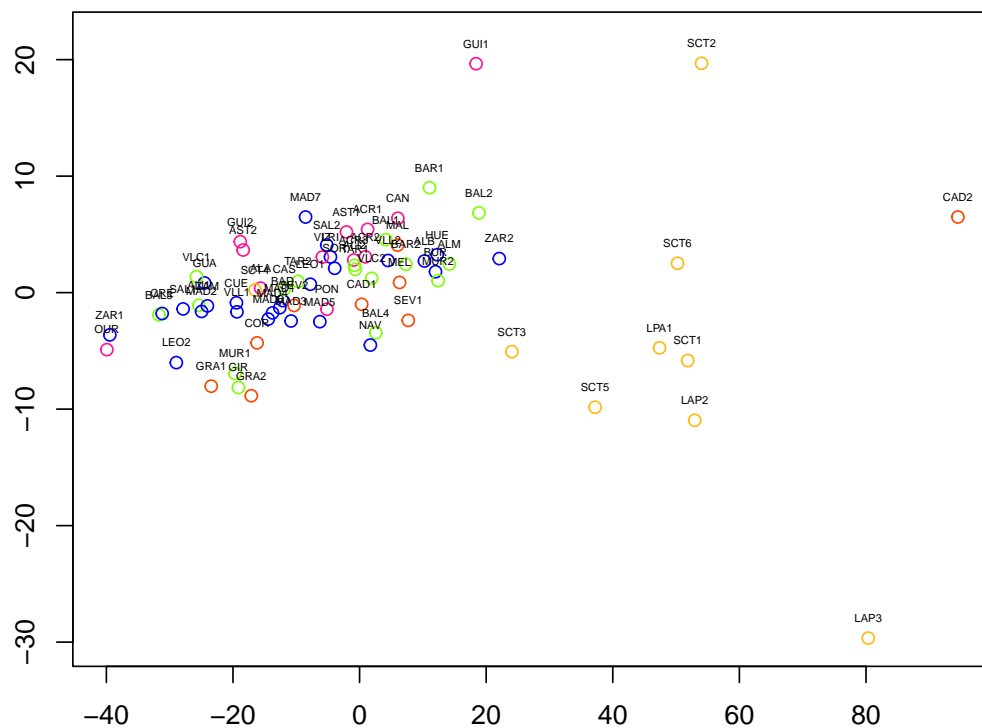


Figura 12: Gráfico de scores de las curvas de la velocidad del viento.

Revisamos nuevamente el gráfico de scores, como se presenta en la Figura 12. Se pueden identificar inicialmente dos grupos principales: las Islas Canarias y el resto de la península. Las estaciones meteorológicas situadas en las Islas Canarias exhiben una mayor velocidad del viento a lo largo de todo el año, evidenciado por los valores positivos y elevados en la primera componente. Esta tendencia puede atribuirse a la influencia constante de los vientos alisios, y durante los meses de verano, la circulación atmosférica en la región subtropical del Atlántico y la topografía montañosa de las islas favorecen la intensificación de estos vientos

alisios, lo cual se refleja en los valores negativos de la segunda componente en la mayoría de las estaciones de las Canarias.

En relación a las distintas regiones de la península, se observa que las estaciones ubicadas en el centro tienen un menor peso en la primera componente, indicando una velocidad del viento más baja. Además, todo el grupo presenta valores cercanos a cero en la segunda componente, sugiriendo una menor variabilidad en la intensidad del viento entre los meses cálidos y fríos o una distribución más uniforme de las ráfagas de viento a lo largo del año.

Cabe destacar la estación ubicada en Tarifa (CAD2), Cadiz. Tarifa es la ciudad más meridional de la península ibérica, ubicada en el estrecho de Gibraltar, crea un pasillo de viento natural entre el mar Mediterráneo y el océano Atlántico que actúa como un embudo natural que acelera el flujo de viento.

En la Figura 13, se presentan los datos correspondientes al promedio de las precipitaciones diarias (mm) a lo largo del período en las diversas estaciones meteorológicas. Es importante destacar que el conjunto de datos representa el logaritmo de la precipitación. La diferencia logarítmica entre dos días consecutivos proporciona una medida directa del cambio porcentual en la precipitación entre esos dos días. Por ejemplo, si el logaritmo de la precipitación para un día es 0,2763683 y para el siguiente día es 0,3998941, entonces la diferencia logarítmica es $0,3998941 - 0,2763683 = 0,1235258$, lo que indica un aumento de aproximadamente un 12,35 % en la precipitación de un día para otro.

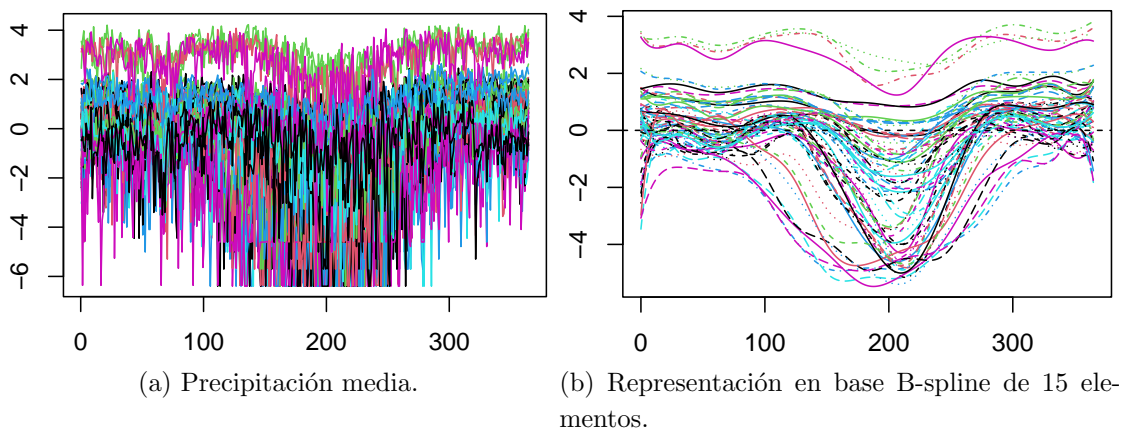


Figura 13: Evolución de la precipitación media diaria.

En la Figura 14, se analiza la distribución de las precipitaciones en diferentes regiones. La zona norte del país se destaca como la región más lluviosa, con una precipitación constante a lo largo de todo el año debido a su clima oceánico. En contraste, tanto en la zona centro como en la zona este, se observan mayores precipitaciones durante los meses fríos, caracterizados por climas con veranos secos y áridos. En el sur de la península y en las Islas Canarias, se evidencia una notable disparidad entre los meses de verano y el resto del año, siendo las precipitaciones más escasas, especialmente en las islas.

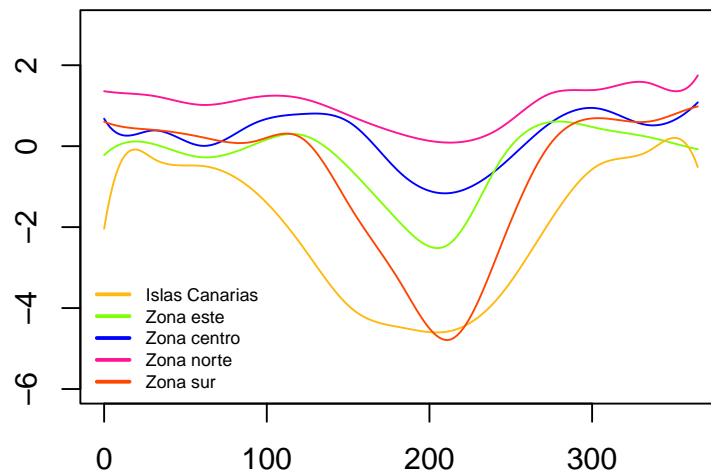


Figura 14: Evolución de la precipitación media en las distintas regiones.

Las componentes principales asociadas a las precipitaciones explican conjuntamente un 94% de la variabilidad, como se ilustra en la Figura 15. La primera componente explica un 85,1%, mientras que la segunda contribuye con un 8,9%. En este contexto, no parece pertinente ni necesario explorar la existencia de una tercera componente principal, ya que, en el mejor de los casos, su contribución sería del 6% en este aspecto.

En relación con la primera, se puede establecer que a medida que el valor asociado a esta componente aumenta, se espera que se registren mayores precipitaciones en la estación correspondiente. Por el contrario, valores más bajos indicarán estaciones ubicadas en lugares con niveles más bajos de precipitación. En relación con la segunda, su interpretación se torna más compleja. Valores cercanos a cero sugieren una distribución relativamente constante de lluvias a lo largo del año. Valores más altos indicarían una mayor variabilidad entre los meses

cálidos y fríos, y por otro lado, valores más bajos sugieren una variabilidad más suave a lo largo del año, con menos diferencias notables entre las estaciones.

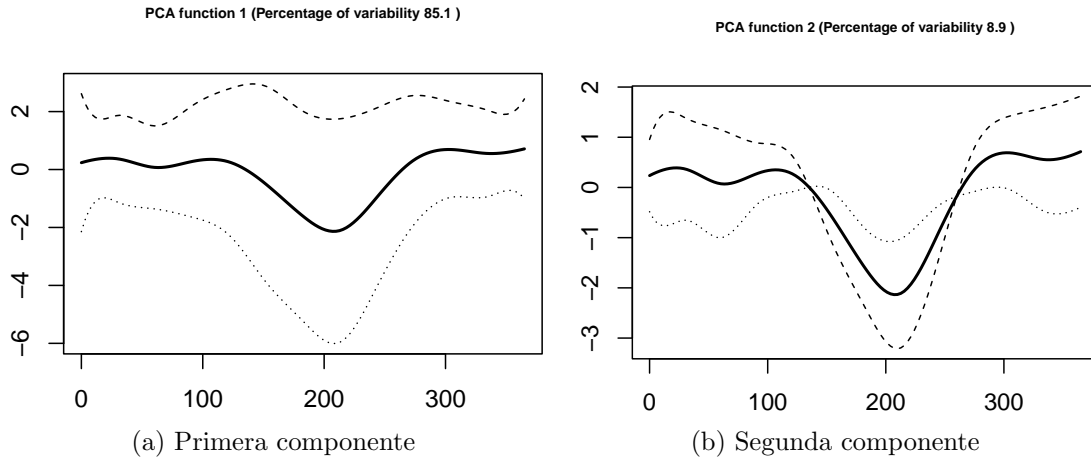


Figura 15: Componentes principales funcionales.

Al analizar el gráfico de scores en la Figura 16, se observan patrones distintivos. En el cuadrante con valores negativos en la primera componente y valores positivos en la segunda, se ubican principalmente estaciones meteorológicas en la zona sur y en las islas, donde las precipitaciones son escasas y se evidencia una marcada sequía durante los meses cálidos. En la nube de puntos con valores cercanos a cero en la primera componente y valores negativos en la segunda, se encuentran las zonas este y centro de la península, donde se destaca una mayor pluviosidad en los meses fríos en comparación con los meses cálidos. En el cuadrante con ambos valores positivos, se localiza la nube correspondiente a la zona norte, caracterizada por ser una región lluviosa a lo largo de todo el año. Destacan estaciones como la de Logroño (LRI), La Rioja, cuya proximidad a áreas montañosas y ubicación geográfica influye en la pluviosidad, así como la estación ubicada en Colmenar Viejo (MAD1), Madrid, cuya altitud supera los 1000 metros.

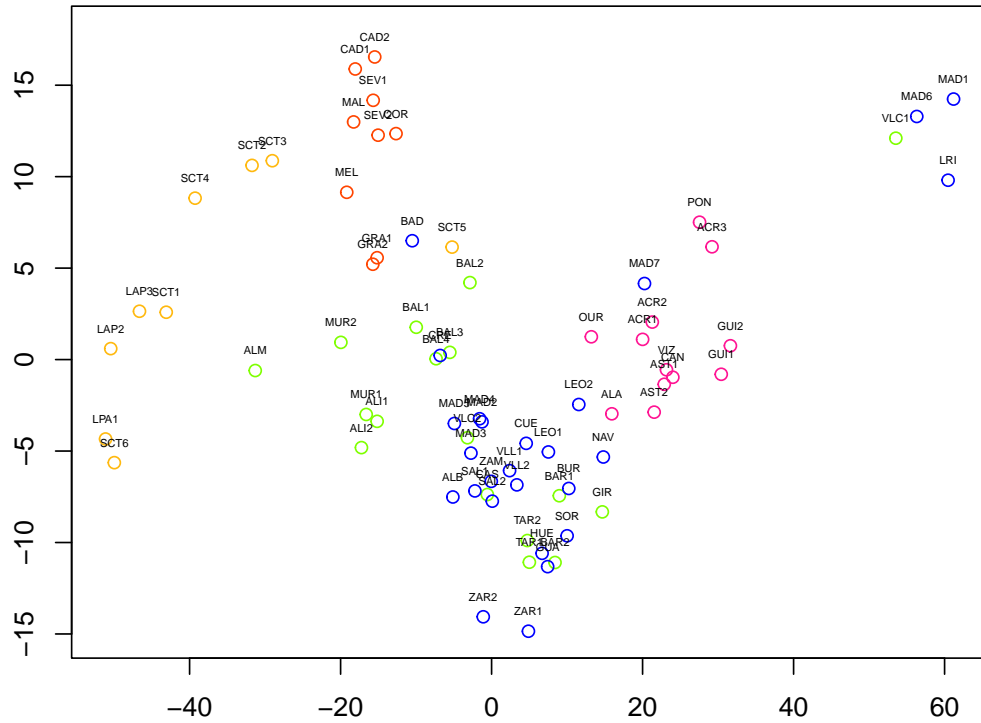


Figura 16: Gráfico de scores de las curvas de precipitación.

4.2. Ejemplo 2: Movimiento Browniano

En la Sección 2.4 fue tratado a nivel teórico el caso del movimiento Browniano. Recordemos que la forma habitual de aproximar un movimiento Browniano era

$$W\left(\frac{k}{K}\right) = \frac{1}{\sqrt{K}} \sum_{i=1}^k N_k,$$

para $N_k \sim \mathcal{N}(0, 1)$ con $k = 1, \dots, K$ e independientes entre sí.

Al aplicar Componentes Principales Funcionales a nivel teórico vimos que los autovalores y autofunciones del movimiento Browniano eran, respectivamente,

$$\lambda_j = \left(\frac{1}{\left(j - \frac{1}{2}\right)^2 \pi^2} \right)$$

y

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

para $j = 1, 2, \dots$

Ahora, procederemos a realizar una simulación de este movimiento Browniano utilizando `R` y calculando las Componentes Principales Funcionales empíricas. Se ha generado el proceso con $K = 70$ pasos aleatorios en el intervalo $[0, 1]$, como se visualiza en la Figura 17, para $n = 100$ realizaciones aleatorias del movimiento Browniano.

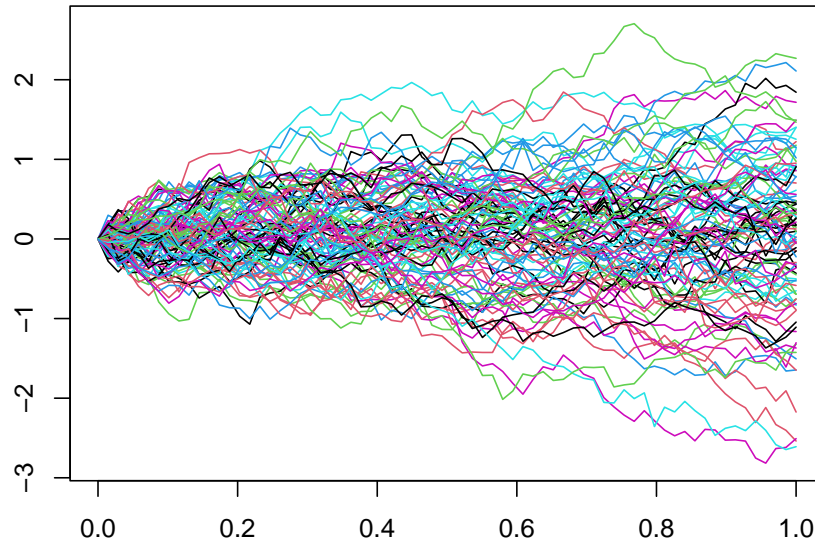


Figura 17: Simulación de $n = 100$ curvas del movimiento Browniano.

Posteriormente, se ha procedido a realizar el Análisis de Componentes Principales a nivel empírico basado en esta muestra de n funciones. El objetivo de este análisis es estudiar los modos de variación más importantes. En el caso específico que nos ocupa, inicialmente se suavizaron las curvas mediante una representación utilizando 11 elementos de la base B-spline mediante la función `fdata2fd`, seguido de la selección de cuatro componentes principales utilizando `pca.fd`, como se muestra en la Figura 18. En dicha figura, se representan las Componentes Principales Funcionales teóricas definidas en la Sección 2.4 y las Componentes Principales Funcionales estimadas obtenidas en `R`.

Nótese, en la Figura 18, que las Componentes Principales Funcionales teóricas (a) y las estimadas (b) son similares en su forma y comportamiento. Este alto grado de similitud sugiere que el modelo teórico y empírico propuestos capturan de manera efectiva la estructura subyacente en FDA, lo cual valida la utilidad de las Componentes Principales Funcionales empíricas en la representación y comprensión de los datos funcionales.

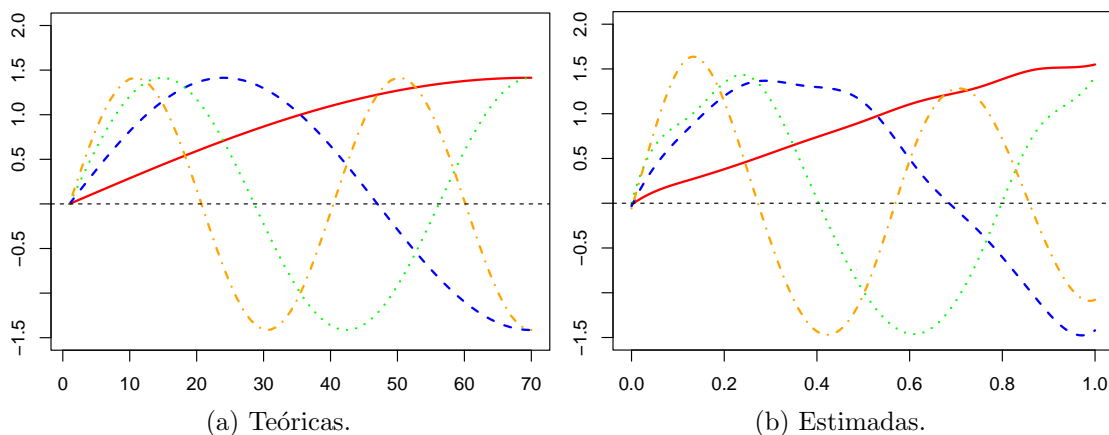


Figura 18: Componentes Principales Funcionales del movimiento Browniano a nivel teórico (a) y a nivel empírico (b).

5. Conclusiones y líneas futuras

El FDA consigue tratar como “funciones” aquellos datos que, tras la observación o experimentación, se perciben con forma de función. Estas técnicas sintetizan mucho mejor la información en nuestros datos que si escogemos la estrategia de tratar las funciones como datos multivariantes. La mayoría de las herramientas a nivel matemático del FDA hacen uso de los espacios de Hilbert. Por tal motivo, la memoria aborda aspectos matemáticos del FDA en dichos espacios, proporcionando un marco teórico para comprender la estructura y las propiedades de los datos funcionales.

Se ha visto que un primer paso habitual es expresar los datos funcionales mediante su aproximación usando bases funcionales. Más precisamente, nos hemos centrado en la utilización de bases B-splines, ya que permiten representar los datos mediante curvas suaves y con control local sobre la forma de las curvas. Esto es especialmente útil para modelar funciones que exhiben diferentes comportamientos en diferentes zonas del intervalo en el que están definidas.

La técnica del FDA que se ha tratado en mayor detalle ha sido el Análisis de Componentes Principales Funcionales, como complemento muy útil e informativo respecto al mero análisis descriptivo funcional. Las Componentes Principales Funcionales permiten estudiar “fuentes de variación” de interés en nuestros datos funcionales y representar los conjuntos de datos funcionales en dimensiones bajas donde resulta más cómoda su visualización. Trás su comprensión y formulación

a nivel teórico-matemático, la obtención de las Componentes Principales Funcionales empíricas se ha llevado a cabo a través de la discretización de las funciones y desde su representación en bases funcionales. Además, se ha mostrado una similitud notable tanto en forma como en comportamiento de dichas componentes teóricas y empíricas. Esta marcada similitud sugiere que tanto el modelo teórico como el empírico propuestos en la memoria capturan de manera efectiva la estructura subyacente en FDA, lo cual valida la utilidad de dichas componentes principales empíricas en la representación y comprensión de datos funcionales.

Finalmente, se han ilustrado los procedimientos revisados en esta memoria mediante su aplicación a conjuntos de datos reales, cuyas características contextuales se comprendían bien, y se han proporcionado resultados en gran medida previsibles pero también se han revelado rasgos inesperados de los datos, brindando información que no se disponía inicialmente.

Como líneas de trabajo futuras se podrían abordar otros problemas típicos en FDA como pueden ser la regresión escalar sobre funciones, los modelos de respuesta funcional, los modelos lineales generalizados funcionales, las series temporales funcionales o modelos de datos funcionales espaciales. Los procedimientos teórico-matemáticos descritos en esta memoria son la base sobre la que se fundamentan todas estas otras técnicas. Otra posible extensión son datos funcionales más generales, como los datos funcionales espacio-temporales, por ejemplo, imágenes que son analizadas en función del tiempo.

Referencias

- [1] Debnath, L. and Mikusinski, P. (2005). Introduction to Hilbert Spaces with Applications. Elsevier.
- [2] Febrero-Bande, M. and de la Fuente, M. O. (2012). Statistical Computing in Functional Data Analysis: The R Package `fda.usc`. Journal of Statistical Software, 51(4), 1–28. <https://doi.org/10.18637/jss.v051.i04>
- [3] Grenander U. (1950). Stochastic processes and statistical inference. Arkiv für Matematik 1:195–277.
- [4] Hsing, T. and Eubank, R. (2015). Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators. Wiley.
- [5] Kokoszka, K. and Reimherr, M. (2017). Introduction to Functional Data Analysis. CRC.
- [6] Laha, R. G. and Roghatgi, V. K. (1979). Probability Theory. Wiley.
- [7] Ramsay J. O. (1982). When the data are functions. Psychometrika 47:379–396.
- [8] Ramsay, J. O. and Dalzell C.J. (1991). Some Tools for Functional Data Analysis. Journal of the Royal Statistical Society. Series B, 53(3), 539–572.
- [9] Ramsay, J., Hooker, G. and Graves, S. (2009). Functional Data Analysis with R and MATLAB. Springer.
- [10] Ramsay, J. O., Hooker, G. and Graves, S. (2024). `fda`: Functional Data Analysis. R package version 6.1.8. <https://CRAN.R-project.org/package=fda>
- [11] Ramsay, J. O. and Silverman, B. W. (2005). Functional Data Analysis. Springer.
- [12] Rao, C. R. (1958). Some statistical methods for comparison of growth curves. Biometrics, 14, 1–17.
- [13] R Core Team (2020). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. , Vienna, Austria. <https://www.R-project.org/>.

- [14] Wasserman, L. (2006). All of nonparametric statistics. Springer Science and Business Media.