



Facultad de Matemáticas

Departamento de Estadística e
Investigación Operativa

— TRABAJO FIN DE GRADO —

Inferencia para datos funcionales

Lucía Manzorro Castrillón
Sevilla, Junio 2016

Agradecimientos

Quiero dar las gracias a todas aquellas personas que, durante este período, me han hecho crecer tanto en lo académico como en lo personal.

En primer lugar, me gustaría expresar mi agradecimiento a María Dolores Jiménez Gamero, mi tutora en este trabajo, por sus consejos, su dedicación y su gran paciencia, pero sobre todo por darme la oportunidad de realizar este proyecto.

Gracias a mis compañeros, a los que hoy puedo decir que son mis amigos, no creo que pudiera haber encontrado mejores personas con las que compartir esta etapa. Aunque esto se acabe estoy segura de que seguiréis formando parte de mi vida.

Y como no, dar las gracias a mi familia por confiar en mí y haberme apoyado en todo momento. Y en especial a mi madre, sin ella estoy segura de que no podría haber llegado hasta aquí.

¡Muchas gracias a todos!

Summary

In recent years, technological advances have made possible to have large masses of data that sometimes correspond to observations of a random phenomenon over a continuous range, that is, functions.

The work will deal with the theory and the practice of statistical methods for situations where the available data are functions (instead of numbers or vectors).

Chapters 1 and 2 are introductory, consisting of definitions and results that will help us to elaborate this memory. In order to develop the techniques in this piece of work, it will be required the functions to have certain properties. Specifically, it will be assumed that they belong to a Hilbert space. Therefore, in Chapter 1, we will present some properties of these spaces, focusing on the space L^2 , while in Chapter 2 we introduce some statistical concepts for functions. One of them is the covariance operator, which will be crucial for further developments.

In Chapter 3 we estimate the mean function, the covariance function and the covariance operator, and study some properties of such estimators.

Sometimes we need to work with the eigenvalues and eigenfunctions of the covariance operator, so in Chapter 4 we will estimate them and study some properties of the estimators, including the asymptotic normality.

In Chapter 5 we will introduce the concept of functional principal components, that allow us to reduce the dimension of infinitely dimensional functional data to a small finite dimension in an optimal way.

In Chapter 6 we will focus on two tests for:

- The equality of mean functions: we will study two methods, proving

theorems giving the asymptotic null distribution of the corresponding test statistics and their convergence in probability under alternatives.

- The equality of covariance operators: we will also study the asymptotic null distribution of the test statistic.

Finally, in Chapter 7, we will implement some of the techniques previously mentioned in the statistical language R.

Resumen

En los últimos años, los avances tecnológicos han permitido disponer de grandes masas de datos que, en ocasiones, se corresponden con observaciones de un fenómeno aleatorio a lo largo de un intervalo continuo, esto es, funciones.

El trabajo versará sobre teoría y práctica de métodos estadísticos para situaciones donde los datos disponibles son funciones (en vez de números o vectores).

Los Capítulos 1 y 2 son introductorios, constando de definiciones y resultados que nos ayudaran a desarrollar esta memoria. Para las técnicas usadas en este trabajo necesitamos que las funciones tengan ciertas propiedades. En nuestro caso que pertenezcan al espacio de Hilbert. Por ello, en el Capítulo 1, veremos algunas propiedades de dichos espacios centrándonos en el espacio L^2 , mientras que en el Capítulo 2 presentamos algunos conceptos estadísticos para funciones. Uno de ellos es el operador covarianza, que será de vital importancia en futuros desarrollos.

En el Capítulo 3 estimaremos la función media, la función covarianza y el operador covarianza, y estudiaremos algunas propiedades de tales estimadores.

En ocasiones necesitamos trabajar con los autovalores y autofunciones del operador covarianza, por ello en el Capítulo 4 los estimaremos y estudiaremos algunas propiedades de los estimadores, incluyendo la normalidad asintótica.

En el Capítulo 5 introduciremos el concepto de componentes principales funcionales, que nos permiten reducir la dimensión de un conjunto de datos funcionales de dimensión infinita a uno pequeño de dimensión finita de una forma óptima.

En el Capítulo 6 nos centraremos en dos contrastes de hipótesis:

- Igualdad de medias: estudiaremos dos métodos, demostrando teoremas

que justifican la distribución asintótica bajo la hipótesis nula de los estadísticos correspondientes, así como su convergencia en probabilidad bajo la hipótesis alternativa.

- Igualdad de operadores covarianza: estudiaremos también la distribución asintótica bajo la hipótesis nula del estadístico correspondiente a este test.

Por último, en el Capítulo 7, implementaremos algunas de las técnicas mencionadas anteriormente en el lenguaje estadístico R.

Índice general

1. Definiciones y resultados preliminares sobre espacios de Hilbert	11
2. Definiciones y resultados preliminares sobre funciones aleatorias en L^2 y el operador covarianza	17
3. Estimación de la función media, función covarianza y operador covarianza	23
4. Estimación de los autovalores y autofunciones del operador covarianza	31
4.1. Estimación de los autovalores y las autofunciones	31
4.2. Normalidad asintótica de las autofunciones	32
5. Componentes Principales Funcionales	35
5.1. Un problema de maximización	35
5.2. Bases ortonormales empíricas óptimas	37
5.3. Determinación de p	38
6. Inferencia de las funciones media y covarianza para dos muestras	41
6.1. Igualdad de medias	41
6.2. Igualdad de operadores covarianza	50
7. Aplicación en \mathbb{R}	57
7.1. Bases B-spline	57
7.2. Datos funcionales en \mathbb{R}	58
7.3. Componentes principales funcionales en \mathbb{R}	60
7.4. Igualdad de medias en \mathbb{R}	61

Capítulo 1

Definiciones y resultados preliminares sobre espacios de Hilbert

En este capítulo introduciremos algunos conceptos fundamentales sobre la teoría de los operadores en un espacio de Hilbert y, a continuación, nos centraremos en el espacio L^2 de funciones de cuadrado integrable, puesto que dicho espacio es suficiente para manejar la mayoría de los procedimientos considerados en este trabajo.

Definición 1.1. *Un espacio métrico es un par (X, d) , donde X es un conjunto arbitrario no vacío y $d : X \times X \rightarrow \mathbb{R}$ una aplicación, llamada distancia o métrica, tal que, para cualesquiera $x, y, z \in X$, se verifica:*

1. $d(x, y) \geq 0$.
2. $d(x, y) = 0 \Leftrightarrow x = y$.
3. $d(x, y) = d(y, x)$.
4. $d(x, z) \leq d(x, y) + d(y, z)$.

Definición 1.2. *Sean (X, d) un espacio métrico y una sucesión $\{a_n\}_{n \in \mathbb{N}} \subseteq X$. Se dice que $\{a_n\}_{n \in \mathbb{N}}$ es una sucesión de Cauchy si para cada $\varepsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que*

$$d(a_n - a_m) < \varepsilon, \text{ si } n, m \geq n_0.$$

Definición 1.3. Un espacio métrico (X, d) se dice que es completo si toda sucesión de Cauchy converge a un elemento de X , es decir, existe un elemento del espacio que es el límite de la sucesión.

Definición 1.4. Sea V un conjunto no vacío. Supongamos que en V hay definida una operación suma, que denotaremos por $+$, y una operación producto por un escalar, que denotaremos por \cdot . Diremos que $(V, +, \cdot)$ es un espacio vectorial si se verifican las siguientes propiedades:

1. Propiedad asociativa ($+$): $(u + v) + w = u + (v + w)$, $\forall u, v, w \in V$.
2. Propiedad conmutativa: $u + v = v + u$, $\forall u, v \in V$.
3. Existencia de elemento neutro: $\exists 0 \in V$ tal que $0 + v = v$, $\forall v \in V$.
4. Existencia de elemento opuesto: $\forall v \in V$, $\exists (-v) \in V$ tal que $v + (-v) = 0$.
5. Propiedad distributiva I: $a \cdot (u + v) = a \cdot u + a \cdot v$, $\forall a \in \mathbb{R}$, $\forall u, v \in V$.
6. Propiedad distributiva II: $(a + b) \cdot v = a \cdot v + b \cdot v$, $\forall a, b \in \mathbb{R}$, $\forall v \in V$.
7. Propiedad asociativa (\cdot): $a \cdot (b \cdot v) = (a \cdot b) \cdot v$, $\forall a, b \in \mathbb{R}$, $\forall v \in V$.
8. Elemento unidad: $1 \cdot v = v$, $\forall v \in V$.

Definición 1.5. Sea V un espacio vectorial. Una aplicación $\langle \cdot, \cdot \rangle : V \times V \rightarrow E$ (E puede ser \mathbb{R} o \mathbb{C}) se dice que es un producto interno (o escalar) si satisface las siguientes propiedades:

1. (aditiva) $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$;
2. (homogénea) $\langle \alpha x, y \rangle = \alpha \langle x, y \rangle$;
3. (hermítica) $\langle x, y \rangle = \overline{\langle y, x \rangle}$;
4. (definida positiva) $\langle x, x \rangle \geq 0$ y $\langle x, x \rangle = 0 \Leftrightarrow x = 0$.

Definición 1.6. Un espacio vectorial dotado de un producto escalar es un espacio prehilbertiano.

Nota 1.1. En los espacios prehilbertianos se define una norma

$$\|x\| = \sqrt{\langle x, x \rangle}.$$

Definición 1.7. Un espacio prehilbertiano que es completo para la norma inducida por el producto escalar es un espacio de Hilbert.

Definición 1.8. Sean H_1 y H_2 dos espacios prehilbertianos, un operador lineal (o transformación lineal) es una función $\Psi : H_1 \rightarrow H_2$ donde a cada elemento de H_1 le corresponde un único elemento de H_2 , y además cumple las siguientes condiciones:

1. $\forall x, y \in H_1$ se cumple que $\Psi(x + y) = \Psi(x) + \Psi(y)$.
2. Sean $\alpha \in \mathbb{R}, x \in H_1$, entonces se tiene que $\Psi(\alpha x) = \alpha \Psi(x)$.

Definición 1.9. Sean H_1 y H_2 dos espacios prehilbertianos dotados de la norma $\|\cdot\|_1$ y $\|\cdot\|_2$, respectivamente. Se define la norma del operador lineal $\Psi : H_1 \rightarrow H_2$ de la siguiente forma:

$$\|\Psi\| = \sup_{x \neq 0} \frac{\|\Psi(x)\|_2}{\|x\|_1}.$$

Definición 1.10. Un operador lineal Ψ se dice que es acotado si $\|\Psi\| < \infty$.

Nota 1.2. En lo que sigue tomaremos $H_1 = H_2 = H$. Todos los resultados pueden ser extendidos fácilmente al caso en el que los operadores actúen entre espacios diferentes.

Observación 1.1. Denotamos por \mathcal{L} al espacio de operadores lineales acotados definidos sobre el espacio de Hilbert H con la norma

$$\|\Psi\|_{\mathcal{L}} = \sup\{\|\Psi(x)\| : \|x\| \leq 1\}.$$

Definición 1.11. Un operador $\Psi \in \mathcal{L}$ se dice que es simétrico si

$$\langle \Psi(x), y \rangle = \langle x, \Psi(y) \rangle, \quad \forall x, y \in H.$$

Definición 1.12. $\Psi \in \mathcal{L}$ es definido positivo si

$$\langle \Psi(x), x \rangle \geq 0, \quad \forall x \in H.$$

Definición 1.13. Una base ortonormal de un espacio de Hilbert H es una base $\{e_j\}_{j \in J}$ de H satisfaciendo:

1. Los elementos están normalizados: $\|e_j\| = 1, \forall j$.
2. Los elementos son ortogonales: $\langle e_j, e_k \rangle = 0, \forall k, j$ tales que $j \neq k$.

Definición 1.14. Una base $\{e_j\}_{j \in J}$ se dice que es numerable cuando existe una aplicación inyectiva de J en \mathbb{N} . Denotaremos a las bases numerables como $\{e_n\}_{n \in \mathbb{N}}$.

Definición 1.15. Un espacio de Hilbert H es separable si admite una base ortonormal numerable.

Definición 1.16. Un operador $\Psi \in \mathcal{L}$ se dice compacto si existen dos bases ortonormales numerables $\{v_n\}_{n \in \mathbb{N}}$ y $\{f_n\}_{n \in \mathbb{N}}$ en H , y $\{\lambda_n\}_{n \in \mathbb{N}}$ una sucesión numerable de escalares que tiende a 0, tal que

$$\Psi(x) = \sum_{n \in \mathbb{N}} \lambda_n \langle x, v_n \rangle f_n, \quad x \in H. \quad (1.1)$$

Podemos suponer sin pérdida de generalidad que los λ_n son todos positivos y si es necesario se cambia f_n por $-f_n$.

Lema 1.1. (Gohberg 1990, página 99) Supongamos que $C, K \in \mathcal{L}$ son dos operadores compactos que admiten la descomposición

$$C(x) = \sum_{n \in \mathbb{N}} \lambda_n \langle x, v_n \rangle f_n, \quad K(x) = \sum_{n \in \mathbb{N}} \gamma_n \langle x, u_n \rangle g_n. \quad (1.2)$$

Entonces, para cada $n \in \mathbb{N}$, $|\gamma_n - \lambda_n| \leq \|K - C\|_{\mathcal{L}}$.

Lema 1.2. (Horváth and Kokoszka 2012, página 34) Supongamos que $C, K \in \mathcal{L}$ son dos operadores compactos con la descomposición (1.2). Si C es simétrico, es decir, $f_n = v_n$ en (1.2), y $\lambda_1 > \dots > \lambda_p > \lambda_{p+1}$, entonces

$$\|u_n - v'_n\| \leq \frac{2\sqrt{2}}{\alpha_n} \|K - C\|_{\mathcal{L}}, \quad 1 \leq n \leq p,$$

donde $\alpha_1 = \lambda_1 - \lambda_2$ y $\alpha_n = \min(\lambda_{n-1} - \lambda_n, \lambda_n - \lambda_{n+1})$, $2 \leq n \leq p$ y $v'_n = \text{signo}(\langle u_n, v_n \rangle) v_n$.

Proposición 1.1. Sea $\{e_j\}_{j \in J}$ es una base ortonormal de H , entonces

$$\|x\|^2 = \langle x, x \rangle = \sum_{j \in J} |\langle x, e_j \rangle|^2.$$

A la identidad anterior se le denomina Identidad de Parseval.

Definición 1.17. El espacio $L^2 = L^2([0, 1])$ es el conjunto de funciones reales medibles x definidas en $[0, 1]$ satisfaciendo $\int_0^1 x^2(t)dt < \infty$.

Proposición 1.2. El espacio L^2 es un espacio de Hilbert con el producto interno

$$\langle x, y \rangle = \int x(t)y(t)dt.$$

Proposición 1.3. L^2 es separable.

Observación 1.2. En este trabajo consideraremos operadores lineales $\Psi : L^2 \rightarrow L^2$, que son de la forma (llamados operadores integrales):

$$\Psi(x)(t) = \int \psi(t, s)x(s)ds, \quad x \in L^2,$$

donde $\psi : [0, 1]^2 \rightarrow \mathbb{R}$, denominado núcleo, satisface:

1. $\iint \psi(t, s)dt ds < \infty$.
2. $\psi(t, s) = \psi(s, t)$.
3. $\sum_{i,j=1}^n \psi(t_i, t_j)c_i c_j \geq 0, \quad \forall c_1, \dots, c_n \in \mathbb{R}, \forall t_1, \dots, t_n \in [0, 1]$.

Definición 1.18. Sea $\Psi : L^2 \rightarrow L^2$. Se dice que $\lambda \in \mathbb{C}$ es un autovalor (o valor propio) de Ψ si existe un función v no nula tal que $\Psi(v) = \lambda v$. La función v se llama autofunción (o función propia) de Ψ asociada al autovalor λ .

Proposición 1.4. (Riesz and Sz.-Nagy 1990, página 243) Sea $\Psi : L^2 \rightarrow L^2$ como en la Observación 1.2. Se tiene que

$$\Psi(t, s) = \sum_{n \in \mathbb{N}} \lambda_n v_n(t)v_n(s) \quad \text{en } L^2([0, 1] \times [0, 1])$$

donde $\{v_n\}_{n \in \mathbb{N}}$ y $\{\lambda_n\}_{n \in \mathbb{N}}$ son las autofunciones y autovalores del operador Ψ , es decir, $\Psi(v_n) = \lambda_n v_n, n \in \mathbb{N}$.

Teorema 1.1 (Teorema de Mercer). *(Riesz and Sz.-Nagy 1990, página 245) Si Ψ es continuo, la expresión anterior es válida para todo $s, t \in [0, 1]$, y la serie converge uniformemente.*

Observación 1.3. *Los operadores considerados pertenecen a una clase más general de operadores, denominados operadores de Hilbert-Schmidt, que a su vez tienen estructura de espacio de Hilbert con el producto escalar*

$$\langle \Psi_1, \Psi_2 \rangle_{\mathcal{S}} = \sum_{j \in J} \langle \Psi_1(e_j), \Psi_2(e_j) \rangle, \quad (1.3)$$

donde $\{e_j\}_{j \in J}$ es una base ortonormal. Se cumple que $\langle \cdot, \cdot \rangle_{\mathcal{S}}$ no depende de la base.

Proposición 1.5. *Un operador Hilbert-Schmidt Ψ simétrico y definido positivo admite la descomposición*

$$\Psi(x) = \sum_{n \in \mathbb{N}} \lambda_n \langle x, v_n \rangle v_n, \quad x \in H \quad (1.4)$$

con v_n ortonormales, que son las autofunciones de Ψ , es decir, $\Psi(v_n) = \lambda_n v_n$.

Proposición 1.6. *Para los operadores integrales considerados en este trabajo (ver Observación 1.2) se tiene que*

$$\|\Psi\|_{\mathcal{S}}^2 = \iint \psi^2(t, s) dt ds = \sum_{n \in \mathbb{N}} \lambda_n^2 < \infty.$$

Proposición 1.7. *(Gohberg 1990, página 144)*

$$\|\Psi\|_{\mathcal{L}} \leq \|\Psi\|_{\mathcal{S}}. \quad (1.5)$$

Teorema 1.2 (Teorema de representación de Riesz). *(Riesz and Sz.-Nagy 1990, página 61) Sea $\Psi : L^2 \rightarrow L^2$ un operador acotado. Entonces existe un único $y \in L^2$ tal que*

$$\Psi(x) = \langle x, y \rangle, \quad \forall x \in L^2.$$

Capítulo 2

Definiciones y resultados preliminares sobre funciones aleatorias en L^2 y el operador covarianza

En este capítulo presentaremos conceptos y resultados estadísticos para funciones. Además introduciremos el operador covarianza, que será necesario para desarrollar la mayoría de las técnicas que veremos en esta memoria.

Definición 2.1. Una familia de subconjuntos de X , representada por Σ , es un σ -álgebra sobre X cuando se cumplen las siguientes propiedades:

1. $\emptyset \in \Sigma$.
2. Si E está en Σ , también está su complementario $\bar{E} = X \setminus E$.
3. Si E_1, E_2, E_3, \dots es una sucesión numerable de elementos de Σ , entonces la unión de todos ellos también está en Σ .

Definición 2.2. Sea Σ el σ -álgebra generado por la clase de todos los subconjuntos abiertos de L^2 . Entonces Σ se denomina σ -álgebra de Borel en L^2 .

Definición 2.3. Un espacio de probabilidad es una tripleta (Ω, \mathcal{A}, P) , donde Ω (llamado espacio muestral) es un conjunto, $\mathcal{A} \subseteq P(\Omega)$ es un σ -álgebra sobre Ω y $P : \mathcal{A} \rightarrow [0, 1]$ satisface

1. $P(\Omega) = 1$.
2. $0 \leq P(E) \leq 1, \forall E \in \mathcal{A}$.
3. Si E_1, \dots, E_n son sucesos disjuntos, es decir, tales que $E_i \cap E_j = \emptyset$ para todo $i \neq j$, entonces $P(\cup_{i=1}^n E_i) = P(E_1) + P(E_2) + \dots + P(E_n)$.

Definición 2.4. Sea X una aplicación de un espacio probabilístico (Ω, \mathcal{A}, P) en un espacio métrico (S, \mathcal{S}) . Si X es medible, es decir, $X^{-1}V \in \mathcal{A} \forall V \in \mathcal{S}$, entonces se dice que es un elemento aleatorio. Diremos que X está definido en su dominio Ω y lo llamaremos elemento aleatorio de S .

Nota 2.1. 1. Si $S = \mathbb{R}^1$, se llama a X variable aleatoria;

2. Si $S = \mathbb{R}^k$, se dice que X es un vector aleatorio;

Nota 2.2. En esta memoria trabajemos con elementos aleatorios que son funciones de L^2 . Hablaremos de funciones aleatorias.

Definición 2.5. Se dice que X es integrable si

$$E\|X\| = E \left[\int X^2(t) dt \right]^{1/2} < \infty.$$

Proposición 2.1. Si X es integrable, existe una única función $\mu \in L^2$ tal que $E\langle y, X \rangle = \langle y, \mu \rangle$ para todo $y \in L^2$, con $\mu(t) = E[X(t)]$ para casi todo $t \in [0, 1]$.

Demostración. La demostración de este teorema es consecuencia directa del Teorema de representación de Riesz (ver Teorema 2.2). \square

Proposición 2.2. (Laha and Rohatgi 1979, página 449) Si $\Psi \in \mathcal{L}$ y X es integrable entonces $E\Psi(X) = \Psi E(X)$.

Definición 2.6. Si X es de cuadrado integrable, i.e.

$$E\|X\|^2 = E \int X^2(t) dt < \infty$$

y $EX = 0$, se define el operador covarianza $C : L^2 \rightarrow L^2$ de X como

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

Proposición 2.3. *Sea X una función aleatoria de cuadrado integrable con $EX = 0$, entonces*

$$C(y)(t) = \int c(t, s)y(s)ds, \quad \text{donde } c(t, s) = E[X(t)X(s)].$$

Demostración. Usando las Proposición 1.2 tenemos que

$$\begin{aligned} C(y)(t) &= E[\langle X, y \rangle X(t)] = E\left[\int X(s)y(s)dsX(t)\right] = \\ &= \int E[X(s)X(t)]y(s)ds = \\ &= \int c(t, s)y(s)ds. \end{aligned}$$

□

Nota 2.3. *Si $EX \neq 0$ se tiene que $c(t, s) = E[\{X(t) - \mu(t)\}\{X(s) - \mu(s)\}]$.*

Proposición 2.4. *Sea X una función aleatoria de cuadrado integrable con $EX = 0$, entonces se satisfacen las siguientes propiedades:*

1. $c(t, s) = c(s, t)$.
2. $\iint c(t, s)y(t)y(s)dtds \geq 0$.

Demostración. 1. Trivial.

$$\begin{aligned} 2. \iint c(t, s)y(t)y(s)dtds &= \iint E[X(t)X(s)]y(t)y(s)dtds = \\ &= E\left[\left(\int X(t)y(t)dt\right)^2\right] \geq 0. \end{aligned}$$

□

Observación 2.1. *De la proposición anterior se deduce que el operador covarianza C es simétrico y definido positivo.*

Proposición 2.5. *Sea X una función aleatoria con $E\|X\|^2 < \infty$ y $EX = 0$. Sean $\{\lambda_n\}_{n \in \mathbb{N}}$ los autovalores del operador covarianza asociados a X , entonces*

1. $\lambda_n \geq 0, \forall n \in \mathbb{N}$.

$$2. \sum_{n \in \mathbb{N}} \lambda_n < \infty.$$

Demostración. 1. Sean $\{v_n\}_{n \in \mathbb{N}}$ las autofunciones del operador covarianza C . La relación $C(v_n) = \lambda_n v_n$ implica que

$$\lambda_n = \langle C(v_n), v_n \rangle = \langle E[\langle X, v_n \rangle X], v_n \rangle = E[\langle X, v_n \rangle^2] \geq 0.$$

2. Las autofunciones $\{v_n\}_{n \in \mathbb{N}}$ son ortogonales, y pueden ser normalizadas para tener norma unitaria, así que $\{v_n\}_{n \in \mathbb{N}}$ forma una base en L^2 . Consecuentemente, por la Identidad de Parseval,

$$\sum_{n \in \mathbb{N}} \lambda_n = \sum_{n \in \mathbb{N}} E[\langle X, v_n \rangle^2] = E\|X\|^2 < \infty.$$

□

Proposición 2.6. (*Bosq 2000, página 36*) $C \in \mathcal{L}(L^2)$ es un operador covarianza si y solo si es simétrico, definido positivo y sus autovalores satisfacen $\sum_{n \in \mathbb{N}} \lambda_n < \infty$.

Proposición 2.7. (*Laha and Rohatgi 1979, página 472*) El operador covarianza C es compacto.

Definición 2.7. Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de funciones aleatorias en L^2 . Diremos que $\{X_n\}_{n \in \mathbb{N}}$ converge en probabilidad a una función aleatoria X si

$$P(\|X_n - X\| > \varepsilon) \rightarrow 0, \quad \forall \varepsilon > 0,$$

lo que representamos por $X_n \xrightarrow{P} X$.

Definición 2.8. Una sucesión de funciones aleatorias, $\{X_n\}_{n \in \mathbb{N}}$, converge de forma casi segura a X cuando se cumple que:

$$P(\lim_{n \rightarrow \infty} \|X_n - X\| = 0) = 1,$$

lo que representamos por $X_n \xrightarrow{c.s.} X$.

Definición 2.9. Sea X una función aleatoria se define la distribución de X como la medida de probabilidad $P = PX^{-1}$ en (S, \mathcal{S}) :

$$P(A) = P(X^{-1}A) = P(X \in A), \quad A \in \mathcal{S}.$$

Definición 2.10. Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de funciones aleatorias en L^2 . Diremos que $\{X_n\}_{n \in \mathbb{N}}$ converge en distribución a una función aleatoria X si la distribución P_n de X_n converge a la distribución P de X , es decir, si

$$\lim_{n \rightarrow \infty} P_n = P,$$

lo que representamos por $X_n \xrightarrow{\mathcal{D}} X$.

Teorema 2.1 (Teorema de Slutsky). (Serfling 2002, página 19) Sean $\{X_n\}_{n \in \mathbb{N}}$, $\{Y_n\}_{n \in \mathbb{N}}$ sucesiones de funciones aleatorias. Si X_n converge en distribución a una función aleatoria X ; e Y_n converge en probabilidad a una constante c , entonces

1. $X_n + Y_n \xrightarrow{\mathcal{D}} X + c$;
2. $X_n Y_n \xrightarrow{\mathcal{D}} cX$;
3. $X_n / Y_n \xrightarrow{\mathcal{D}} X/c$, siempre que $c \neq 0$.

Nota 2.4. En el enunciado del teorema de Slutsky, la condición “ Y_n converge en probabilidad a una constante c ” puede ser reemplazada por “ Y_n converge en distribución a una constante c ”. Estas dos condiciones son equivalentes debido a propiedades de la convergencia de funciones aleatorias.

Teorema 2.2 (Teorema de la aplicación continua). (Billingsley 1999, página 29) Sea $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de funciones aleatorias en L^2 y X una función aleatoria. Si $X_n \xrightarrow{\mathcal{D}} X$ y $P(X \in D_h) = 0$, entonces $h(X_n) \xrightarrow{\mathcal{D}} h(X)$, donde h es un conjunto medible y D_h es el conjunto de puntos de discontinuidad de h .

Definición 2.11. Una función aleatoria X se dice que es Gaussiana si para cada $y \in L^2$ la función aleatoria $\langle X, y \rangle$ se distribuye según una normal.

Definición 2.12. Un proceso estocástico $\Gamma = \{\Gamma(t)\}_{t \in T}$, donde T es un conjunto arbitrario, tal que $E\Gamma(t)^2 < \infty$, se denomina proceso Gaussiano si todas las funciones aleatorias

$$\sum_{i=1}^k \alpha_i \Gamma(t_i), \quad \alpha_1, \dots, \alpha_k \in L^2, \quad t_1, \dots, t_k \in T,$$

son funciones aleatorias Gaussianas.

Teorema 2.3 (Teorema Central del Límite). (Bosq 2000, página 51) Supongamos que $\{X_i\}_{i=1}^N$ es una sucesión de funciones aleatorias independientes e idénticamente distribuidos (iid) con media cero en un espacio de Hilbert tal que $E\|X_i\|^2 < \infty$. Entonces

$$N^{-1/2} \sum_{i=1}^N X_i \xrightarrow{\mathcal{D}} Z,$$

donde Z es una función aleatoria Gaussiana con el operador covarianza

$$C(x) = E[\langle Z, x \rangle Z] = E[\langle X_1, x \rangle X_1].$$

Teorema 2.4 (Teorema de Karhunen-Loève). (Bosq 2000, página 25) Sea $\Gamma = \{\Gamma(t)\}_{0 \leq t \leq 1}$ un proceso Gaussiano en L^2 , con un operador covarianza C , tal que $E\Gamma(t)^2 < \infty$. Entonces

$$\Gamma(t) = \sum_{n \in \mathbb{N}} \sqrt{\lambda_n} N_n v_n(t), \quad t \in [0, 1], \quad (2.1)$$

donde N_n son variables aleatorias independientes distribuidas según $N(0, 1)$ y $\{\lambda_n\}_{n \in \mathbb{N}}$ y $\{v_n\}_{n \in \mathbb{N}}$ son los autovalores y autofunciones del operador C .

Teorema 2.5 (Ley Fuerte de los Grandes Números). (Bosq 2000, página 47) Supongamos que $\{X_i\}_{i=1}^N$ es una sucesión de funciones aleatorias iid en un espacio de Hilbert tal que $E\|X_i\|^2 < \infty$. Entonces

$$N^{-1} \sum_{i=1}^N X_i \xrightarrow{c.s.} \mu.$$

Capítulo 3

Estimación de la función media, función covarianza y operador covarianza

En las aplicaciones, observamos una muestra que consiste en N funciones X_1, \dots, X_N . Veremos cada función como una realización de una función aleatoria X , las cuales tienen la misma distribución que esa función X .

Para demostrar los resultados de este capítulo supondremos que

Condición 3.1. *Las observaciones X_1, X_2, \dots, X_N son iid en L^2 y tienen la misma distribución que X , que supondremos de cuadrado integrable.*

Definición 3.1. *Dada una muestra X_1, \dots, X_n , iid a X y un estimador $T(X_1, \dots, X_n)$, del parámetro poblacional θ , el sesgo es:*

$$E(T) - \theta.$$

Definición 3.2. *Un estimador se dirá que es insesgado si*

$$E(T) = \theta.$$

Definimos los siguientes parámetros:

$$\mu(t) = E[X(t)] \text{ (función media);}$$

$$c(t, s) = E[\{X(t) - \mu(t)\}\{X(s) - \mu(s)\}] \text{ (función covarianza);}$$

$C = E[\langle (X - \mu), \cdot \rangle (X - \mu)]$ (operador covarianza).

Consideramos los siguientes estimadores:

$$\hat{\mu}(t) = N^{-1} \sum_{i=1}^N X_i(t) \text{ (media muestral);}$$

$$\hat{c}(t, s) = N^{-1} \sum_{i=1}^N [X_i(t) - \hat{\mu}(t)][X_i(s) - \hat{\mu}(s)] \text{ (covarianza muestral);}$$

$$\hat{C}(x) = N^{-1} \sum_{i=1}^N \langle X_i - \hat{\mu}, x \rangle (X_i - \hat{\mu}), \quad x \in L^2 \text{ (operador covarianza muestral).}$$

El objetivo de este capítulo será estudiar las propiedades de estos estimadores.

Lema 3.1. Si $X_1, X_2 \in L^2$ son independientes, de cuadrado integrable y $EX_1 = 0$ entonces $E[\langle X_1, X_2 \rangle] = 0$.

Demostración. Sabemos que como $X_1, X_2 \in L^2$ son independientes

$$E[X_1(s)X_2(s)] = E[X_1(s)]E[X_2(s)].$$

Es claro que como $EX_1 = 0$ se tiene que

$$\begin{aligned} E[\langle X_1, X_2 \rangle] &= E \int X_1(s)X_2(s)ds = \int E[X_1(s)X_2(s)]ds = \\ &= \int E[X_1(s)]E[X_2(s)]ds = 0. \end{aligned}$$

□

Teorema 3.1. Si se satisface la Condición 3.1, entonces $E\hat{\mu} = \mu$ (i.e., $\hat{\mu}$ es un estimador insesgado de μ) y $E\|\hat{\mu} - \mu\|^2 = O(N^{-1})$.

Demostración. Para cada i , y para casi todo $t \in [0, 1]$, $EX_i(t) = \mu(t)$, por lo tanto $E\hat{\mu} = \mu$ en L^2 . Observemos que

$$E\|\hat{\mu} - \mu\|^2 = N^{-2} \sum_{i,j=1}^N E[\langle (X_i - \mu), (X_j - \mu) \rangle].$$

Por el Lema 3.1, si $i \neq j$ entonces $E[\langle (X_i - \mu), (X_j - \mu) \rangle] = 0$ y por consiguiente

$$E\|\hat{\mu} - \mu\|^2 = N^{-2} \sum_{i=1}^N E\|X_i - \mu\|^2 = N^{-1} E\|X - \mu\|^2.$$

□

Proposición 3.1. *Bajo la condición $E\|X - \mu\|^2 < \infty$ se tiene que*

$$c(\cdot, \cdot) \in L^2([0, 1] \times [0, 1]).$$

Demostración.

$$\iint c^2(t, s) dt ds = \iint E[(X(t) - \mu(t))(X(s) - \mu(s))]^2 dt ds.$$

Se tiene por la desigualdad de Cauchy-Schwartz que

$$\begin{aligned} & \iint E[(X(t) - \mu(t))(X(s) - \mu(s))]^2 dt ds \leq \\ & \leq \iint E[(X(t) - \mu(t))^2] E[(X(s) - \mu(s))^2] dt ds = \\ & = \left(\int E[(X(t) - \mu(t))^2] dt \right)^2 = \\ & = \left(E \int (X(t) - \mu(t))^2 dt \right)^2 = \\ & = (E\|X - \mu\|^2)^2. \end{aligned}$$

□

Proposición 3.2. *Si se satisface la Condición 3.1, entonces $\hat{c}(t, s)$ no es un estimador insesgado de $c(t, s)$.*

Demostración. Por la Nota 2.3 tenemos que

$$\begin{aligned} c(t, s) &= E[\{X(t) - \mu(t)\}\{X(s) - \mu(s)\}] = \\ &= E[X(t)X(s)] - E[X(t)]\mu(s) - E[X(s)]\mu(t) + \mu(t)\mu(s) = \\ &= E[X(t)X(s)] - \mu(t)\mu(s) - \mu(s)\mu(t) + \mu(t)\mu(s) = \\ &= E[X(t)X(s)] - \mu(t)\mu(s). \end{aligned}$$

Por otro lado, sabemos que

$$\begin{aligned}
E[\hat{c}(t, s)] &= E \left[N^{-1} \sum_{i=1}^N \{X_i(t) - \hat{\mu}(t)\} \{X_i(s) - \hat{\mu}(s)\} \right] = \\
&= E \left[N^{-1} \sum_{i=1}^N \{X_i(t)X_i(s) - X_i(t)\hat{\mu}(s) - X_i(s)\hat{\mu}(t) + \hat{\mu}(t)\hat{\mu}(s)\} \right] = \\
&= E \left[\frac{1}{N} \sum_{i=1}^N X_i(t)X_i(s) \right] - E \left[\frac{1}{N} \sum_{i=1}^N X_i(t)\hat{\mu}(s) \right] - \\
&\quad - E \left[\frac{1}{N} \sum_{i=1}^N X_i(s)\hat{\mu}(t) \right] + E \left[\frac{1}{N} \sum_{i=1}^N \hat{\mu}(t)\hat{\mu}(s) \right].
\end{aligned}$$

Desarrollemos cada uno de los sumandos de la expresión anterior:

- $E \left[\frac{1}{N} \sum_{i=1}^N X_i(t)X_i(s) \right] = E[X(t)X(s)] = c(t, s) + \mu(t)\mu(s).$
- $E \left[\frac{1}{N} \sum_{i=1}^N X_i(t)\hat{\mu}(s) \right] =$
 $= E[\hat{\mu}(t)\hat{\mu}(s)] =$
 $= \frac{1}{N^2} \sum_{i,j=1}^N E[X_i(t)X_j(s)] =$
 $= \frac{1}{N^2} \sum_{i,j=1}^N \begin{cases} E[X_i(t)X_j(s)] = E[X_i(t)]E[X_j(s)] = \mu(t)\mu(s) & \text{si } i \neq j \\ E[X_i(t)X_i(s)] = c(t, s) + \mu(t)\mu(s) & \text{si } i = j \end{cases} =$
 $= \frac{1}{N^2} [N(N-1)\mu(t)\mu(s) + Nc(t, s) + N\mu(t)\mu(s)] =$
 $= \mu(t)\mu(s) + \frac{1}{N}c(t, s).$

- Del mismo modo que en el caso anterior

$$E \left[\frac{1}{N} \sum_{i=1}^N X_i(s)\hat{\mu}(t) \right] = E[\hat{\mu}(s)\hat{\mu}(t)] = \mu(t)\mu(s) + \frac{1}{N}c(t, s).$$

- $E \left[\frac{1}{N} \sum_{i=1}^N \hat{\mu}(t)\hat{\mu}(s) \right] = E[\hat{\mu}(t)\hat{\mu}(s)].$

Por tanto,

$$\begin{aligned}
 E[\hat{c}(t, s)] &= E \left[\frac{1}{N} \sum_{i=1}^N X_i(t) X_i(s) \right] - E[\hat{\mu}(t)\hat{\mu}(s)] - E[\hat{\mu}(t)\hat{\mu}(s)] + E[\hat{\mu}(t)\hat{\mu}(s)] = \\
 &= E \left[\frac{1}{N} \sum_{i=1}^N X_i(t) X_i(s) \right] - E[\hat{\mu}(t)\hat{\mu}(s)] = \\
 &= c(t, s) + \mu(t)\mu(s) - \mu(t)\mu(s) - \frac{1}{N}c(t, s) = \\
 &= \frac{N-1}{N}c(t, s) \quad (\text{en } L^2([0, 1] \times [0, 1])).
 \end{aligned}$$

□

Observación 3.1. *El sesgo de \hat{c} es asintóticamente despreciable.*

Nota 3.1. *Reemplazar μ por $\hat{\mu}$ tiene, en general, un efecto insignificante, y en el trabajo teórico, es conveniente suponer que μ es conocida e igual a cero. Esto simplifica muchas fórmulas.*

Desde ahora, excepto cuando se diga explícitamente, asumiremos que las observaciones tienen media cero. Tenemos por lo tanto

$$\hat{c}(t, s) = N^{-1} \sum_{i=1}^N X_i(t) X_i(s); \quad \hat{C}(x) = N^{-1} \sum_{i=1}^N \langle X_n, x \rangle X_n$$

y

$$\hat{C}(x)(t) = \int \hat{c}(t, s)x(s)ds, \quad x \in L^2.$$

Teorema 3.2. *Si $E\|X\|^4 < \infty$, $EX = 0$, y se satisface la Condición 3.1, entonces*

$$E\|\hat{C}\|_{\mathcal{S}}^2 \leq E\|X\|^4.$$

Demostración. Por la desigualdad triangular

$$E\|\hat{C}\|_{\mathcal{S}}^2 = E \left\| N^{-1} \sum_{i=1}^N \langle X_i, \cdot \rangle X_i \right\|_{\mathcal{S}}^2 \leq E\|\langle X, \cdot \rangle X\|_{\mathcal{S}}^2.$$

Por (1.3), para toda base ortonormal $\{e_n\}_{n \in \mathbb{N}}$,

$$\|\langle X, \cdot \rangle X\|_{\mathcal{S}}^2 = \sum_{n \in \mathbb{N}} \|\langle X, e_n \rangle X\|^2 = \|X\|^2 \sum_{n \in \mathbb{N}} |\langle X, e_n \rangle|^2 = \|X\|^4.$$

Donde la última igualdad se tiene por la Identidad de Parseval. \square

Lema 3.2. *Sea X una función aleatoria en L^2 con media μ y $E\|X\|^2 \leq \infty$, entonces*

$$E\|X - \mu\|^2 \leq E\|X\|^2.$$

Demostración.

$$\begin{aligned} \|X - \mu\|^2 &= \langle X - \mu, X - \mu \rangle = \langle X, X \rangle - \langle X, \mu \rangle - \langle \mu, X \rangle + \langle \mu, \mu \rangle = \\ &= \|X\|^2 - \langle X, \mu \rangle - \langle \mu, X \rangle + \|\mu\|^2. \end{aligned}$$

Por la Proposición 2.1 sabemos que $E\langle X, \mu \rangle = \langle \mu, \mu \rangle = \|\mu\|^2$. Luego

$$\begin{aligned} E\|X - \mu\|^2 &= E\|X\|^2 - \|\mu\|^2 - \|\mu\|^2 + \|\mu\|^2 = \\ &= E\|X\|^2 - \|\mu\|^2 \leq \\ &\leq E\|X\|^2. \end{aligned}$$

La última desigualdad se tiene ya que $\|\mu\|^2 \geq 0$. \square

Teorema 3.3. *Si $E\|X\|^4 < \infty$, $EX = 0$ y se satisface la Condición 3.1, entonces*

$$E\|\hat{C} - C\|_S^2 \leq N^{-1}E\|X\|^4.$$

Demostración. Por (1.3), para toda base ortonormal $\{e_n\}_{n \in \mathbb{N}}$,

$$\begin{aligned} \|\hat{C} - C\|_S^2 &= \sum_{n \in \mathbb{N}} \left\| \frac{1}{N} \sum_{i=1}^N \langle X_i, e_n \rangle X_i - E[\langle X, e_n \rangle X] \right\|^2 = \\ &= \sum_{n \in \mathbb{N}} \left\langle \frac{1}{N} \sum_{i=1}^N \{ \langle X_i, e_n \rangle X_i - E[\langle X_i, e_n \rangle X_i] \}, \right. \\ &\quad \left. \frac{1}{N} \sum_{j=1}^N \{ \langle X_j, e_n \rangle X_j - E[\langle X_j, e_n \rangle X_j] \} \right\rangle \\ &= \frac{1}{N^2} \sum_{n \in \mathbb{N}} \sum_{i=1}^N \sum_{j=1}^N \langle \{ \langle X_i, e_n \rangle X_i - E[\langle X_i, e_n \rangle X_i] \} \times \\ &\quad \times \{ \langle X_j, e_n \rangle X_j - E[\langle X_j, e_n \rangle X_j] \} \rangle. \end{aligned}$$

Por el Lema 3.1, para $i \neq j$,

$$E\langle \{\langle X_i, e_n \rangle X_i - E[\langle X_i, e_n \rangle X_i]\}, \{\langle X_j, e_n \rangle X_j - E[\langle X_j, e_n \rangle X_j]\} \rangle = 0.$$

Por tanto,

$$E\|\hat{C} - C\|_S^2 = \frac{1}{N^2} \sum_{n \in \mathbb{N}} \sum_{i=1}^N E\|\langle X_i, e_n \rangle X_i - E[\langle X_i, e_n \rangle X_i]\|^2.$$

Además por el Lema 3.2 tenemos que

$$\frac{1}{N^2} \sum_{n \in \mathbb{N}} \sum_{i=1}^N E\|\langle X_i, e_n \rangle X_i - E[\langle X_i, e_n \rangle X_i]\|^2 \leq \frac{1}{N} \sum_{n \in \mathbb{N}} E\|\langle X, e_n \rangle X\|^2.$$

Luego,

$$\begin{aligned} E\|\hat{C} - C\|_S^2 &\leq \frac{1}{N} \sum_{n \in \mathbb{N}} E\|\langle X, e_n \rangle X\|^2 = N^{-1} E \sum_{n \in \mathbb{N}} \|\langle X, e_n \rangle X\|^2 = \\ &= N^{-1} E \left[\|X\|^2 \sum_{n \in \mathbb{N}} |\langle X, e_n \rangle|^2 \right] = N^{-1} E\|X\|^4. \end{aligned}$$

la última igualdad se tiene por la Identidad de Parseval. □

Capítulo 4

Estimación de los autovalores y autofunciones del operador covarianza

4.1. Estimación de los autovalores y las autofunciones

A menudo tenemos que calcular los autovalores y las autofunciones de C , pero la interpretación de estos parámetros, y su estimación, debe ser tratada con cuidado. Los autovalores deben ser identificables, por lo que supondremos que $\lambda_1 > \lambda_2 > \dots$. En la práctica, sólo es posible estimar los p autovalores mayores, y asumiremos que $\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}$, lo que implica que los primeros p autovalores son distintos de cero. Las autofunciones $\{v_n\}_{n \in \mathbb{N}}$ están definidas por $Cv_n = \lambda_n v_n$, luego, si v_n es una autofunción, también lo es av_n , para cualquier escalar a no nulo. Los $\{v_n\}_{n \in \mathbb{N}}$ están normalizados, es decir, $\|v_n\| = 1$, pero esto no determina el signo de v_n . Por tanto, si \hat{v}_n es un estimador calculado a partir de los datos, únicamente podemos esperar que $\hat{c}_n \hat{v}_n$ sea cercano a v_n , donde

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

Hay que tener en cuenta que \hat{c}_n no puede ser calculado a partir de los datos, por lo que hay que asegurarse de que los estadísticos con los que queremos trabajar no dependen de \hat{c}_n .

Definimos las autofunciones y autovalores muestrales como

$$\int \hat{c}(t, s)\hat{v}_n(s)ds = \hat{\lambda}_n\hat{v}_n(t), \quad n = 1, 2, \dots, N.$$

El siguiente teorema implica que, bajo condiciones de regularidad, las autofunciones y autovalores muestrales estiman consistentemente las autofunciones y autovalores poblacionales.

Teorema 4.1. *Si $E\|X\|^4 < \infty$, $EX = 0$, se satisface la Condición 3.1 y*

$$\lambda_1 > \lambda_2 > \dots > \lambda_p > \lambda_{p+1}, \quad (4.1)$$

entonces, para cada $1 \leq n \leq p$,

$$\limsup_{N \rightarrow \infty} NE[\|\hat{c}_n\hat{v}_n - v_n\|^2] < \infty, \quad \limsup_{N \rightarrow \infty} NE[|\lambda_n - \hat{\lambda}_n|^2] < \infty.$$

Demostración. Por (1.5) se tiene que $\|\hat{C} - C\|_{\mathcal{L}} \leq \|\hat{C} - C\|_{\mathcal{S}}$. Por el Teorema 3.3, $E\|\hat{C} - C\|_{\mathcal{S}}^2 = O(N^{-1})$. Por lo que la prueba de este teorema es consecuencia de los Lemas 1.1 y 1.2. \square

4.2. Normalidad asintótica de las autofunciones

Para finalizar este capítulo daremos la normalidad asintótica de $N^{1/2}(\hat{c}_n\hat{v}_n - v_n)$ y de las variables aleatorias $N^{1/2}(\lambda_n - \hat{\lambda}_n)$. Sea

$$Z_N(t, s) = N^{1/2}(\hat{c}(t, s) - c(t, s)),$$

donde $\hat{c}(t, s), c(t, s)$ pueden estar centradas con la función media, o descentradas si la función media es asumida cero.

El siguiente teorema establece la convergencia en distribución de $Z_N(\cdot, \cdot)$ en el espacio $L^2([0, 1] \times [0, 1])$ asumiendo que las observaciones tienen media cero.

Teorema 4.2. *Si se satisface la Condición 3.1 con $EX(t) = 0$ y $E\|X\|^4 < \infty$, entonces $Z_N(t, s)$ converge en distribución en $L^2([0, 1] \times [0, 1])$ a un proceso Gaussiano $\Gamma(t, s)$ con $E\Gamma(t, s) = 0$ y*

$$E[\Gamma(t, s)\Gamma(t', s')] = E[X(t)X(s)X(t')X(s')] - c(t, s)c(t', s').$$

Demostración. Escribiendo

$$Z_N(t, s) = N^{-1/2} \sum_{i=1}^N [X_i(t)X_i(s) - c(t, s)],$$

observamos que $Z_N(t, s)$ es un proceso normalizado de sumas parciales independientes e idénticamente distribuidas, tomando valores en $L^2([0, 1] \times [0, 1])$. Entonces el Teorema Central del Límite en un espacio de Hilbert se satisface si

$$E \iint (X(t)X(s))^2 dt ds < \infty.$$

Esta condición se da porque

$$\begin{aligned} E \iint \{X(t)X(s)\}^2 dt ds &= E \int X^2(t) dt \int X^2(s) ds \leq \\ &\leq \left\{ E \left[\int X^2(t) dt \right]^2 \left[\int X^2(s) ds \right]^2 \right\}^{1/2} = \\ &= (E\|X\|^4)^2 < \infty. \end{aligned}$$

Si llamamos $Y_i(u) = X_i(t)X_i(s)$ tenemos que

$$\begin{aligned} Z_N(t, s) &= N^{-1/2} \sum_{i=1}^N [X_i(t)X_i(s) - E\{X_i(t)X_i(s)\}] = \\ &= N^{-1/2} \sum_{i=1}^N [Y_i(u) - E\{Y_i(u)\}]. \end{aligned}$$

Sabemos por el Teorema Central del Límite que Γ tiene la misma covarianza que Y_i , por lo tanto

$$\begin{aligned} cov[\Gamma(t, s), \Gamma(t', s')] &= E[\Gamma(t, s)\Gamma(t', s')] - E[\Gamma(t, s)]E[\Gamma(t', s')] = \\ &= E[\Gamma(t, s)\Gamma(t', s')] = \\ &= E[Y_i(u)Y_i(u')] - E[Y_i(u)]E[Y_i(u')] = \\ &= E[X_i(t)X_i(s)X_i(t')X_i(s')] - E[X_i(t)X_i(s)]E[X_i(t')X_i(s')] = \\ &= E[X_i(t)X_i(s)X_i(t')X_i(s')] - c(t, s)c(t', s'). \end{aligned}$$

□

Condición 4.1. La función aleatoria $X \in L^2$ que tiene la misma distribución que los X_i satisface las siguientes condiciones:

$$C1: \text{ Para todo } \kappa > 0, \sup_{0 \leq t \leq 1} E|X(t)|^\kappa < \infty,$$

C2: Existe $\gamma > 0$ tal que para todo $\kappa > 0$

$$\sup_{0 \leq t, s \leq 1} E[|t - s|^{-\gamma} |X(t) - X(s)|^\kappa] < \infty,$$

C3: Para cada entero $r \geq 0$, la secuencia $\{\lambda_n^{-1} E\langle X, v_n \rangle^{2r}\}$ está acotada.

Teorema 4.3. (Hall and Hosseini-Nasab 2006) Si se satisfacen las Condiciones 3.1 y 4.1 y además (4.1), entonces para $1 \leq j \leq p$,

$$N^{1/2}(\hat{\lambda}_n - \lambda_n) = \iint Z_N(t, s) v_n(t) v_n(s) dt ds + o_p(1)$$

y

$$\sup_{0 \leq t \leq 1} |N^{1/2}\{\hat{c}_n \hat{v}_n(t) - v_n(t)\} - \hat{T}_n(t)| = o_p(1),$$

donde

$$\hat{T}_n(t) = \sum_{k \neq n} (\lambda_n - \lambda_k)^{-1} v_k(t) \iint Z_N(t, s) v_n(t) v_n(s) dt ds.$$

Nota 4.1. Sean $\{X_n\}_{n \in \mathbb{N}}$ una sucesión de funciones aleatorias y X una función aleatorias, entonces:

1. Si $X_n \xrightarrow{P} 0$, escribimos $X_n = o_p(1)$,
2. Si $X_n \xrightarrow{D} X$, escribimos $X_n = O_p(1)$.

Capítulo 5

Componentes Principales Funcionales

El Análisis de Componentes Principales Funcionales (ACPF) es una técnica que permite reducir la dimensión de un conjunto de datos funciones de dimensión infinita a uno más pequeño de dimensión finita perdiendo la menor cantidad de información posible.

Un aspecto clave en el ACPF es la interpretación, ya que ésta no viene dada a priori, por lo que será vital importancia el conocimiento que el experto tenga sobre la materia de investigación.

5.1. Un problema de maximización

En esta sección presentaremos algunos resultados preliminares que son fundamentales para el resto del capítulo. Empezaremos con un caso vectorial, y luego nos iremos al espacio L^2 .

Teorema 5.1. *Sea \mathbf{A} una matriz simétrica $p \times p$. Entonces existe una matriz ortonormal $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_p]$ cuyas columnas son los autovectores de \mathbf{A} , es decir,*

$$\mathbf{U}^T \mathbf{U} = \mathbf{I} \text{ y } \mathbf{A} \mathbf{u}_n = \lambda_n \mathbf{u}_n.$$

Por otra parte,

$$\mathbf{U}^T \mathbf{A} \mathbf{U} = \mathbf{\Lambda} = \text{diag}[\lambda_1, \dots, \lambda_p].$$

La ortonormalidad de \mathbf{U} es equivalente a la afirmación de que los vectores $\mathbf{u}_1, \dots, \mathbf{u}_p$ forman una base ortonormal en el espacio R^p . El teorema anterior implica que $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, representación conocida como descomposición espectral de \mathbf{A} . Esto se usa para el siguiente problema de maximización.

Supongamos que \mathbf{A} es simétrica y definida positiva con autovalores distintos organizados en orden decreciente: $\lambda_1 > \dots > \lambda_p$. Queremos encontrar un vector unitario \mathbf{x} tal que $\mathbf{x}^T \mathbf{A} \mathbf{x}$ es máximo. Por la descomposición espectral, $\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{y}^T \mathbf{\Lambda} \mathbf{y}$, donde $\mathbf{y} = \mathbf{U}^T \mathbf{x}$. Como \mathbf{U} es ortonormal, $\|\mathbf{y}\| = \|\mathbf{x}\|$, es suficiente encontrar un vector unitario \mathbf{y} tal que $\mathbf{y} \mathbf{\Lambda} \mathbf{y}^T$ sea máximo, y por tanto establecemos $\mathbf{x} = \mathbf{U} \mathbf{y}$. Ya que $\mathbf{y} \mathbf{\Lambda} \mathbf{y}^T = \sum_{n=1}^p \lambda_n y_n^2$, claramente, el máximo se alcanza para $\mathbf{y} = [1, 0, \dots, 0]^T$, y $\mathbf{x} = \mathbf{u}_1$ con el máximo siendo λ_1 .

La idea anterior puede ser extendida a un espacio separable de Hilbert. Supongamos que Ψ es un operador Hilbert-Schmidt en L^2 simétrico y definido positivo. Hemos visto en el Capítulo 3 que el operador covarianza C y su equivalente operador covarianza muestral \hat{C} están en esta clase, siempre que $E\|X\|^4 < \infty$. El operador Ψ entonces admite la descomposición (1.4), y el problema de maximización $\langle \Psi(x), x \rangle$ sujeto a $\|x\| = 1$ es trivial porque

$$\langle \Psi(x), x \rangle = \left\langle \sum_{n \in \mathbb{N}} \lambda_n \langle x, v_n \rangle v_n, x \right\rangle = \sum_{n \in \mathbb{N}} \lambda_n \langle x, v_n \rangle^2.$$

Por la Identidad de Parseval, debemos maximizar la expresión anterior, sujeta a $\sum_{n \in \mathbb{N}} \langle x, v_n \rangle^2 = 1$. Para asegurarnos la unicidad, supongamos que $\lambda_1 > \lambda_2 > \dots$, por lo que tomaremos $\langle x, v_1 \rangle = 1$ y $\langle x, v_n \rangle = 0$ para $n > 1$. Por tanto, $\langle \Psi(x), x \rangle$ es máximo en $x = v_1$ (o $-v_1$), y este máximo es λ_1 . Supongamos ahora que queremos maximizar $\langle \Psi(x), x \rangle$ sujeto no solo a $\|x\| = 1$, sino también a $\langle x, v_1 \rangle = 0$. Por tanto queremos encontrar una función con norma unitaria que sea ortogonal a la función encontrada en el primer paso. Tal función claramente satisface $\langle \Psi(x), x \rangle = \sum_{n=2}^{\infty} \lambda_n \langle x, v_n \rangle^2$ y $\sum_{n=2}^{\infty} \langle x, v_n \rangle^2 = 1$, por lo que $x = v_2$, y el máximo ahora es λ_2 . Repitiendo este proceso, llegamos al siguiente teorema.

Teorema 5.2. *Supongamos que Ψ es un operador Hilbert-Schmidt simétrico y definido positivo con autofunciones v_n y autovalores λ_n , satisfaciendo $\lambda_1 > \dots > \lambda_p > \lambda_{p+1}$. Entonces*

$$\sup\{\langle \Psi(x), x \rangle : \|x\| = 1, \langle x, v_n \rangle = 0, 1 \leq n \leq i-1, i < p\} = \lambda_i,$$

y el supremo se alcanza si $x = v_i$, el cual es único salvo signo.

5.2. Bases ortonormales empíricas óptimas

En esta sección aplicaremos el desarrollo anterior al siguiente problema. Supongamos que observamos las funciones x_1, x_2, \dots, x_N . En esta sección no es necesario ver estas funciones como funciones aleatorias, pero podemos pensar en ellas como realizaciones observadas de funciones aleatorias en L^2 . Fijamos un entero p mucho más pequeño que N , normalmente será un número de un solo dígito. Queremos encontrar una base ortonormal u_1, u_2, \dots, u_p , tal que

$$\hat{S}^2 = \sum_{i=1}^N \left\| x_i - \sum_{k=1}^p \langle x_i, u_k \rangle u_k \right\|^2$$

sea mínimo. Una vez que hemos encontrado la base, podemos reemplazar cada función x_i por $\sum_{k=1}^p \langle x_i, u_k \rangle u_k$, la cual será una buena aproximación. En vez de trabajar con funciones x_i de dimensión infinita, podemos trabajar con vectores p -dimensionales

$$x_i = [\langle x_i, u_1 \rangle, \langle x_i, u_2 \rangle, \dots, \langle x_i, u_p \rangle]^T.$$

Las funciones u_n se denominan colectivamente *bases ortonormales empíricas óptimas* o *componentes ortonormales naturales*, las palabras “empírica” y “natural” hacen hincapié en que se calculan directamente a partir de los datos.

Las funciones u_1, u_2, \dots, u_p que minimizan \hat{S}^2 son iguales (salvo signo) a las autofunciones normalizadas del operador covarianza muestral. Para ver esto, supongamos primero que $p = 1$, en decir, queremos encontrar u con $\|u\| = 1$ que minimiza

$$\begin{aligned} \sum_{i=1}^N \|x_i - \langle x_i, u \rangle u\|^2 &= \sum_{i=1}^N \|x_i\|^2 - 2 \sum_{i=1}^N \langle x_i, u \rangle^2 + \sum_{i=1}^N \langle x_i, u \rangle^2 \|u\|^2 = \\ &= \sum_{i=1}^N \|x_i\|^2 - \sum_{i=1}^N \langle x_i, u \rangle^2, \end{aligned}$$

es decir, tenemos que maximizar $\sum_{i=1}^N \langle x_i, u \rangle^2 = \langle \hat{C}(u), u \rangle = \sum_{n \in \mathbb{N}} \hat{\lambda}_n \langle u, \hat{v}_n \rangle^2$. Por el Teorema 5.2, concluimos que $u = \hat{v}_1$.

El caso general se trata de manera análoga. Como

$$\hat{S}^2 = \sum_{i=1}^N \|x_i\|^2 - \sum_{i=1}^N \sum_{k=1}^p \langle x_i, u_k \rangle^2,$$

necesitamos maximizar

$$\begin{aligned} \sum_{i=1}^N \sum_{k=1}^p \langle x_i, u_k \rangle^2 &= \sum_{k=1}^p \langle \hat{C}(u_k), u_k \rangle = \\ &= \sum_{n \in \mathbb{N}} \hat{\lambda}_n \langle u_1, \hat{v}_n \rangle^2 + \sum_{n \in \mathbb{N}} \hat{\lambda}_n \langle u_2, \hat{v}_n \rangle^2 + \dots + \sum_{n \in \mathbb{N}} \hat{\lambda}_n \langle u_p, \hat{v}_n \rangle^2. \end{aligned}$$

Por el Teorema 5.2 la suma no puede exceder $\sum_{k=1}^p \hat{\lambda}_k$ y este máximo se alcanza si $u_1 = \hat{v}_1, u_2 = \hat{v}_2, \dots, u_p = \hat{v}_p$.

5.3. Determinación de p

Supongamos que X_1, X_2, \dots, X_N son observaciones funcionales. Las autofunciones del operador covarianza muestral \hat{C} se llaman componentes principales principales empíricas (CPFE) de los datos X_1, X_2, \dots, X_N . Si esas observaciones siguen la misma distribución que una función aleatoria X de cuadrado integrable, definimos las componentes funcionales principales (CFP) como las autofunciones del operador covarianza C .

Observamos que bajo la condición $EX_i = 0$, tenemos el estadístico

$$\frac{1}{N} \sum_{i=1}^N \langle X_i, x \rangle^2 = \langle \hat{C}(x), x \rangle.$$

Queremos encontrar x que maximice $\langle \hat{C}(x), x \rangle$. Claramente, hay que imponer una restricción en la norma de x , por lo que si exigimos que $\|x\| = 1$, tenemos por el Teorema 4.3 que $x = \hat{v}_1$ es la primera CFPE. Ahora, queremos encontrar otra x , ortogonal a \hat{v}_1 . Por el Teorema 4.3, se tiene que $x = \hat{v}_2$. Observamos que como los $\hat{v}_j, j = 1, \dots, N$, forman una base en \mathbb{R}^N ,

$$\frac{1}{N} \sum_{i=1}^N \|X_i\|^2 = \frac{1}{N} \sum_{i=1}^N \sum_{n=1}^N \langle X_i, \hat{v}_n \rangle^2 = \sum_{n=1}^N \frac{1}{N} \sum_{i=1}^N \langle X_i, \hat{v}_n \rangle^2 = \sum_{n=1}^N \hat{\lambda}_n.$$

En la mayoría de las aplicaciones, es importante determinar un valor p para el cual los datos se pueden sustituir por la aproximación $\sum_{n=1}^p \langle \hat{v}_n, X_i \rangle \hat{v}_n$. Un

método conocido es el “scree plot” (“scree” es una palabra inglesa que significa “sedimentación” (término geológico que hace referencia a los residuos que se acumulan en la parte inferior de una ladera rocosa)). Para aplicarlo se representan los sucesivos autovalores $\hat{\lambda}_n$ contra n . Este método sugiere encontrar n donde el decrecimiento de los autovalores parece estabilizarse. Este punto suele ser seleccionado como el valor p . A la derecha de este valor, se encuentra el “factor scree”. El método que funciona mejor para las aplicaciones que nos interesan en este trabajo es el “método CPV”. Elegimos un valor p para el cual $CPV(p)$ excede un valor deseado, 85% es un valor recomendado. El valor del $CPV(p)$ (porcentaje de varianza acumulado) se calcula de la siguiente forma

$$CPV(p) = \frac{\sum_{k=1}^p \hat{\lambda}_k}{\sum_{k=1}^N \hat{\lambda}_k} = \frac{\sum_{k=1}^p \hat{\lambda}_k}{\frac{1}{N} \sum_{i=1}^N \|X_i\|^2}.$$

Capítulo 6

Inferencia de las funciones media y covarianza para dos muestras

6.1. Igualdad de medias

En esta sección presentaremos procedimientos para contrastar la igualdad de las funciones media, junto con los teoremas que los justifican asintóticamente.

Consideramos dos muestras X_1, \dots, X_N y X_1^*, \dots, X_M^* . Supondremos que satisfacen el modelo

$$X_i(t) = \mu(t) + \varepsilon_i(t), \quad 1 \leq i \leq N$$

y

$$X_i^*(t) = \mu^*(t) + \varepsilon_i^*(t), \quad 1 \leq i \leq M.$$

Queremos realizar el contraste de hipótesis

$$H_0 : \mu = \mu^* \text{ en } L^2 \quad \text{vs} \quad H_1 : \mu \neq \mu^* \text{ en } L^2.$$

Para ello asumiremos que

$$\text{las dos muestras son independientes,} \tag{6.1}$$

$$\varepsilon_1, \dots, \varepsilon_N \text{ son iid con } E\varepsilon_1(t) = 0 \text{ y } E\|\varepsilon_1\|^4 < \infty \tag{6.2}$$

y de manera similar

$$\varepsilon_1^*, \dots, \varepsilon_M^* \text{ son iid con } E\varepsilon_1^*(t) = 0 \text{ y } E\|\varepsilon_1^*\|^4 < \infty. \quad (6.3)$$

Es importante resaltar que no estamos suponiendo que los ε_i^* tienen la misma distribución que los ε_i .

Para contrastar H_0 vs H_1 estudiaremos dos métodos.

Método I: Como

$$\bar{X}_N(t) = \frac{1}{N} \sum_{i=1}^N X_i(t) \text{ y } \bar{X}_M^*(t) = \frac{1}{M} \sum_{i=1}^M X_i^*(t)$$

son estimadores insesgados y consistentes de $\mu(t)$ y $\mu^*(t)$, respectivamente, es natural rechazar la hipótesis nula si el estadístico

$$U_{N,M} = \frac{NM}{N+M} \int_0^1 (\bar{X}_N(t) - \bar{X}_M^*(t))^2 dt$$

es “grande”, es decir, la región crítica es de la forma,

$$\text{RC: rechazar } H_0 \text{ si } U_{N,M} \geq c \text{ con } c \text{ tal que } P_{H_0}(U_{N,M} > c) = \alpha. \quad (6.4)$$

Es claro que no podemos determinar c satisfaciendo (6.4) porque no hemos hecho ninguna hipótesis distribucional y por tanto desconocemos la distribución nula de $U_{N,M}$. Para solucionar este problema aproximamos por su distribución asintótica bajo H_0 , que viene dada en el siguiente teorema.

Teorema 6.1. *Si se satisface H_0 (6.1)-(6.3) y*

$$\frac{N}{N+M} \rightarrow \theta \text{ con algún } \theta \in [0, 1], \quad (6.5)$$

entonces

$$U_{N,M} \xrightarrow{\mathcal{D}} \int_0^1 \Gamma^2(t) dt, \quad (6.6)$$

donde $\{\Gamma(t)\}_{0 \leq t \leq 1}$ es un proceso Gaussiano con $E\Gamma(t) = 0$ y

$$E[\Gamma(t)\Gamma(s)] = (1-\theta)c(t,s) + \theta c^*(t,s) = z(t,s), \quad (6.7)$$

donde $c(t,s) = \text{Cov}(X_1(t), X_1(s))$ y $c^*(t,s) = \text{Cov}(X_1^*(t), X_1^*(s))$.

Demostración. Es claro que

$$\bar{X}_N(t) - \bar{X}_M(t) = \{\bar{X}_N(t) - \mu(t)\} - \{\bar{X}_M^*(t) - \mu(t)\}.$$

Por el Teorema Central del Límite hay dos funciones Gaussianas independientes Γ_1 y Γ_2 tales que

$$\sqrt{N} \{\bar{X}_N(t) - \mu(t)\} \xrightarrow{\mathcal{D}} \Gamma_1 \quad \text{y} \quad \sqrt{M} \{\bar{X}_M^*(t) - \mu(t)\} \xrightarrow{\mathcal{D}} \Gamma_2.$$

$$\begin{aligned} U_{N,M} &= \int_0^1 \left[\sqrt{\frac{NM}{N+M}} \{\bar{X}_N(t) - \bar{X}_M^*(t) \pm \mu(t)\} \right]^2 dt = \\ &= \int_0^1 \left[\sqrt{\frac{NM}{N+M}} \{\bar{X}_N(t) - \mu(t)\} - \sqrt{\frac{NM}{N+M}} \{\bar{X}_M^*(t) - \mu(t)\} \right]^2 dt = \\ &= \int_0^1 \left[\sqrt{\frac{M}{N+M}} \sqrt{N} \{\bar{X}_N(t) - \mu(t)\} - \sqrt{\frac{N}{N+M}} \sqrt{M} \{\bar{X}_M^*(t) - \mu(t)\} \right]^2 dt = \\ &= \int_0^1 I^2 dt. \end{aligned}$$

Por tanto

$$I \xrightarrow{\mathcal{D}} \Gamma = \sqrt{1-\theta}\Gamma_1 + \sqrt{\theta}\Gamma_2.$$

Como Γ_1 y Γ_2 son funciones Gaussianas independientes, se tiene que Γ es un proceso Gaussiano con media 0 y varianza $(1-\theta)c(t,s) + \theta c^*(t,s)$, donde $c(t,s)$ y $c^*(t,s)$ son las funciones covarianza de Γ_1 y Γ_2 respectivamente.

Por lo que el resultado sigue del teorema de la aplicación continua (ver Teorema 2.2). \square

La distribución asintótica depende de las funciones de covarianza c y c^* . Recordemos que por (2.1)

$$\Gamma(t) = \sum_{k \in \mathbb{N}} \tau_k^{1/2} N_k \varphi_k(t),$$

donde N_1, N_2, \dots son variables aleatorias independientes $N(0,1)$, $\tau_1 \geq \tau_2 \geq \dots$ son los autovalores y $\varphi_1, \varphi_2, \dots$ las autofunciones asociadas al operador determinado por $(1-\theta)c + \theta c^*$. Claramente,

$$\int_0^1 \Gamma^2(t) dt = \sum_{k,s \in \mathbb{N}} \tau_k^{1/2} \tau_s^{1/2} N_k N_s \int_0^1 \varphi_k(t) \varphi_s(t) dt = \sum_{k \in \mathbb{N}} \tau_k N_k^2.$$

Por tanto para obtener una aproximación razonable a $\int_0^1 \Gamma^2(t)dt$ solo necesitamos estimar $\{\tau_k\}_{k \in \mathbb{N}}$. Esto puede hacerse utilizando los autovalores $\{\hat{\tau}_k\}_{k \in \mathbb{N}}$ de la función de covarianza muestral

$$\begin{aligned} \hat{z}_{N,M} &= \frac{M}{N+M} \frac{1}{N} \sum_{i=1}^N [X_i(t) - \bar{X}_N(t)][X_i(s) - \bar{X}_N(s)] + \\ &+ \frac{N}{N+M} \frac{1}{M} \sum_{i=1}^M [X_i^*(t) - \bar{X}_M^*(t)][X_i^*(s) - \bar{X}_M^*(s)]. \end{aligned}$$

La suma $\sum_{k=1}^d \hat{\tau}_k \chi_1^2$, para un valor d suficientemente grande, proporciona una aproximación a $\int_0^1 \Gamma^2(t)$.

La consistencia del test se sigue del siguiente resultado.

Teorema 6.2. *Supongamos que satisface (6.1)-(6.3), (6.5) y*

$$\kappa = \int_0^1 [\mu(t) - \mu^*(t)]^2 dt > 0,$$

entonces $U_{N,M} \xrightarrow{P} \infty$.

Demostración.

$$\frac{N+M}{NM} U_{N,M} = \int_0^1 [\{\bar{X}_N(t) - \mu(t)\} - \{\bar{X}_M^*(t) - \mu^*(t)\} + \{\mu(t) - \mu^*(t)\}]^2 dt.$$

Por la Ley Fuerte de los Grandes Números sabemos que

$$\bar{X}_N(t) - \mu(t) \xrightarrow{c.s.} 0 \text{ y } \bar{X}_M^*(t) - \mu^*(t) \xrightarrow{c.s.} 0.$$

Por lo que

$$\bar{X}_N(t) - \mu(t) \xrightarrow{P} 0 \text{ y } \bar{X}_M^*(t) - \mu^*(t) \xrightarrow{P} 0.$$

Luego

$$\frac{N+M}{NM} U_{N,M} = \int_0^1 [\mu(t) - \mu^*(t)]^2 dt + o_p(1) \xrightarrow{P} \kappa.$$

Es claro que como $\kappa > 0$, entonces $U_{N,M} \xrightarrow{P} \infty$. □

Método II: Este método está basado en proyecciones, utilizaremos las proyecciones sobre el espacio determinado por las autofunciones asociadas a los mayores autovalores del operador $Z = (1 - \theta)C + \theta C^*$. Nótese que

$$Z(x)(t) = \int z(t, s)y(s)ds, \text{ con } z \text{ definido en (6.7).}$$

Supongamos que los autovalores de Z satisfacen

$$\tau_1 > \tau_2 > \dots > \tau_d > \tau_{d+1}. \quad (6.8)$$

Sean $\varphi_1, \dots, \varphi_{d+1}$ las correspondientes autofunciones. Queremos proyectar las observaciones sobre el espacio generado por $\varphi_1, \dots, \varphi_d$. Como esas funciones son desconocidas, utilizamos las correspondientes autofunciones de $\hat{Z}_{N,M}$, denotadas por $\hat{\varphi}_1, \dots, \hat{\varphi}_d$. Ahora proyectamos $\bar{X}_N - \bar{X}_M^*$ sobre el espacio lineal generado por $\hat{\varphi}_1, \dots, \hat{\varphi}_d$. Sean

$$\hat{a}_i = \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_i \rangle, \quad 1 \leq i \leq d,$$

y $\hat{\mathbf{a}} = (\hat{a}_1, \dots, \hat{a}_d)^T$. Veremos que en las condiciones del Teorema 6.1 el vector $(NM/(N + M))^{1/2}\hat{\mathbf{a}}$ es aproximadamente normal d -variante salvo signos aleatorios. Veremos que la matriz de varianzas-covarianzas asintótica de $(NM/(N + M))^{1/2}\hat{\mathbf{a}}$ es $Q = \{Q(i, j), 1 \leq i, j \leq d\}$, donde

$$Q(i, j) = (1 - \theta)E\langle X_1 - \mu, \varphi_i \rangle \langle X_1 - \mu, \varphi_j \rangle + \theta E\langle X_1^* - \mu^*, \varphi_i \rangle \langle X_1^* - \mu^*, \varphi_j \rangle,$$

$1 \leq i, j \leq d$. Se tiene que

$$\begin{aligned} Q(i, j) &= \int_0^1 \int_0^1 (1 - \theta)E[\{X_1(t) - \mu(t)\}\{X_1(s) - \mu(s)\}]\varphi_i(t)\varphi_j(s)dt ds + \\ &+ \int_0^1 \int_0^1 \theta E[\{X_1^*(t) - \mu^*(t)\}\{X_1^*(s) - \mu^*(s)\}]\varphi_i(t)\varphi_j(s)dt ds = \\ &= \int_0^1 \int_0^1 z(t, s)\varphi_i(t)\varphi_j(s)dt ds = \\ &= \begin{cases} \tau_i, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases}. \end{aligned} \quad (6.9)$$

Debido a (6.8) y (6.9), podemos considerar los siguientes estadísticos para contrastar H_0

$$T_{N,M}^{(1)} = \frac{NM}{N + M} \sum_{k=1}^d \hat{a}_k^2 / \hat{\tau}_k$$

y

$$T_{N,M}^{(2)} = \frac{NM}{N+M} \sum_{k=1}^d \hat{a}_k^2.$$

Teorema 6.3. Si H_0 es cierta y (6.1)-(6.3), (6.5) y (6.8) también lo son, entonces

$$T_{N,M}^{(1)} \xrightarrow{\mathcal{D}} \chi_d^2$$

y

$$T_{N,M}^{(2)} \xrightarrow{\mathcal{D}} \sum_{k=1}^d \tau_k N_k^2,$$

donde χ_d^2 es una variable aleatoria chi-cuadrado con d grados de libertad, y N_1, N_2, \dots, N_d son variables aleatorias independientes $N(0, 1)$.

Demostración. Bajo H_0 , X_1, \dots, X_N y X_1^*, \dots, X_N^* tienen la misma media μ . Por el Teorema Central del Límite y bajo H_0 :

$$\sqrt{N}(\bar{X}_N - \mu) \xrightarrow{\mathcal{D}} \Gamma_1$$

y

$$\sqrt{M}(\bar{X}_M^* - \mu) \xrightarrow{\mathcal{D}} \Gamma_2$$

donde Γ_1 y Γ_2 son procesos Gaussianos centrados con funciones de covarianza $c(t, s)$ (la misma que X) y $c^*(t, s)$ (la misma que X^*), respectivamente.

Como

$$\sqrt{\frac{NM}{N+M}}(\bar{X}_N - \bar{X}_M^* + \mu - \mu) = \sqrt{\frac{M}{N+M}}\sqrt{N}(\bar{X}_N - \mu) - \sqrt{\frac{NM}{N+M}}\sqrt{M}(\bar{X}_M^* - \mu)$$

se tiene que

$$\sqrt{\frac{NM}{N+M}}(\bar{X}_N - \bar{X}_M^*) \xrightarrow{\mathcal{D}} \sqrt{1-\theta}\Gamma_1 + \theta\Gamma_2 = \Gamma$$

donde Γ es un proceso Gaussiano centrado con función de covarianza $(1 - \theta)c(t, s) + \theta c^*(t, s)$.

Por definición de proceso Gaussiano

$$\left(\frac{NM}{N+M}\right)^{1/2} (\langle \bar{X}_N - \bar{X}_M^*, \varphi_1 \rangle, \dots, \langle \bar{X}_N - \bar{X}_M^*, \varphi_d \rangle)^T \xrightarrow{\mathcal{D}} N_d(0, \Sigma),$$

donde $N_d(0, \Sigma)$ es un vector aleatorio normal d -variante con media 0 y covarianza Σ .

Recordar que si Γ es un proceso Gaussiano centrado (con función de covarianza $z(t, s)$) con operador covarianza Z

$$\langle \Gamma, x \rangle \sim N(0, \sigma^2), \quad \text{con } \sigma^2 = \langle Zx, x \rangle.$$

Por lo que

$$\Sigma = (\sigma_{ij}), \quad \text{donde } \sigma_{ij} = \langle Z\varphi_i, \varphi_j \rangle$$

De donde se tiene que

$$\langle Z\varphi_i, \varphi_j \rangle = \langle \tau_i \varphi_i, \varphi_j \rangle = \begin{cases} \tau_i, & \text{si } i = j \\ 0, & \text{si } i \neq j \end{cases},$$

donde τ_i y φ_i son los autovalores y autofunciones de Z respectivamente. Podemos observar que $\Sigma = Q$.

Los $\varphi_1, \dots, \varphi_d$ son desconocidos. Veamos qué ocurre si aproximamos φ_i por los $\hat{\varphi}_i$. Como

$$\int_0^1 \int_0^1 (\hat{Z}_{N,M}(t, s) - Z(t, s))^2 dt ds = o_p(1),$$

el Lema 1.2 y la desigualdad (1.5) implican que

$$\max_{1 \leq i \leq d} |\hat{\tau}_i - \tau_i| = o_p(1)$$

$$\max_{1 \leq i \leq d} \|\hat{\varphi}_i - \hat{c}_i \varphi_i\| = o_p(1),$$

donde $\hat{c}_i, \dots, \hat{c}_d$ son signos aleatorios. Lo que realmente tenemos es

$$\begin{aligned} & \left(\frac{NM}{N+M} \right)^{1/2} (\langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_1 \rangle, \dots, \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_d \rangle)^T = \\ & = \left(\frac{NM}{N+M} \right)^{1/2} (\langle \bar{X}_N - \bar{X}_M^*, \hat{c}_1 \varphi_1 \rangle, \dots, \langle \bar{X}_N - \bar{X}_M^*, \hat{c}_d \varphi_d \rangle)^T + \\ & + \left(\frac{NM}{N+M} \right)^{1/2} (\langle \bar{X}_N - \bar{X}_M^*, \varphi_1 - \hat{c}_1 \varphi_1 \rangle, \dots, \langle \bar{X}_N - \bar{X}_M^*, \varphi_d - \hat{c}_d \varphi_d \rangle)^T. \end{aligned}$$

Veamos que el segundo término converge en probabilidad a 0

$$\begin{aligned} & \left(\frac{NM}{N+M} \right)^{1/2} (\langle \bar{X}_N - \bar{X}_M^*, \varphi_1 - \hat{c}_1 \varphi_1 \rangle, \dots, \langle \bar{X}_N - \bar{X}_M^*, \varphi_d - \hat{c}_d \varphi_d \rangle)^T \leq \\ & \leq \left\| \left(\frac{NM}{N+M} \right)^{1/2} (\bar{X}_N - \bar{X}_M^*) \right\| \|\varphi_i - \hat{c}_i \varphi_i\| = O_p(1) o_p(1) = o_p(1). \end{aligned}$$

Ahora veamos qué ocurre con el primer término. Sabemos que $Cov(aX, aX) = a^2 Cov(X, X)$ para cualquier a fijo, por lo que como Q es diagonal, y $(\hat{c}_i)^2 = 1$, se tiene que

$$\left(\frac{NM}{N+M} \right)^{1/2} (\langle \bar{X}_N - \bar{X}_M^*, \hat{c}_1 \varphi_1 \rangle, \dots, \langle \bar{X}_N - \bar{X}_M^*, \hat{c}_d \varphi_d \rangle)^T \xrightarrow{\mathcal{D}} N_d(0, Q).$$

En resumen, por el teorema de Slutsky (ver Teorema 2.1)

$$\left(\frac{NM}{N+M} \right)^{1/2} (\langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_1 \rangle, \dots, \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_d \rangle)^T \xrightarrow{\mathcal{D}} N_d(0, Q).$$

Sabemos que si $X \xrightarrow{\mathcal{D}} N_d(\mu, \Sigma)$ entonces

1. $(X - \mu)' \Sigma^{-1} (X - \mu) \xrightarrow{\mathcal{D}} \chi_d^2$.
2. $(X - \mu)' (X - \mu) \xrightarrow{\mathcal{D}} \sum_{i=1}^d \lambda_i N_i^2$, donde $\lambda_1, \dots, \lambda_d$ son los autovalores de Σ .

Luego

$$\begin{aligned} T_{N,M}^{(1)} &= \frac{NM}{N+M} \sum_{k=1}^d \hat{a}_k^2 / \hat{\tau}_k \xrightarrow{\mathcal{D}} \chi_d^2, \\ T_{N,M}^{(2)} &= \frac{NM}{N+M} \sum_{k=1}^d \hat{a}_k^2 \xrightarrow{\mathcal{D}} \sum_{i=1}^d \tau_i N_i^2. \end{aligned}$$

□

Es claro que $T_{N,M}^2$ es una proyección de $U_{N,M}$ donde sólo usamos los primeros d términos en la expansión de $\bar{X}_N - \bar{X}_M^*$. El estadístico $T_{N,M}^1$ es asintóticamente distribución libre.

La consistencia de los tests basados en T^1 y T^2 puede ser fácilmente establecida en paralelo al Teorema 6.2:

Teorema 6.4. *Si se satisfacen (6.1)-(6.3), (6.5) y (6.8), y $\mu - \mu^*$ no es ortogonal al espacio generado por $\varphi_1, \dots, \varphi_d$, entonces $T_{N,M}^{(1)} \xrightarrow{P} \infty$ y $T_{N,M}^{(2)} \xrightarrow{P} \infty$.*

Demostración.

$$\begin{aligned} \left(\frac{NM}{N+M}\right)^{1/2} \hat{a}_i &= \left(\frac{NM}{N+M}\right)^{1/2} \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_i \rangle = \\ &= \left(\frac{NM}{N+M}\right)^{1/2} [\langle \bar{X}_N - \mu, \hat{\varphi}_i \rangle - \langle \bar{X}_M^* - \mu^*, \hat{\varphi}_i \rangle + \langle \mu - \mu^*, \hat{\varphi}_i \rangle] \end{aligned}$$

Como vimos en el demostración del Teorema 6.2

$$\bar{X}_N - \mu = o_p(1) \text{ y } \bar{X}_M^* - \mu^* = o_p(1).$$

Por lo que

$$\hat{a}_i = \langle \mu - \mu^*, \hat{\varphi}_i \rangle + o_p(1) \xrightarrow{P} \langle \mu - \mu^*, \hat{\varphi}_i \rangle.$$

Sabemos que

$$\langle \mu - \mu^*, \hat{\varphi}_i \rangle = \langle \mu - \mu^*, \hat{c}_i \varphi_i \rangle + \langle \mu - \mu^*, \hat{\varphi}_i - \hat{c}_i \varphi_i \rangle$$

y

$$\langle \mu - \mu^*, \hat{\varphi}_i - \hat{c}_i \varphi_i \rangle \xrightarrow{P} 0.$$

Luego

$$\hat{a}_i = \langle \mu - \mu^*, \hat{\varphi}_i \rangle + o_p(1) = \langle \mu - \mu^*, \hat{c}_i \varphi_i \rangle + o_p(1) \xrightarrow{P} \langle \mu - \mu^*, \hat{c}_i \varphi_i \rangle.$$

Por tanto como $\mu - \mu^*$ no es ortogonal al espacio generado por $\varphi_1, \dots, \varphi_d$ se tiene que $\langle \mu - \mu^*, \hat{c}_i \varphi_i \rangle^2 > 0$ para algún i , por lo que

$$\left(\frac{NM}{N+M}\right) \hat{a}_i^2 \xrightarrow{P} \infty$$

y

$$\left(\frac{NM}{\hat{\tau}_i(N+M)}\right) \hat{a}_i^2 \xrightarrow{P} \infty$$

para algún $i = 1, \dots, d$. Entonces

$$T_{M,N}^{(2)} \xrightarrow{P} \infty \quad \text{y} \quad T_{M,N}^{(1)} \xrightarrow{P} \infty.$$

□

La diferencia entre los tests basados en $U_{N,M}$ y $T_{N,M}^{(1)}$, $T_{N,M}^{(2)}$ es que los dos últimos solo ven la diferencia entre μ y μ^* en un espacio d -dimensional. Si $\mu = \mu^*$ en este subespacio entonces H_0 no será rechazada. No obstante, esto es de poco interés práctico si las primeras d τ_k explican un alto porcentaje de la varianza de la diferencia.

6.2. Igualdad de operadores covarianza

En esta sección presentaremos un procedimiento asintótico para probar la igualdad de operadores covarianza en dos muestras independientes.

Consideramos dos muestras, X_1, X_2, \dots, X_N y $X_1^*, X_2^*, \dots, X_M^*$. Las funciones en cada muestra son elementos de L^2 iid con media cero y las dos muestras son independientes.

Consideramos los operadores covarianza,

$$C(x) = E[\langle X, x \rangle X], \quad C^*(x) = E[\langle X^*, x \rangle X^*],$$

donde X tiene la misma distribución que los X_i y X^* la misma distribución que los X_j^* . Queremos realizar el contraste de hipótesis

$$H_0 : C = C^* \quad \text{vs} \quad H_1 : C \neq C^*.$$

En el Teorema 6.5, supondremos que X y X^* son elementos Gaussianos de L^2 , lo que significa que la igualdad de las covarianzas implica la igualdad en distribución. Por tanto, bajo la hipótesis adicional de normalidad, H_0 establece que los X_i tienen la misma distribución que los X_j^* .

Denotamos por \hat{C} y \hat{C}^* a los equivalentes muestrales de C y C^* , y por \hat{R} , al operador covarianza muestral de la muestra conjunta, es decir,

$$\begin{aligned} \hat{R}(x) &= \frac{1}{N+M} \left\{ \sum_{i=1}^N \langle X_i, x \rangle X_i + \sum_{j=1}^M \langle X_j^*, x \rangle X_j^* \right\} = \\ &= \hat{\theta} \hat{C}(x) + (1 - \hat{\theta}) \hat{C}^*(x), \quad x \in L^2, \end{aligned}$$

donde

$$\hat{\theta} = \frac{N}{N+M}.$$

El operador \hat{R} tiene $N+M$ autofunciones, que llamaremos $\hat{\phi}_k$, $1 \leq k \leq N+M$. Denotamos

$$\hat{\lambda}_k = \frac{1}{N} \sum_{n=1}^N \langle X_n, \hat{\phi}_k \rangle^2, \quad \hat{\lambda}_k^* = \frac{1}{M} \sum_{m=1}^M \langle X_m^*, \hat{\phi}_k \rangle^2.$$

Darse cuenta que los $\hat{\lambda}_k$ y los $\hat{\lambda}_k^*$ no son los autovalores de los operadores \hat{C} y \hat{C}^* , sino las varianzas muestrales de los coeficientes de X y X^* con respecto al sistema ortonormal $\{\hat{\phi}_k, 1 \leq k \leq N+M\}$ formado por las autofunciones del operador \hat{R} .

El estadístico del contraste de hipótesis es

$$\hat{T} = \frac{N+M}{2} \hat{\theta}(1-\hat{\theta}) \sum_{i,j=1}^p \frac{\langle (\hat{C} - \hat{C}^*) \hat{\phi}_i, \hat{\phi}_j \rangle^2}{(\hat{\theta} \hat{\lambda}_i + (1-\hat{\theta}) \hat{\lambda}_i^*) + (1-\hat{\theta} \hat{\lambda}_j^*)},$$

donde p es fijo.

Teorema 6.5. *Supongamos que X y X^* son elementos Gaussianos centrados de L^2 tales que $E\|X\|^4 < \infty$ y $E\|X^*\|^4 < \infty$. Supongamos también que $\hat{\theta} \rightarrow \theta \in (0, 1)$, cuando $N, M \rightarrow \infty$. Entonces, bajo H_0 ,*

$$\hat{T} \xrightarrow{\mathcal{D}} \chi_{p(p+1)/2}^2, \quad N, M \rightarrow \infty,$$

donde $\chi_{p(p+1)/2}^2$ es una variable aleatoria chi-cuadrado con $p(p+1)/2$ grados de libertad.

Demostración. Definimos los operadores aleatorios

$$C_i(x) = \langle X_i, x \rangle X_i, \quad C_j^* = \langle X_j^*, x \rangle X_j^*, \quad x \in L^2.$$

Los C_i forman una secuencia de elementos iid en el espacio de Hilbert \mathcal{S} de operadores Hilbert-Schmidt actuando en L^2 y lo mismo podemos decir para los C_j^* . Bajo H_0 , los C_i y los C_j^* tienen la misma media C . También tienen el

mismo operador covarianza, que es un operador actuando en \mathcal{S} dado por

$$\begin{aligned}
\mathcal{B}(\Psi) &= E[\langle C_i - C, \Psi \rangle_{\mathcal{S}}(C_i - C)] = \\
&= E[\langle C_i - C, \Psi \rangle_{\mathcal{S}} C_i] - E[\langle C_i - C, \Psi \rangle_{\mathcal{S}} C] = \\
&= E[\langle C_i, \Psi \rangle_{\mathcal{S}} C_i] - E[\langle C, \Psi \rangle_{\mathcal{S}} C_i] - E[\langle C_i, \Psi \rangle_{\mathcal{S}} C] + E[\langle C, \Psi \rangle_{\mathcal{S}} C] = \\
&= E[\langle C_i, \Psi \rangle_{\mathcal{S}} C_i] - \langle C, \Psi \rangle_{\mathcal{S}} E(C_i) - \langle E(C_i), \Psi \rangle_{\mathcal{S}} C + \langle C, \Psi \rangle_{\mathcal{S}} C = \\
&= E[\langle C_i, \Psi \rangle_{\mathcal{S}} C_i] - \langle C, \Psi \rangle_{\mathcal{S}} C - \langle C, \Psi \rangle_{\mathcal{S}} C + \langle C, \Psi \rangle_{\mathcal{S}} C = \\
&= E[\langle C_i, \Psi \rangle_{\mathcal{S}} C_i] - \langle C, \Psi \rangle_{\mathcal{S}} C, \quad \Psi \in \mathcal{S}.
\end{aligned} \tag{6.10}$$

Bajo H_0 el segundo término $\langle C, \Psi \rangle_{\mathcal{S}} C$ es el mismo para ambos operadores. Por (1.3)

$$\begin{aligned}
E[\langle C_i, \Psi \rangle_{\mathcal{S}} C_i] &= E \left[\sum_{n \in \mathbb{N}} \langle C_i(e_n), \Psi(e_n) \rangle C_i \right] = \\
&= \sum_{n \in \mathbb{N}} E [\langle X_i, e_n \rangle X_i, \Psi(e_n) \langle X_i, e_n \rangle X_i] = \\
&= \sum_{n \in \mathbb{N}} E [\langle X_i, e_n \rangle^2 \langle X_i, \Psi(e_n) \rangle X_i].
\end{aligned}$$

La hipótesis de que X y X^* son elementos Gaussianos de L^2 y $C = C^*$ implican que los X_i y los X_j^* tienen la misma distribución, por lo que

$$E [\langle X_i, e_n \rangle^2 \langle X_i, \Psi(e_n) \rangle X_i] = E [\langle X_j^*, e_n \rangle^2 \langle X_j^*, \Psi(e_n) \rangle X_j^*].$$

Queremos aplicar el Teorema Central del Límite en el espacio de Hilbert \mathcal{S} a los operadores C_i . Por el Teorema 2.3, debemos verificar que $E \|C_i\|_{\mathcal{S}}^2 < \infty$. Esto se cumple porque, por la Identidad de Parseval,

$$E \|C_i\|_{\mathcal{S}}^2 = E \sum_{n \in \mathbb{N}} \|\langle X_i, e_n \rangle X_i\|^2 = E \left[\|X_i\|^2 \sum_{n \in \mathbb{N}} |\langle X_i, e_n \rangle|^2 \right] = E \|X_i\|^4.$$

Por tanto obtenemos,

$$N^{1/2}(\hat{C} - C) \xrightarrow{\mathcal{D}} Z_1, \quad M^{1/2}(\hat{C}^* - C^*) \xrightarrow{\mathcal{D}} Z_2, \tag{6.11}$$

donde Z_1 es una función Gaussiana definida sobre \mathcal{S} con el mismo operador covarianza, C_1 , y Z_2 es una función Gaussiana con el mismo operador covarianza, C_2 . Por tanto, Z_1 y Z_2 son independientes, y, bajo H_0 ambos tienen

operador covarianza igual a \mathcal{B} . Para cada $1 \leq i, j \leq p$, definimos las variables aleatorias

$$W_{N,M}(i, j) = \left\langle [(N + M)\hat{\theta}(1 - \hat{\theta})]^{1/2}(\hat{C} - \hat{C}^*)\hat{c}_i\hat{\phi}_i, \hat{c}_j\hat{\phi}_j \right\rangle,$$

donde

$$\hat{c}_i = \text{signo}(\langle \hat{\phi}_i, \phi_i \rangle),$$

por tanto

$$\hat{T} = \sum_{i,j=1}^p \frac{W_{N,M}^2(i, j)}{2(\theta\hat{\lambda}_i + (1 - \hat{\theta})\hat{\lambda}_i^*)(\hat{\theta}\hat{\lambda}_j + (1 - \hat{\theta})\hat{\lambda}_j^*)}. \quad (6.12)$$

Observemos que bajo H_0 ,

$$\begin{aligned} W_{N,M}(i, j) &= \left\langle \left[(N + M) \frac{N}{N + M} \frac{M}{N + M} \right]^{1/2} (\hat{C} - \hat{C}^* - C + C^*) \hat{c}_i \hat{\phi}_i, \hat{c}_j \hat{\phi}_j \right\rangle = \\ &= \left\langle \left[\left(\frac{M}{N + M} \right)^{1/2} N^{1/2} (\hat{C} - C) - \left(\frac{N}{N + M} \right)^{1/2} M^{1/2} (\hat{C}^* - C^*) \right] \hat{c}_i \hat{\phi}_i, \hat{c}_j \hat{\phi}_j \right\rangle = \\ &= \left\langle \left[(1 - \hat{\theta})^{1/2} N^{1/2} (\hat{C} - C) - \hat{\theta}^{1/2} M^{1/2} (\hat{C}^* - C^*) \right] \hat{c}_i \hat{\phi}_i, \hat{c}_j \hat{\phi}_j \right\rangle. \end{aligned}$$

Por el Teorema 4.1, bajo H_0 , $\hat{c}_i \hat{\phi}_i \xrightarrow{P} v_i$ y $\hat{\lambda}_i \xrightarrow{P} \lambda_i$, con los v_i siendo las autofunciones de C (y de C^*) y λ_i sus autovalores. Entonces, por (6.11) ya que la convergencia en probabilidad implica convergencia en distribución llegamos a

$$W_{N,M}(i, j) \xrightarrow{D} \langle Z v_i, v_j \rangle, \quad Z = (1 - \theta)^{1/2} Z_1 - \theta^{1/2} Z_2.$$

Como Z_1 y Z_2 son operadores aleatorios independientes en \mathcal{S} , vemos que la varianza del operador Z es igual a

$$(1 - \theta)\mathcal{B} + \theta\mathcal{B} = \mathcal{B}.$$

Por (6.12) obtenemos que

$$\begin{aligned} \hat{T} &\xrightarrow{D} \sum_{i,j=1}^p \frac{\langle Z(v_i), v_j \rangle^2}{2(\theta\lambda_i + (1 - \theta)\lambda_i)(\theta\lambda_j + (1 - \theta)\lambda_j)} = \\ &= \sum_{i,j=1}^p \frac{\langle Z(v_i), v_j \rangle^2}{2\lambda_i\lambda_j} = \\ &= \sum_{i=j} \frac{\langle Z(v_i), v_i \rangle^2}{2\lambda_i^2} + \sum_{i<j} \frac{\langle Z(v_i), v_j \rangle^2}{2\lambda_i\lambda_j} + \sum_{i>j} \frac{\langle Z(v_i), v_j \rangle^2}{2\lambda_i\lambda_j} = \\ &= \sum_{i=j} \frac{\langle Z(v_i), v_i \rangle^2}{2\lambda_i^2} + \sum_{i<j} \frac{\langle Z(v_i), v_j \rangle^2 + \langle Z(v_j), v_i \rangle^2}{2\lambda_i\lambda_j}. \end{aligned} \quad (6.13)$$

Para identificar la distribución del lado derecho de (6.13), es conveniente representar el operador Z en términos de los operadores V_{ij} definidos por

$$V_{ij}(x) = \langle v_i, x \rangle v_j.$$

Por el Lema 6.1,

$$\begin{aligned} Z &\stackrel{d}{=} \sqrt{2} \sum_{i \in \mathbb{N}} \lambda_i \zeta_{ii} V_{ii} + \sum_{i < j} \sqrt{\lambda_i \lambda_j} \zeta_{ij} (V_{ij} + V_{ji}) = \\ &= \sqrt{2} \sum_{i \in \mathbb{N}} \lambda_i \zeta_{ii} V_{ii} + \sum_{i < j} \sqrt{\lambda_i \lambda_j} \zeta_{ij} V_{ij} + \sum_{i > j} \sqrt{\lambda_i \lambda_j} \zeta_{ij} V_{ij} \end{aligned} \quad (6.14)$$

donde los ζ_{ij} son normales estándar iid.

Por (6.14) tenemos que

$$Z(v_k) = \sqrt{2} \lambda_k \zeta_{kk} V_{kk}(v_k) + \sum_{i < j} \sqrt{\lambda_i \lambda_j} \zeta_{ij} (V_{ij}(v_k) + V_{ji}(v_k)).$$

Los v_i son ortogonales, lo que implica que

$$V_{ii}(v_k) = \langle v_i, v_k \rangle v_i = 0, \quad \text{si } i \neq k.$$

$$V_{ij}(v_k) = \langle v_i, v_k \rangle v_j = 0, \quad \text{si } i \neq k.$$

$$V_{ji}(v_k) = \langle v_j, v_k \rangle v_i = 0, \quad \text{si } j \neq k.$$

Luego

$$Z(v_k) = \sqrt{2} \lambda_k \zeta_{kk} v_k + \sum_{i=k < j} \sqrt{\lambda_k \lambda_j} \zeta_{kj} v_j + \sum_{i < j=k} \sqrt{\lambda_i \lambda_k} \zeta_{ik} v_i,$$

y entonces

$$\langle Z(v_k), v_n \rangle = \begin{cases} \sqrt{2} \lambda_k \zeta_{kk}, & \text{si } k = n, \\ \sqrt{\lambda_k \lambda_n} \zeta_{kn}, & \text{si } k < n, \\ \sqrt{\lambda_k \lambda_n} \zeta_{nk}, & \text{si } k > n. \end{cases} \quad (6.15)$$

Usando (6.13) y (6.15) vemos que

$$\hat{T} \stackrel{D}{\rightarrow} \sum_{i=j} \zeta_{ii}^2 + \sum_{i < j} \frac{\zeta_{ij}^2 + \zeta_{ji}^2}{2} \stackrel{d}{=} \sum_{i=j} \zeta_{ii}^2 + \sum_{i < j} \zeta_{ij}^2 \stackrel{d}{=} \chi_{p(p+1)/2}^2.$$

□

Lema 6.1. *Bajo las hipótesis del Teorema 6.5,*

$$\mathcal{B} = \sum_{i \in \mathbb{N}} (\sqrt{2}\lambda_i)^2 \langle V_{ii}, \cdot \rangle_{\mathcal{S}} V_{ii} + \sum_{i < j} \lambda_i \lambda_j \langle V_{ij} + V_{ji}, \cdot \rangle_{\mathcal{S}} (V_{ij} + V_{ji}).$$

Demostración. Trabajaremos con la expansión (2.1)

$$X_n = \sum_{i \in \mathbb{N}} \sqrt{\lambda_i} \xi_{ni} v_i,$$

donde los λ_i son los autovalores de C y $\{\xi_{ni}\}_{i \in \mathbb{N}}$ son secuencias independientes de variables aleatorias iid $N(0, 1)$.

$$\begin{aligned} C_n = \langle X_n, \cdot \rangle X_n &= \left\langle \sum_{i \in \mathbb{N}} \sqrt{\lambda_i} \xi_{ni} v_i, \cdot \right\rangle \sum_{j \in \mathbb{N}} \sqrt{\lambda_j} \xi_{nj} v_j = \\ &= \sum_{i, j \in \mathbb{N}} \sqrt{\lambda_i \lambda_j} \xi_{ni} \xi_{nj} \langle v_i, \cdot \rangle v_j = \\ &= \sum_{i, j \in \mathbb{N}} \sqrt{\lambda_i \lambda_j} \xi_{ni} \xi_{nj} V_{ij}. \end{aligned} \quad (6.16)$$

Por la Proposición 1.5, como C es simétrico y definido positivo se tiene que

$$C = \sum_{i \in \mathbb{N}} \lambda_i \langle \cdot, v_i \rangle v_i = \sum_{i \in \mathbb{N}} \lambda_i \langle v_i, \cdot \rangle v_i = \sum_{i \in \mathbb{N}} \lambda_i V_{ii}. \quad (6.17)$$

Sea Ψ un operador Hilbert-Schmidt arbitrario. Entonces, por (6.10), (6.16) y (6.17)

$$\begin{aligned} \mathcal{B}(\Psi) &= E[\langle C_n, \Psi \rangle_{\mathcal{S}} C_n] - \langle C, \Psi \rangle_{\mathcal{S}} C = \\ &= E \left[\left\langle \sum_{i, j \in \mathbb{N}} \sqrt{\lambda_i \lambda_j} \xi_{ni} \xi_{nj} V_{ij}, \Psi \right\rangle_{\mathcal{S}} \sum_{k, l \in \mathbb{N}} \sqrt{\lambda_k \lambda_l} \xi_{nk} \xi_{nl} V_{kl} \right] - \\ &\quad - \left\langle \sum_{i \in \mathbb{N}} \lambda_i V_{ii}, \Psi \right\rangle_{\mathcal{S}} \sum_{j \in \mathbb{N}} \lambda_j V_{jj} = \\ &= \sum_{i, j, k, l} E[\xi_{ni} \xi_{nj} \xi_{nl} \xi_{nk}] \sqrt{\lambda_i \lambda_j \lambda_l \lambda_k} \langle V_{ij}, \Psi \rangle_{\mathcal{S}} V_{lk} - \sum_{i, j} \lambda_i \lambda_j \langle V_{ii}, \Psi \rangle_{\mathcal{S}} V_{jj}. \end{aligned}$$

El valor esperado $E[\xi_{ni} \xi_{nj} \xi_{nl} \xi_{nk}]$ es cero a menos que haya dos pares de índices iguales, o todos los índices sean iguales. Veamos esto con detalle,

- si existiera un índice distinto de los demás, tendríamos cuatro opciones. Si por ejemplo, el índice i fuera distinto al resto, puesto que los ξ_n son independientes tendríamos que

$$E[\xi_{ni}\xi_{nj}\xi_{nl}\xi_{nk}] = E[\xi_{ni}]E[\xi_{nj}\xi_{nl}\xi_{nk}] = 0 \cdot E[\xi_{nj}\xi_{nl}\xi_{nk}] = 0.$$

Por tanto

$$\sum_{i,j,k,l} E[\xi_{ni}\xi_{nj}\xi_{nl}\xi_{nk}] \sqrt{\lambda_i\lambda_j\lambda_l\lambda_k} \langle V_{ij}, \Psi \rangle_S V_{lk} = 0.$$

- Si los índices fueran iguales dos a dos, habría $\binom{4}{2} = 6$ opciones. Si por ejemplo, $i = j$ y $k = l$,

$$E[\xi_{ni}\xi_{nj}\xi_{nl}\xi_{nk}] = E[\xi_{ni}\xi_{nj}]E[\xi_{nl}\xi_{nk}] = E[\chi_1^2]E[\chi_1^2] = 1 \cdot 1 = 1.$$

Luego

$$\begin{aligned} \sum_{i,j,k,l} E[\xi_{ni}\xi_{nj}\xi_{nl}\xi_{nk}] \sqrt{\lambda_i\lambda_j\lambda_l\lambda_k} \langle V_{ij}, \Psi \rangle_S V_{lk} &= \sum_{i \neq j} \lambda_i\lambda_j \langle V_{ij}, \Psi \rangle_S V_{ij} + \\ &+ \sum_{i \neq j} \lambda_i\lambda_j \langle V_{ij}, \Psi \rangle_S V_{ji} + \sum_{i \neq j} \lambda_i\lambda_j \langle V_{ii}, \Psi \rangle_S V_{jj}. \end{aligned}$$

- Si todos los índices fueran iguales tendríamos que

$$\begin{aligned} E[\xi_{ni}\xi_{nj}\xi_{nl}\xi_{nk}] &= E[(\xi_{ni}^2)^2] = E[(\chi_1^2)^2] = \\ &= \text{var}[\chi_1^2] + (E[\chi_1^2])^2 = 2 \cdot 1 + 1 = 3, \end{aligned}$$

por lo que

$$\sum_{i,j,k,l} E[\xi_{ni}\xi_{nj}\xi_{nl}\xi_{nk}] \sqrt{\lambda_i\lambda_j\lambda_l\lambda_k} \langle V_{ij}, \Psi \rangle_S V_{lk} = 3 \sum_i \lambda_i^2 \langle V_{ii}, \Psi \rangle_S V_{ii}.$$

Tenemos entonces

$$\begin{aligned} \mathcal{B}(\Psi) &= \sum_{i \neq j} \lambda_i\lambda_j [\langle V_{ii}, \Psi \rangle_S V_{jj} + \langle V_{ij}, \Psi \rangle_S V_{ij} + \langle V_{ij}, \Psi \rangle_S V_{ji}] + \\ &+ 3 \sum_i \lambda_i^2 \langle V_{ii}, \Psi \rangle_S V_{ii} - \sum_i \lambda_i^2 \langle V_{ii}, \Psi \rangle_S V_{ii} - \sum_{i \neq j} \lambda_i\lambda_j \langle V_{ii}, \Psi \rangle_S V_{jj} = \\ &= 2 \sum_i \lambda_i^2 \langle V_{ii}, \Psi \rangle_S V_{ii} + \sum_{i \neq j} \lambda_i\lambda_j [\langle V_{ij}, \Psi \rangle_S V_{ij} + \langle V_{ij}, \Psi \rangle_S V_{ji}]. \end{aligned}$$

La prueba finaliza reordenando los términos. □

Capítulo 7

Aplicación en R

En este capítulo describiremos una forma de llevar a cabo algunos de los procedimientos vistos en esta memoria con el lenguaje de programación R. Para ello utilizaremos el paquete `fda`.

7.1. Bases B-spline

Los datos funcionales para N funciones con índices $i = 1, \dots, N$ vienen de un conjunto de valores discretos y_{i1}, \dots, y_{iJ} . Lo primero que debemos hacer es convertir esos valores en una función X_i con valores $X_i(t)$ que pueden ser calculados para cualquier valor t , donde t es un argumento unidimensional. Observamos las funciones de t en los valores $t_j, j = 1, \dots, J$.

Para hacer esto, necesitamos especificar una base. Una base es un sistema de un conjunto de funciones conocidas ϕ_k que son ortogonales entre sí y que tienen la propiedad de que pueden aproximar bien cualquier función tomando una combinación lineal de un número K suficientemente grande de esas funciones. Expresaremos una observación funcional X_i como

$$X_i(t) \approx \sum_{k=1}^K c_{ik} \phi_k(t), \quad (7.1)$$

donde $\phi_k, k = 1, \dots, K$, son las funciones de base y $t = [a, b]$ es el intervalo donde recaen las observaciones.

Determinaremos los coeficientes c_{ik} minimizando el criterio de mínimos cua-

drados

$$RSS = \sum_{j=1}^J [y_{ij} - \sum_{k=1}^K c_{ik} \phi_k(t_j)]^2,$$

siendo RSS la suma residual de cuadrados.

Trabajaremos con las bases B-spline. Para ello examinaremos primero la estructura de una función spline. El primer paso es dividir el intervalo en el que la función va ser aproximada en L subintervalos separados por valores π_l , $l = 1, \dots, L - 1$ que se denominan nudos.

Sobre cada intervalo, un spline es un polinomio de orden m . El orden de un polinomio es el número de constantes requeridas para definirlo, y este es uno más su grado.

Polinomios adyacentes deben unirse en el nudo que los separa, por ello los valores de la función se construyen para ser iguales en su unión.

Para construir una función spline especificaremos un sistema de funciones de base $\phi_k(t)$, que tendrá las siguientes propiedades:

- Cada función $\phi_k(t)$ es una función spline con orden m y una secuencia de nudos τ .
- Puesto que un múltiplo de una función spline es una función spline, y sumas y diferencias de splines son también splines, cualquier combinación lineal de esas funciones base es una función spline.
- Cualquier función spline definida por m y τ se puede expresar como una combinación lineal de esas funciones de base.

Aunque hay muchas formas de construir tales sistemas, las bases B-splines desarrolladas por de Boor (2001) son las más populares, y el código para trabajar con ellas está disponible en un amplio rango de lenguajes de programación, incluido R.

7.2. Datos funcionales en R

Trabajaremos con el conjunto de datos `pinch` el cual se encuentra en la librería `fda`. `pinch` contiene $J = 151$ mediciones de fuerza de tracción para $N = 20$

replicaciones cada 2 milisegundos para 300 milisegundos.

```
library(fda)
y<-pinch
x=1:dim(pinch)[1]
base<-create.bspline.basis(rangeval=c(1,151),nbasis=69)
```

donde `rangeval` es un vector de longitud 2 definiendo el intervalo sobre el que los datos se pueden evaluar y `nbasis` es una variable que especifica el número de funciones de base. Una vez se crea la base, los datos se convierten en objetos funcionales, para esto se usa la función `Data2fd`.

```
data.fd<-Data2fd(argvals=x,y, basisobj=base)
```

con `argvals` el conjunto de valores donde recaen las observaciones, y matriz que contiene los valores de las curvas y `basisobj` vector numérico que indica los nudos de la base B-spline. El código presentado abajo produce la Figura 7.1.

```
plot.fd(data.fd)
```

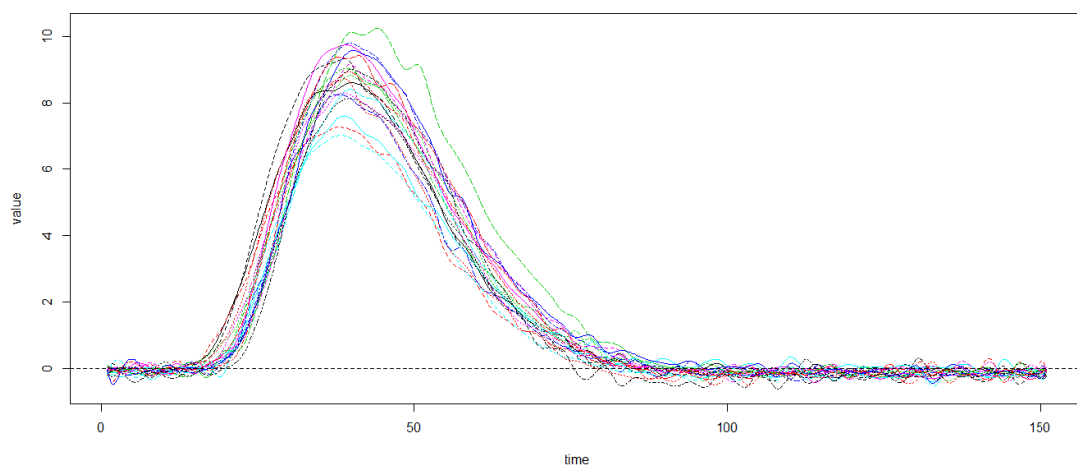


Figura 7.1: Datos funcionales.

7.3. Componentes principales funcionales en R

La función de R `pca.fd` calcula las CPFE \hat{v}_j , sus correspondientes autovalores $\hat{\lambda}_j$ y las puntuaciones $\langle X_i - \bar{X}_N, \hat{v}_j \rangle$. Para calcular esto hacemos

```
pca<-pca.fd(data.fd, nharm = 3, centerfns = TRUE)
CPV<-100*sum(pca$values[1:3])/sum(pca$values)
```

El argumento `nharm` especifica el número de p de CPFE que serán estimadas. `centerfns = TRUE` significa que las CPFE y las puntuaciones han sido calculadas para funciones centradas $X_i - \bar{X}_N$.

Hemos obtenido un $CPV = 93,1185$, por lo que $p = 3$ es un valor recomendado, aunque si lo calculamos para $p = 2$ nos saldría un valor de $88,2335$, es decir, con 2 componentes principales nuestros datos también quedarían suficientemente explicados.

Una vez que `pca` ha sido creado, obtendremos $\hat{v}_1, \dots, \hat{v}_p$ escribiendo la orden `pca$harmonics`, $\hat{\lambda}_1, \dots, \hat{\lambda}_p$ con `pca$values`. Las puntuaciones $\langle X_i - \bar{X}_N, \hat{v}_j \rangle$, $i = 1, 2, \dots, N$, $j = 1, 2, \dots, p$ están en la matriz $N \times p$ `pca$scores`.

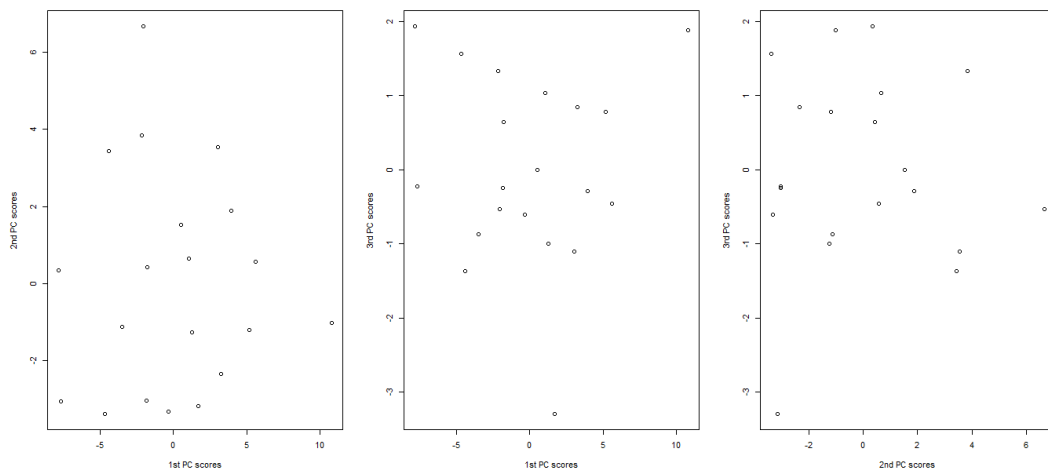


Figura 7.2: Gráficos de dispersión de las puntuaciones.

El panel de la izquierda de la Figura 7.2 muestra el gráfico de dispersión de los pares $(\langle X_i - \bar{X}_N, \hat{v}_1 \rangle, \langle X_i - \bar{X}_N, \hat{v}_2 \rangle)$. Los otros dos paneles muestran los gráficos

de dispersión de las otras dos combinaciones. Dicha figura ha sido creada con el siguiente código

```
par(mfrow=c(1,3))
plot(pca$scores[,1], pca$scores[,2], xlab="1st PC scores",
     ylab="2nd PC scores")
plot(pca$scores[,1], pca$scores[,3], xlab="1st PC scores",
     ylab="3rd PC scores")
plot(pca$scores[,2], pca$scores[,3], xlab="2nd PC scores",
     ylab="3rd PC scores")
par(mfrow=c(1,1))
```

7.4. Igualdad de medias en R

Para realizar el contraste de hipótesis

$$H_0 : \mu = \mu^* \text{ en } L^2 \quad \text{vs} \quad H_1 : \mu \neq \mu^* \text{ en } L^2.$$

usaremos las técnicas vistas en el método II del capítulo anterior. Para ello calcularemos el estadístico $T_{N,M}^{(1)}$. Recordemos que

$$T_{N,M}^{(1)} = \frac{NM}{N+M} \sum_{k=1}^d \hat{a}_k^2 / \hat{\tau}_k,$$

siendo $\hat{a}_k = \langle \bar{X}_N - \bar{X}_M^*, \hat{\varphi}_k \rangle$, $1 \leq k \leq d$ con $\hat{\tau}_k$ y $\hat{\varphi}_k$ los autovalores y autofunciones de $\hat{Z}_{N,M}$.

Recordemos también que

$$\begin{aligned} \hat{z}_{N,M} &= \frac{M}{N+M} \frac{1}{N} \sum_{i=1}^N [X_i(t) - \bar{X}_N(t)][X_i(s) - \bar{X}_N(s)] + \\ &+ \frac{N}{N+M} \frac{1}{M} \sum_{i=1}^M [X_i^*(t) - \bar{X}_M^*(t)][X_i^*(s) - \bar{X}_M^*(s)]. \end{aligned}$$

Consideramos las muestras $y1$ e $y2$ con $N = 10$ y $M = 8$ respectivamente. Primero, calculemos los datos centrados, para ello usaremos la función `center.fd`.

```

library(fda)
set.seed(1234)
N=10
M=8
x<-1:100/100

y1<-sin(x)+matrix(rnorm(100*N)/4, ncol=N)
base1<-create.bspline.basis(rangeval=c(0,1),nbasis=7)
data.fd1<-Data2fd(x,y1, basisobj=base1)
mean.function1<-mean.fd(data.fd1)
plot.fd(data.fd1)
plot.fd(mean.function1, add=TRUE)
data.fd1<-center.fd(data.fd1)
plot.fd(data.fd1)

```

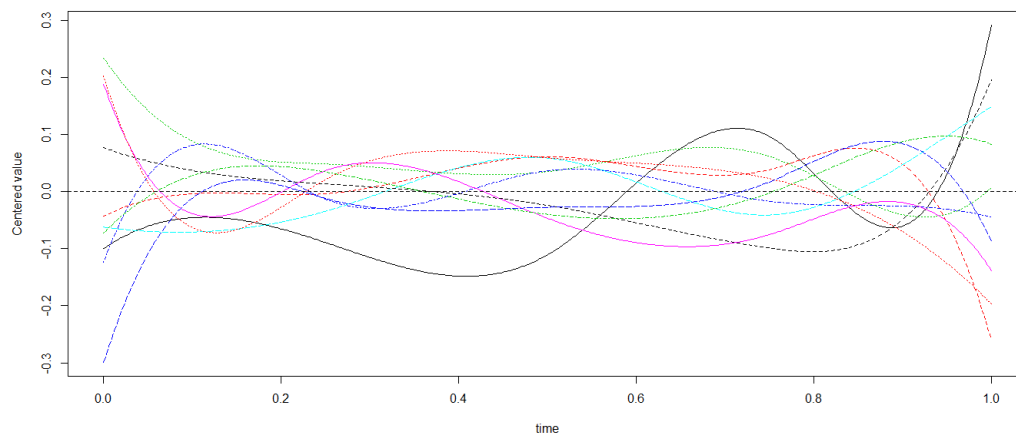


Figura 7.3: Datos funcionales centrados de la primera muestra.

```

y2<-cos(x)+matrix(rnorm(100*M)/4, ncol=M)
base2<-create.bspline.basis(rangeval=c(0,1),nbasis=7)
data.fd2<-Data2fd(x,y2, basisobj=base2)
mean.function2<-mean.fd(data.fd2)
plot.fd(data.fd2)
plot.fd(mean.function2, add=TRUE)
data.fd2<-center.fd(data.fd2)
plot.fd(data.fd2)

```

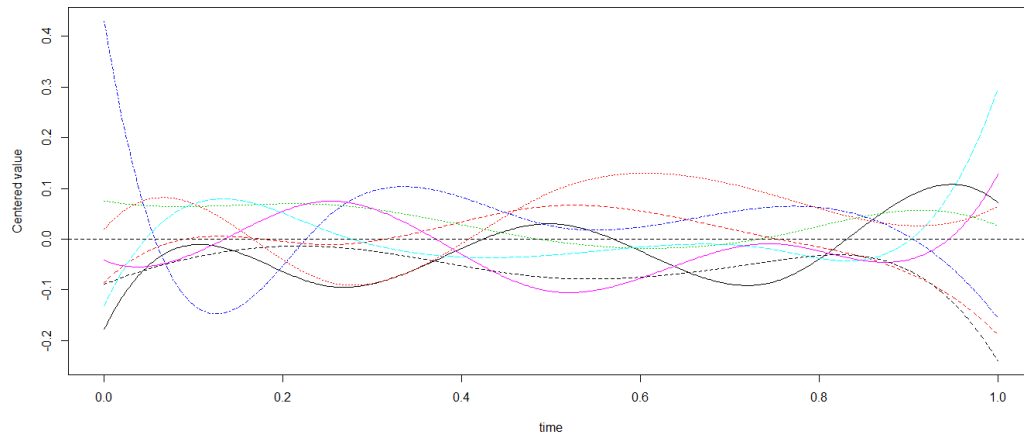


Figura 7.4: Datos funcionales centrados de la segunda muestra.

Agrupamos ahora ambas muestras multiplicando los datos de la primera por $\sqrt{M/N}$ y los de la segunda por $\sqrt{N/M}$. Almacenamos sus autovalores y autofunciones con la función `pca.fd`.

```
data.fd1$coefs<-sqrt(M/N)*data.fd1$coefs
data.fd2$coefs<-sqrt(N/M)*data.fd2$coefs
```

```
data.fd1$coefs=cbind(data.fd1$coefs,data.fd2$coefs)
pca<-pca.fd(data.fd1, nharm = 5, centerfns = TRUE)
CPV<-100*sum(pca$values[1:5])/sum(pca$values)
```

Tomamos 5 componentes principales puesto que el primer valor para el cual el *CVP* supera el 85%.

Una vez hecho esto ya podemos pasar a calcular el vector **a**

```
restamedia=mean.function1$coefs- mean.function2$coefs
```

```
autf<-pca$harmonics$coefs
a1<-sum(restamedia*autf[,1])
a2<-sum(restamedia*autf[,2])
a3<-sum(restamedia*autf[,3])
a4<-sum(restamedia*autf[,4])
a5<-sum(restamedia*autf[,5])
a=rbind(a1,a2,a3,a4,a5)
```

Calculemos ahora el estadístico y su correspondiente p-valor con un nivel de significación $\alpha = 0,05$. Por el Teorema 6.3 sabemos que $T_{N,M}^{(1)}$ se distribuye según una chi-cuadrado con $d = 5$ grados de libertad.

```
T1=((M*N)/(N+M))*sum(a^2*(1/pca$values[1:5]))
pvalor=1-pchisq(T1,5)
```

Observamos que p-valor = 0 < 0,05 = α , luego rechazamos la hipótesis nula. Por lo que hay evidencias para afirmar que las medias de ambas muestras son distintas.

Bibliografía

- [1] Billingsley, P. *Convergence of Probability Measures*. Second Edition, Wiley, New York (1999).
- [2] Bosq, D. *Linear Processes in Function Spaces*. Springer, New York (2000).
- [3] de Boor, C. *A Practical Guide to Splines*. Revised Edition. Springer, New York (2001).
- [4] Gohberg, I., Golberg, S. y Kaashoek, M.A. *Classes of Linear Operators*. Classes of Linear Operators. Operator Theory: Advances and Applications, volume 49. Birkhäuser (1990).
- [5] Hall, P. and Hosseini-Nasab, M. On properties of functional principal components. *Journal of the Royal Statistical Society (B)*, 68, 109–126. (2006)
- [6] Horváth, L. y Kokoszka, P. *Inference for Functional Data with Applications*. Springer (2012).
- [7] Laha, R.G. y Rohatgi, V.K. *Probability Theory*. John Wiley & Sons (1979).
- [8] Leon, S. *Linear Algebra with Applications*. Pearson (2006).
- [9] Ramsay, J.O. y Silverman B.W. *Functional Data Analysis*. Springer (2005).
- [10] Riesz, F. y Sz.-Nagy, B. *Functional analysis*. Dover (1990).
- [11] Serfling, R. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons (2002).