



FACULTADE DE MATEMÁTICAS

Traballo Fin de Grao

El Modelo de Regresión Lineal Funcional

Luis García Romarís

Junio 2025

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

GRAO DE MATEMÁTICAS

Traballo Fin de Grao

El Modelo de Regresión Lineal Funcional

Luis García Romarís

Junio 2025

UNIVERSIDADE DE SANTIAGO DE COMPOSTELA

Trabajo propuesto

Área de Coñecemento: Estadística e Investigación Operativa
Título: Modelo de regresión lineal funcional
Breve descripción do contido
Se trata de revisar algunos modelos de regresión con covariables funcionales y respuesta escalar o funcional.
Recomendacións
El guión de trabajo a seguir es el siguiente: <ol style="list-style-type: none">1) Una revisión del modelo lineal general de regresión.2) Aspectos metodológicos y prácticos del análisis de datos funcionales.3) Modelos de regresión lineal funcional con covariable funcional.4) Modelos de regresión lineal funcional con respuesta funcional.5) Aplicaciones a datos reales o simulados.
Outras observacións
El trabajo se desarrollará a lo largo del curso con la siguiente planificación: parte 1) hasta finales de diciembre, parte 2) hasta finales de febrero, partes 3) y 4) hasta finales de abril y parte 5) hasta finales de mayo. Se dedicará junio para revisiones finales, presentación y aspectos varios.

Índice

Resumen	VIII
Introducción	XI
1. Revisión de la regresión lineal con variables escalares	1
1.1. Introducción a la regresión	1
1.2. Formulación del modelo de regresión lineal general	2
1.3. Estimación del vector de coeficientes	4
1.4. Inferencia sobre los coeficientes	4
2. Aspectos teóricos y metodológicos del análisis de datos funcionales	7
2.1. Elección del espacio funcional. Espacios de Hilbert	8
2.2. Elementos aleatorios en un espacio de Hilbert	10
2.3. Esperanza y operador de covarianza de funciones aleatorias	11
2.3.1. Esperanza y operador de covarianza en el espacio L^2	15
2.3.2. Estimadores empíricos	15
2.4. Análisis de componentes principales funcionales	18
2.4.1. Componentes principales funcionales estimadas	21
2.5. De datos discretos a datos funcionales: técnicas de presuavizado	22
2.5.1. Expansión en base. Suavizado por mínimos cuadrados	24
2.5.2. Suavizado mediante penalización	26

3. Los modelos de regresión lineal funcional	29
3.1. Modelo de regresión lineal funcional de respuesta escalar	30
3.1.1. Diferencias con respecto a la regresión euclidiana	31
3.1.2. Estimación por medio de una expansión en base	34
3.1.3. Estimación mediante penalización	35
3.1.4. Estimación basada en componentes principales funcionales	36
3.1.5. Contraste de significación	38
3.2. Modelo de regresión lineal funcional de respuesta funcional	41
3.2.1. Estimación basada en componentes principales funcionales	43
3.2.2. Estimación mediante penalización	45
3.2.3. Contraste de significación	47
4. Aplicación a datos de CGM	51
Referencias	59
A. Fundamentos teóricos del análisis funcional	61
A.1. Subespacios. Ortogonalidad	61
A.2. Propiedades básicas de operadores	62
A.3. Operadores compactos. Subclases	62
B. Código	65

Resumen

El análisis de datos funcionales es la rama de la estadística que tiene por objeto el estudio de funciones como objetos probabilísticos, expandiendo la visión tradicional ocupada únicamente con escalares o vectores. El objetivo de este trabajo es proporcionar una introducción a la estadística de los datos funcionales, estableciendo sus bases teóricas y prestando especial atención al análisis de la regresión. Se formulan dos modelos de regresión lineal funcional, según el carácter escalar o funcional de la respuesta, y se presentan distintas técnicas de estimación y contraste de significación aplicables a cada uno de los modelos.

Abstract

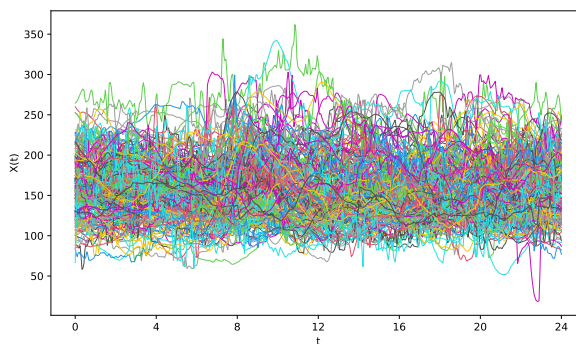
Functional data analysis is the branch of statistics concerned with the study of functions as probabilistic objects, rather than scalars or vectors, as is the case in the traditional statistical framework. The aim of this work is to provide an introduction to the statistics of functional data, establishing its theoretical foundations with regression as a centric point. Two functional linear regression models are formulated, depending on whether the response variable is scalar or functional, and a range of estimation methods as well as significance testing procedures applicable to each model are presented.

Introducción

En las últimas décadas, numerosos y frenéticos avances tecnológicos han abierto las puertas a nuevas formas de adquirir datos. En campos como la medicina, la meteorología o el análisis financiero es cada vez más habitual trabajar con mediciones de alta frecuencia, sea en el dominio del tiempo, el espacio u otros tipos de magnitudes continuas. En muchas ocasiones, esto permite caracterizar los sucesos subyacentes observados como objetos funcionales, tales como curvas, superficies, u objetos similares de mayor dimensión.

Desde comienzos de este siglo ha estado ganando cada vez mayor relevancia el uso de dispositivos subcutáneos que permiten tomar mediciones regulares y frecuentes de los niveles de glucosa en sangre de pacientes diabéticos [10]. Esta reciente tendencia de la medicina personalizada, que se conoce como monitorización continua de glucosa (CGM, por sus siglas en inglés *Continuous Glucose Monitoring*) es una clara fuente de datos funcionales. Es razonable considerar los datos proporcionados por esta técnica no como observaciones discretas de los niveles de glucosa $X(t_j)$, $j = 1, \dots, J$, sino como una única curva $X(t)$, $t \in [0, 24]$, que registra la tendencia de la concentración de glucosa en el tiempo a lo largo de, por ejemplo, un día completo.

En la figura de la derecha pueden observarse las curvas diarias del nivel de glucosa en sangre de un número de pacientes. Esta base de datos será empleada a lo largo del trabajo con motivos ilustrativos. Una descripción más detallada de los datos se encuentra en el Capítulo 4.



El análisis de datos funcionales responde a la necesidad de un marco estadístico que permita trabajar con la cantidad emergente y en crecimiento constante de datos de este tipo. Conceptos bien conocidos y ampliamente estudiados de la estadística multivariante —como la clasificación, técnicas de reducción de dimensionalidad, y detección de datos atípicos, entre otros— deben ser reformulados para poder aplicarse en el ámbito de los recién llegados datos funcionales.

Con el objetivo de desarrollar la teoría básica de la regresión lineal en datos funcionales, en este trabajo introduciremos conceptos teóricos fundamentales del análisis de datos funcionales, como el operador de covarianza y el análisis de componentes principales funcionales. Objetos como estos fundamentarán los métodos de estimación y de contraste propuestos en el estudio de la regresión.

De manera más precisa, en el Capítulo 1 se recuerdan las nociones básicas de la regresión lineal en el contexto escalar, con la intención de fijar la notación y establecer conceptos paralelos a los que trataremos en capítulos posteriores.

En el Capítulo 2 nos adentramos propiamente en el estudio del análisis de datos funcionales. Comenzaremos introduciendo el concepto de función aleatoria, redefiniendo la distribución gaussiana y adaptando objetos estadísticos fundamentales, como lo son la media y la matriz de covarianzas, a este nuevo contexto. El análisis de componentes principales funcionales, una herramienta especialmente útil en el ámbito funcional, será tratado también. Se cerrará el capítulo con una sección dedicada al tratamiento de datos funcionales tal y como surgen en el mundo real, estudiando técnicas de presuavizado.

El Capítulo 3 se dedica a presentar la materia que da propiamente título a este trabajo, la regresión lineal con datos funcionales. Exploraremos dos modelos de regresión lineal funcional, diferenciados por el carácter escalar o funcional de la respuesta. Para cada uno de ellos, se desarrollará una variedad de técnicas de estimación de la función coeficiente y se propondrá un procedimiento para el contraste de significación de la misma.

Finalmente, en el Capítulo 4 se aplicarán las técnicas desarrolladas a un estudio sobre la ya mencionada base de datos de CGM. Ajustaremos un modelo de regresión lineal funcional y contrastaremos su significación, con el fin de ilustrar la metodología propuesta a lo largo del trabajo.

Esta memoria es acompañada por dos anexos. En el primero de ellos se recogen algunas nociones básicas del análisis funcional en espacios de Hilbert empleadas en el curso del trabajo, que resultan útiles, aunque no imprescindibles, para el seguimiento del texto. En el segundo se lista el código empleado para efectuar el estudio de los datos de CGM en el Capítulo 4.

Capítulo 1

Revisión de la regresión lineal con variables escalares

Dedicaremos el primer capítulo de este texto a revisar brevemente el modelo de regresión lineal general, con variables escalares, con el objetivo de más adelante trasladar las ideas que surgen en este contexto al análisis de datos funcionales.

La primera sección se dedica a introducir los conceptos generales del estudio de la regresión. En la segunda nos centramos en los modelos de regresión lineal, presentándolos de manera jerárquica en creciente complejidad, y en las dos secciones que la siguen nos ocupamos de las cuestiones de estimación e inferencia acerca de los parámetros del modelo lineal general.

1.1. Introducción a la regresión

La regresión es un modelo estadístico que describe una variable dependiente, que llamamos respuesta, en función de una variable independiente, que llamamos variable explicativa o covariable. El objetivo de la regresión es analizar la relación entre ambas variables para llevar a cabo propósitos explicativos y predictivos.

Consideraremos entonces dos variables aleatorias: la variable respuesta Y y la variable explicativa X , que serán siempre escalares a lo largo de este capítulo. El supuesto básico de la regresión consiste en que la variabilidad de la respuesta Y puede descomponerse de la siguiente forma: una parte de ella depende de manera determinística de la realización de la variable explicativa X y la segunda se trata de una componente estocástica que trataremos como ruido o error. Esto es, podemos escribir la relación entre ambas variables como

$$Y = m(X) + \varepsilon,$$

donde la función m se denomina *función de regresión*, y para nuestros propósitos se tratará de la media de Y condicionada a X :

$$m(x) = \mathbb{E}[Y | X = x].$$

Existen otros métodos de regresión que pretenden explicar otras medidas de tendencia central, como la regresión cuantil o la regresión modal, pero nosotros nos centraremos exclusivamente en la regresión habitual que trabaja con la media.

1.2. Formulación del modelo de regresión lineal general

Existen diversas formas de tratar la regresión, cada una adecuada dependiendo del contexto y la naturaleza de la relación entre la variable respuesta y explicativa. A lo largo de este trabajo, la protagonista será la regresión lineal, en la que dicha relación se modela como una dependencia lineal, tal y como su nombre sugiere.

Dentro de la regresión lineal podemos encontrarnos diferentes modelos con crecientes grados de complejidad. El primero que presentamos es el *modelo de regresión lineal simple* (MLS). En este se considera una única variable explicativa escalar X , y se asume que la función de regresión viene dada por $m(x) = \beta_0 + \beta_1 x$, siendo los coeficientes β_0 y β_1 parámetros desconocidos.

Con esto, el MLS se puede formular en términos poblacionales como

$$Y = \beta_0 + \beta_1 X + \varepsilon.$$

Una escritura alternativa en el contexto muestral, si se realizan n observaciones emparejadas (x_i, Y_i) de las variables explicativa y respuesta, es la siguiente:

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, \dots, n,$$

donde Y_i es la observación de la respuesta del i -ésimo individuo, x_i el valor de la covariable para el mismo y ε_i el error asociado a dicho individuo.

Las hipótesis más habituales del MLS, las cuales son comunes a todos los modelos lineales, son las siguientes: linealidad, homocedasticidad y normalidad de los errores e independencia de las observaciones. Adicionalmente notamos que de manera habitual las realizaciones x_i de la covariable se asumen conocidas, y por tanto se tratan como valores determinísticos. Este enfoque se conoce como diseño fijo y simplifica en gran medida las tareas de inferencia.

El modelo de regresión lineal simple es excesivamente sencillo en multitud de situaciones. El primer paso en complejidad consiste en permitir múltiples variables explicativas (o, equivalen-

temente, considerar una covariable vectorial $X \in \mathbb{R}^{p-1}$), lo cual nos lleva al llamado *modelo de regresión lineal múltiple* (MLM).

Pretendemos modelar Y en función de varias covariables X_1, \dots, X_{p-1} . Entonces, la formulación a nivel poblacional del modelo será la siguiente:

$$Y = \beta_0 + \beta_1 X_1 + \dots + \beta_{p-1} X_{p-1} + \varepsilon,$$

y, de nuevo, esta puede reescribirse en términos muestrales como sigue:

$$Y_i = \beta_0 + \beta_1 x_{i,1} + \dots + \beta_{p-1} x_{i,p-1} + \varepsilon_i = \beta_0 + \sum_{j=1}^{p-1} \beta_j x_{i,j} + \varepsilon_i. \quad (1.1)$$

Con el objetivo de simplificar la escritura se introduce la siguiente notación matricial:

$$Y = X\beta + \varepsilon, \quad (1.2)$$

donde ahora denotamos $Y = (Y_1, \dots, Y_n)' \in \mathbb{R}^n$, $\beta = (\beta_0, \dots, \beta_{p-1})' \in \mathbb{R}^p$ y $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)' \in \mathbb{R}^n$, y la *matriz de diseño* $X \in \mathcal{M}_{n \times p}$ viene dada por:

$$X = \begin{pmatrix} 1 & x_{1,1} & \dots & x_{1,p-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & \dots & x_{n,p-1} \end{pmatrix}$$

A partir de la presentación matricial del MLM se puede elevar la generalidad del modelo una vez más. En lugar de exigir que la matriz de diseño esté conformada por columnas que contienen las observaciones de las covariables, en el *modelo de regresión lineal general* la forma de la matriz X tiene total libertad.

Este modelo se formula exactamente como en la Ecuación (1.2) sin asumir una estructura determinada de la matriz de diseño. Para cerrar su planteamiento, una forma compacta de recoger las hipótesis del modelo es la siguiente:

$$\varepsilon \sim \mathcal{N}_n(\vec{0}, \sigma^2 \mathbf{I}_n).$$

El modelo lineal general tiene como casos particulares el MLS y MLM mencionados anteriormente, y además es capaz de acoger cómodamente formulaciones de otros modelos aparentemente diferenciados, como los modelos ANOVA y ANCOVA.

Por último, puede tomarse un paso final para expandir el marco de la regresión lineal: el modelo lineal generalizado. Este modelo permite transformar el predictor lineal $X\beta$ con el fin de modelar otras clases de respuestas para las que el modelo lineal general no es satisfactorio. Por ejemplo, la regresión logística, en la que Y es una variable aleatoria Bernoulli, es un caso particular del modelo lineal generalizado.

1.3. Estimación del vector de coeficientes

Antes de dar comienzo a esta sección, recordamos que todo el estudio a continuación se lleva a cabo en el marco del modelo lineal general, por lo que todos los modelos particulares, como el MLS y el MLM, se benefician también de este desarrollo.

El primer objetivo fundamental del análisis de datos en regresión es siempre la estimación de los parámetros del modelo. En el caso de los modelos lineales, la metodología más extendida con este fin es la de mínimos cuadrados: se buscan los coeficientes que minimicen la suma de distancias cuadráticas entre las observaciones y los valores ajustados.

En el caso del MLM, el estimador de mínimos cuadrados viene dado mediante la siguiente condición:

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (Y_i - x_i \beta)^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2, \quad (1.3)$$

donde $\hat{Y}_i = x_i \hat{\beta}$ es el valor ajustado correspondiente a la observación i -ésima, denotada por $x_i = (1, x_{i,1}, \dots, x_{i,p-1})$, de las covariables.

En el marco del modelo lineal general, la Ecuación 1.3 se expresa de la siguiente forma:

$$\hat{\beta} = \arg \min_{\beta \in \mathbb{R}^p} (Y - \hat{Y})'(Y - \hat{Y}),$$

siendo $\hat{Y} = X\beta$ el vector de valores ajustados.

Para hallar el valor óptimo se puede proceder calculando los puntos críticos de la función objetivo. Este desarrollo conduce a las llamadas ecuaciones normales de la regresión:

$$(X'X)\beta = X'Y, \quad (1.4)$$

de las cuales se deduce que, en caso de que la matriz $X'X$ sea invertible (lo cual ocurre siempre que X tenga rango máximo), el estimador de mínimos cuadrados viene dado por:

$$\hat{\beta} = (X'X)^{-1}X'Y.$$

1.4. Inferencia sobre los coeficientes

El motivo por el que el estimador $\hat{\beta}$ dado por el método de mínimos cuadrados se usa de manera tan extendida es que, bajo las hipótesis de los modelos lineales, coincide con el de máxima verosimilitud, y tiene además un número de propiedades deseables para la inferencia.

El estimador hereda las propiedades comunes a todos los estimadores de máxima verosimilitud: consistencia, eficiencia asintótica... Por otro lado, bajo las hipótesis del modelo $\hat{\beta}$ es

insesgado y sigue una distribución normal: $\hat{\beta} \sim \mathcal{N}_p(\beta, \sigma^2(X'X)^{-1})$. De hecho, $\hat{\beta}$ es el estimador lineal insesgado óptimo (BLUE) de β ¹. La distribución conocida de $\hat{\beta}$ facilita en gran medida la construcción de intervalos o regiones de confianza y simplifica la tarea de contrastar la significación de los coeficientes.

Los contrastes de significación son herramientas esenciales en el análisis de regresión, ya que permiten evaluar de manera rigurosa si una covariable ejerce una influencia real sobre la respuesta.

El contraste de significación de un cierto coeficiente β_k se formula como el contraste de la hipótesis nula:

$$H_0 : \beta_k = 0,$$

frente a

$$H_1 : \beta_k \neq 0.$$

La distribución conocida del estimador $\hat{\beta}_k \sim \mathcal{N}(\beta_k, \sigma^2(X'X)_{kk}^{-1})$ permite enfocarlo como un simple test-T.

En caso de pretender contrastar la significación del modelo globalmente (o bien de ciertos coeficientes de forma simultánea), esto es, contrastar la hipótesis nula

$$H_0 : \beta_k = 0, \forall k \in \{1, \dots, p-1\}$$

frente a

$$H_1 : \exists k \in \{1, \dots, p-1\} / \beta_k \neq 0,$$

el procedimiento de contraste apropiado es un F-test.

¹Por esto se entiende que la elección de $\hat{\beta}$ minimiza la varianza de las proyecciones del estimador sobre todo vector $c \in \mathbb{R}^p$, es decir, para todo $\tilde{\beta}$ perteneciente a la clase de estimadores insesgados de β , $\text{Var}(c'\hat{\beta}) \leq \text{Var}(c'\tilde{\beta})$.

Capítulo 2

Aspectos teóricos y metodológicos del análisis de datos funcionales

En este capítulo dejamos atrás el mundo euclídeo y nos adentramos en el ámbito del análisis de datos funcionales (FDA, por sus siglas en inglés *Functional Data Analysis*). Tal y como prometía la introducción de este texto, de ahora en adelante cobrarán protagonismo objetos estadísticos de carácter infinito-dimensional, especialmente curvas o funciones, y para tratar con ellos será necesario presentar nuevas técnicas estadísticas que permitan su estudio.

La teoría del FDA es relativamente reciente. La publicación de la primera edición del libro de Ramsay y Silverman [18], en el año 1997, se considera el detonante de la explosiva expansión de esta área de investigación. De hecho, es en esta monografía donde se acuña el nombre por el que hoy se conoce la disciplina. El trabajo de Ramsay y Silverman no se centra en establecer la teoría matemática con absoluto rigor, sino en presentar conceptos avanzados de forma asequible para investigadores de otras áreas y profundizar en sus aplicaciones.

Desde entonces, numerosas publicaciones en el área han ido surgiendo para cubrir distintos rincones de la teoría estadística en datos funcionales: [1], sobre series funcionales en espacios de Banach; [8], sobre aspectos no paramétricos del FDA; [11] y [12], acerca de cuestiones de inferencia (contrastes de hipótesis) y fundamentos teóricos, respectivamente, en el contexto de espacios de Hilbert; entre otros. Una última referencia es el libro de texto *Introduction to Functional Data Analysis* [14], el cual se siguió durante todo el proceso de documentación para la elaboración de este trabajo. A estos libros se suma una gran cantidad de artículos de los últimos 15 o 20 años que traen avances al ámbito del FDA.

Dejando atrás las notas históricas, hagamos un breve comentario acerca de los contenidos de este capítulo. Comenzaremos introduciendo el marco teórico en el que nos moveremos, espacios

de funciones y funciones aleatorias serán tratados en las dos primeras secciones. En la tercera definiremos los conceptos análogos a la media y varianza de una variable aleatoria para funciones aleatorias, los cuales son, como cabría esperar, pilares esenciales sobre los que se apoya la teoría de la regresión. En la parte final de esta sección enunciaremos algunos teoremas fundamentales de la estadística en el contexto de datos funcionales. Las secciones cuarta y quinta se centran en conceptos de interés más aplicado (aunque las ideas que presentaremos en ellas serán instrumentales para el desarrollo de ciertas cuestiones en cuanto a la estimación en la regresión), como lo son las componentes principales funcionales y el suavizado de datos funcionales. En particular, las componentes principales funcionales tendrán una presencia constante en el trabajo a partir de este capítulo.

2.1. Elección del espacio funcional. Espacios de Hilbert

La primera cuestión que debemos determinar es el contexto matemático en el que se desarrollará la teoría estadística de datos funcionales. Como el objetivo del FDA es el estudio de funciones como objetos estadísticos, la elección del espacio de funciones en el que trabajaremos es de evidente importancia.

Lo más habitual es que dicho espacio sea un espacio de Hilbert, o bien el caso más general de un espacio de Banach. El estudio de los espacios de Banach definitivamente tiene interés, pero la ausencia de un producto interior añade una considerable dosis de complejidad que se escapa del alcance de este trabajo. Remitimos a un lector interesado en la teoría de elementos aleatorios en espacios de Banach a [1]. De ahora en adelante trataremos exclusivamente con espacios de Hilbert, que además serán siempre reales.

Consideramos que es valioso dar un breve repaso de la definición de espacio de Hilbert. Un espacio prehilbertiano H es un espacio vectorial, posiblemente de dimensión infinita (y lo será para las aplicaciones que nos conciernen en este texto), dotado de un *producto escalar* o *interior* $\langle \cdot, \cdot \rangle: H \times H \rightarrow \mathbb{R}$. Este producto escalar define la geometría del espacio y permite hablar de algunos conceptos a los que estamos acostumbrados en espacios euclídeos, como la ortogonalidad. Un producto escalar debe cumplir tres propiedades:

1. Linealidad en el primer argumento: $\langle \alpha f + g, h \rangle = \alpha \langle f, h \rangle + \langle g, h \rangle$
2. Simetría: $\langle f, g \rangle = \langle g, f \rangle$
3. Positividad y definición: $\langle f, f \rangle \geq 0$ y $\langle f, f \rangle = 0 \iff f = 0$.

donde en las tres condiciones $\alpha \in \mathbb{R}$, $f, g, h \in H$. Notamos que de 1 y 2 se sigue la bilinealidad del producto escalar.

Gracias a las condiciones enunciadas, un producto escalar define una norma de manera na-

tural:

$$\begin{aligned}\|\cdot\|: H &\longrightarrow [0, \infty) \\ f &\longmapsto \|f\| = \sqrt{\langle f, f \rangle}\end{aligned}$$

Con esto, el espacio prehilbertiano H es además un espacio normado $(H, \|\cdot\|)$. Esta norma a su vez induce en H una estructura de espacio métrico mediante la distancia $d(f, g) = \|f - g\|$ y dota a H de una topología.

Definición 2.1. Un espacio de Hilbert es un espacio prehilbertiano $(H, \langle \cdot, \cdot \rangle)$ completo con la topología inducida por el producto escalar.

En el análisis funcional es absolutamente fundamental el concepto de *operador*. Un operador (lineal) entre espacios de Hilbert H y W es una aplicación lineal $T: H \longrightarrow W$. Diremos que T es *acotado*, y lo denotaremos por $T \in \mathcal{LB}(H, W)$, si tiene norma de operadores finita:

$$\|T\| := \sup_{\|h\|_H \leq 1} \|T(h)\|_W < \infty.$$

En este texto siempre trataremos con operadores acotados y por lo tanto en ocasiones nos referiremos a ellos simplemente como operadores. Es sencillo probar que la propiedad de acotación $\|T\| < \infty$ equivale a la continuidad de T . Son operadores de especial importancia los *funcionales*, aquellos con codominio \mathbb{R} , los cuales conforman el dual continuo $H^* = \mathcal{LB}(H, \mathbb{R})$ del espacio. Para los endomorfismos usaremos la notación simplificada $\mathcal{LB}(H) = \mathcal{LB}(H, H)$.

Muchos otros conceptos y resultados propios del análisis funcional son instrumentales para el desarrollo de la teoría del FDA. Sin embargo, debido a restricciones de espacio, no es posible presentar en el cuerpo principal del texto todos los fundamentos necesarios de la teoría de espacios de Hilbert. Por este motivo, hemos decidido incluir en el Anexo A las definiciones y los teoremas pertinentes, de modo que el lector pueda consultarlos cuando sea necesario.

El espacio de Hilbert (de dimensión infinita) más habitual es el espacio L^2 de funciones cuadrado-Lebesgue-integrables sobre un espacio de medida $\mathcal{D} \equiv (\mathcal{D}, \Sigma, \nu)$, donde Σ se trata de un σ -álgebra sobre el conjunto \mathcal{D} y ν una medida en (\mathcal{D}, Σ) :

$$L^2(\mathcal{D}) = \left\{ f: \mathcal{D} \longrightarrow \mathbb{R} / f \text{ medible y } \int_{\mathcal{D}} f(x)^2 d\nu(x) < \infty \right\},$$

donde la integral que aparece es la integral de Lebesgue. Este espacio se dota del siguiente producto escalar:

$$\langle f, g \rangle = \int_{\mathcal{D}} f(x)g(x) d\nu(x).$$

Es importante notar que para que $L^2(\mathcal{D})$ así definido sea en efecto un espacio de Hilbert es necesario cocientarlo por la relación de equivalencia que identifica funciones iguales en casi todo

punto. Nosotros realizaremos esta identificación de manera informal y no nos preocuparemos por otros detalles similares acerca de la teoría de la medida, considerando, por ejemplo, que las funciones siempre están definidas en todo el intervalo.

Como la mayoría de datos funcionales se presentan en forma de curvas, el espacio $L^2([a, b])$ será el que manejemos en la práctica, y habitualmente consideraremos que el intervalo de definición es $[0, 1]$ por comodidad. De todas formas, la teoría inicial se desarrollará para un espacio de Hilbert separable¹ arbitrario, por lo que se extiende también a otros espacios que puedan resultar de interés, como los espacios de Sobolev.

A lo largo de este capítulo, sea $(H, \langle \cdot, \cdot \rangle)$ un espacio de Hilbert separable. Como para las aplicaciones que nos interesan H será un espacio de funciones, nos referiremos a sus elementos como tal.

2.2. Elementos aleatorios en un espacio de Hilbert

Buscamos realizar un estudio estadístico que involucre funciones aleatorias, por lo que debemos comenzar dando una definición apropiada de este concepto. Para ello, sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad, que asumiremos que se trata del espacio común sobre el que se definen todos los elementos aleatorios de ahora en adelante.

Antes de dar la definición de un elemento aleatorio sobre un espacio de Hilbert, recordamos que una *variable aleatoria* X es una aplicación medible $X: \Omega \rightarrow \mathbb{R}$, donde el codominio \mathbb{R} es dotado de la σ -álgebra de Borel. De manera completamente análoga se tiene la siguiente definición:

Definición 2.2. Decimos que una aplicación $X: \Omega \rightarrow H$ es una *función aleatoria* si es medible. Esto es, si

$$X^{-1}(C) \in \mathcal{A}, \quad \forall C \in \mathcal{G},$$

donde \mathcal{G} es la σ -álgebra de Borel de H (con la topología inducida por el producto escalar).

Notemos que si X es una función aleatoria, por la continuidad de $\|\cdot\|$ se tiene que $\|X\|: \Omega \rightarrow \mathbb{R}$ es una variable aleatoria.

Toda función aleatoria X induce una medida de probabilidad μ en el espacio en el que toma valores, la cual llamamos su *distribución*, definida como:

$$\mu(C) = \mathbb{P}(X \in C) = \mathbb{P}(\{\omega \in \Omega / X(\omega) \in C\}) = \mathbb{P}(X^{-1}(C)), \quad C \in \mathcal{G}$$

¹Decimos que un espacio topológico H es separable si posee un subespacio numerable denso en él. En el caso de espacios de Hilbert, esta condición equivale a que H admita una base de Hilbert (Definición A.5) numerable.

Definimos a continuación la noción de convergencia en distribución para espacios de Hilbert, la cual es un poco más abstracta que en el caso de variables aleatorias.

Definición 2.3. Sean μ y μ_n , $n \in \mathbb{N}$, medidas de probabilidad en (H, \mathcal{G}) . Decimos que μ_n converge débilmente a μ si $\int_H f(h) d\mu_n(h) \rightarrow \int_H f(h) d\mu(h)$ cuando $n \rightarrow \infty$, para toda función continua y acotada $f: H \rightarrow \mathbb{R}$.

Sean X y X_n , $n \in \mathbb{N}$, funciones aleatorias en H . Decimos que X_n converge en distribución a X si μ_n converge débilmente a μ , siendo μ_n y μ las distribuciones de X_n y X , respectivamente.

A continuación mostramos un resultado que nos da una caracterización especialmente útil de la distribución de funciones aleatorias. Una demostración de dicho resultado puede encontrarse en [12].

Teorema 2.4. Sea X una aplicación $X: \Omega \rightarrow H$. Entonces,

1. X es medible si y solo si $\langle X, f \rangle$ es medible para todo $f \in H$.
2. Si X es una función aleatoria, su distribución está unívocamente determinada por las distribuciones, llamadas marginales, de $\langle X, f \rangle$ con $f \in H$.

El teorema anterior motiva la siguiente definición:

Definición 2.5. Sea X una función aleatoria en H . Diremos que X es *gaussiana* si $\langle X, f \rangle$ es una variable aleatoria normal para todo $f \in H$.

Es posible definir la distribución gaussiana de manera más abstracta a través del funcional característico (véase [14]). Nosotros escogemos tratar directamente con esta caracterización por simplicidad. Con estos conocimientos acerca de la convergencia en distribución y la distribución gaussiana de funciones aleatorias, estamos cerca de encontrarnos en condiciones de enunciar el teorema central del límite para elementos aleatorios en un espacio de Hilbert. Sin embargo, para ello tendremos que esperar hasta el final de la siguiente sección, cuando hayamos finalmente presentado la media y el operador de covarianza (junto con sus estimadores empíricos) de funciones aleatorias.

2.3. Esperanza y operador de covarianza de funciones aleatorias

Resulta de gran importancia definir las nociones de esperanza y covarianza de funciones aleatorias, ya que son esenciales para el estudio de la regresión. De nuevo, lo que la teoría del FDA pretende es adaptar estos conceptos ya existentes al contexto de las funciones aleatorias, conservando sus propiedades definatorias.

Comenzamos presentando la definición de la esperanza de un elemento aleatorio en un espacio de Hilbert.

Definición 2.6. Una función aleatoria X se dice *integrable* si $\mathbb{E}\|X\| < \infty$. Para toda función aleatoria integrable existe una función $\mu \in H$ que verifica la siguiente propiedad:

$$\mathbb{E} \langle X, h \rangle = \langle \mu, h \rangle, \quad \forall h \in H.$$

Diremos entonces que μ es la *esperanza* o *media* de X , y lo denotaremos por $\mathbb{E}[X] = \mu$.

Notamos que la condición de finitud del primer momento de $\|X\|$ garantiza la existencia de μ , pero no se trata de una condición necesaria. El argumento que exponemos para la demostración de la suficiencia recae sobre el teorema de representación de Riesz (Teorema A.7), el cual es uno de los teoremas básicos de la teoría de espacios de Hilbert.

Demostración. Sea $h \in H$ arbitrariamente fijado. Consideremos la variable aleatoria $\langle X, h \rangle$. Esta es en efecto una función medible de Ω en \mathbb{R} , pues X es medible y $\langle \cdot, h \rangle : H \rightarrow \mathbb{R}$ es continua. Además tiene primer momento finito, ya que

$$\mathbb{E} |\langle X, h \rangle| \leq \mathbb{E} (\|X\| \|h\|) = \|h\| \mathbb{E} \|X\| < \infty,$$

donde en la primera desigualdad hemos aplicado la desigualdad de Cauchy-Schwarz (junto con la monotonía de la esperanza), y en la última la hipótesis de integrabilidad de X .

Con esto el siguiente funcional está bien definido:

$$L: h \in H \mapsto L(h) = \mathbb{E} \langle X, h \rangle \in \mathbb{R}$$

Además, es trivialmente lineal pues lo son el producto escalar y la esperanza de variables aleatorias. De la desigualdad anterior se sigue también que es acotado, pues $|L(h)| \leq \mathbb{E} \|X\| \|h\| \Rightarrow \|L\| \leq \mathbb{E} \|X\| < \infty$.

Con esto L pertenece al dual topológico de H , $L \in H^*$, y el teorema de representación de Riesz garantiza la existencia de un elemento $\mu \in H$ que lo caracteriza de la siguiente forma: $L(h) = \langle \mu, h \rangle$, $\forall h \in H$. Comprobamos siguiendo la definición de L que μ verifica la propiedad buscada. \square

Observación 2.7. Esta demostración ilustra cómo la media de una función aleatoria puede entenderse no solo como una función, sino también como un funcional de forma natural.

Proposición 2.8. Sean X e Y funciones aleatorias integrables en H . La esperanza de funciones aleatorias verifica las siguientes propiedades:

1. $\|\mathbb{E}[X]\| \leq \mathbb{E}\|X\|$.
2. $\mathbb{E}[\lambda X + \zeta Y] = \lambda \mathbb{E}[X] + \zeta \mathbb{E}[Y]$.
3. Sea W un espacio de Hilbert y $T \in \mathcal{LB}(H, W)$. Entonces, $T(X)$ es una función aleatoria integrable en W con media $\mathbb{E}[T(X)] = T(\mathbb{E}[X])$.

La segunda y tercera de estas propiedades son una extensión de la propiedad de linealidad de la esperanza de variables aleatorias.

Centrémonos ahora en el otro momento más importante: la varianza, o en el caso de funciones aleatorias, el operador de covarianza. Este se define como sigue:

Definición 2.9. Una función aleatoria X se dice *cuadrado-integrable* si $\mathbb{E}\|X\|^2 < \infty$. El operador de covarianza de una función aleatoria cuadrado-integrable se define como:

$$C: H \longrightarrow H$$

$$h \longmapsto C(h) = \mathbb{E}[\langle X - \mu, h \rangle (X - \mu)],$$

donde denotamos $\mu = \mathbb{E}[X]$. En caso de existir posibilidad de confusión, por ejemplo si se trabaja con más de una función aleatoria, se indicará la variable referida a través de un subíndice: $C \equiv C_X$. Si X es una función aleatoria cuadrado-integrable, C es un operador lineal acotado en H .

Demostración. Veamos que $C \in \mathcal{LB}(H)$. Primero notamos que C está bien definido ya que, para todo $h \in H$, $\langle X - \mu, h \rangle (X - \mu)$ es una función aleatoria integrable:

$$\begin{aligned} \mathbb{E}\|\langle X - \mu, h \rangle (X - \mu)\| &= \mathbb{E}\left[|\langle X - \mu, h \rangle| \|X - \mu\|\right] \\ &\leq \mathbb{E}\left[\|X - \mu\|^2\right] \|h\| \\ &= (\mathbb{E}\|X\|^2 - \|\mu\|^2) \|h\| < \infty \end{aligned}$$

Además, C es trivialmente lineal por serlo el producto escalar y la esperanza de funciones aleatorias. Resta comprobar que es acotado. Por la primera propiedad de la Proposición 2.8, del desarrollo anterior se deduce que $\|C\| \leq \mathbb{E}\|X\|^2 - \|\mu\|^2 < \infty$. \square

El operador de covarianza es una adaptación natural del segundo momento (centrado) de una función aleatoria si, en lugar de intentar generalizar la varianza, consideramos la matriz de covarianzas de una variable aleatoria vectorial. En efecto, cuando el espacio de Hilbert es $H = \mathbb{R}^d$, la aplicación lineal inducida por la matriz de varianzas-covarianzas coincide con el operador, por lo que podemos identificar ambos conceptos.

La siguiente proposición muestra como, al igual que la esperanza de funciones aleatorias, el operador de covarianza tiene una interpretación natural en términos del producto escalar.

Proposición 2.10. *Sea X una función aleatoria y C su operador de covarianza. Entonces, para todo $f, g \in H$ se tiene:*

$$\langle C(f), g \rangle = \langle f, C(g) \rangle = \mathbb{E}[\langle X - \mu, f \rangle \langle X - \mu, g \rangle] = \text{Cov}(\langle X, f \rangle, \langle X, g \rangle).$$

En particular, la forma cuadrática asociada a C viene dada por:

$$\langle C(h), h \rangle = \mathbb{E} \langle X - \mu, h \rangle^2 = \text{Var} \langle X - \mu, h \rangle = \text{Var} \langle X, h \rangle, \quad h \in H.$$

La siguiente caracterización de los operadores de covarianza es de especial importancia:

Teorema 2.11. *Un operador $C \in \mathcal{LB}(H)$ es el operador de covarianza de una función aleatoria cuadrado-integrable si y solo si*

1. C es autoadjunto: $\langle C(f), g \rangle = \langle f, C(g) \rangle, \forall f, g \in H$;
2. C es semidefinido positivo: $\langle C(f), f \rangle \geq 0, \forall f \in H$;
3. C es nuclear, esto es, sus autovalores satisfacen $\sum_{k=1}^{\infty} \lambda_k < \infty$.

La tercera propiedad nos dice, en particular, que todo operador de covarianza es Hilbert-Schmidt, y con ello compacto (Definiciones A.13 y A.11). Una consecuencia notable de este hecho es que C nunca será (salvo en el caso de dimensión finita) un operador invertible.

Observación 2.12. Por ser C un operador autoadjunto, induce una descomposición ortogonal del espacio de la siguiente forma:

$$H = \ker C \oplus \overline{\text{Im } C},$$

pues se tiene $(\text{Im } C)^\perp = \ker C$.

La principal y más importante consecuencia de esta descomposición es que $X - \mu \in \overline{\text{Im } C}$ c.s. Consecuentemente, el estudio probabilístico de la función aleatoria puede reducirse al subespacio $\overline{\text{Im } C} \subset H$. Aquí podemos diferenciar dos casos:

- Si $\dim \overline{\text{Im } C} = d < \infty$ entonces la función aleatoria $X - \mu$ puede ser identificada con un vector aleatorio en \mathbb{R}^d , y su estudio puede llevarse a cabo con las herramientas conocidas del caso multivariante.
- En caso contrario, $\dim \overline{\text{Im } C} = \infty$, $\overline{\text{Im } C}$ es a su vez un espacio de Hilbert separable de dimensión infinita (con la estructura heredada de H), por lo que la teoría que desarrollamos puede aplicarse de manera inmediata a $X - \mu$ como un elemento aleatorio en dicho espacio.

Veamos una breve demostración de esta propiedad. Para simplificar la notación supondremos que $\mu = \mathbb{E}[X] = 0$. Sea $\{\eta_j\}_{j \in \mathbb{N}}$ una base de Hilbert de $\ker C$. Así, dado $h \in H$ se tiene que $h \in \overline{\text{Im } C}$ si y solo si $\langle h, \eta_j \rangle = 0, \forall j \in \mathbb{N}$. Ahora, por la Proposición 2.10 tenemos que $\text{Var} \langle X, \eta_j \rangle = \langle \eta_j, C(\eta_j) \rangle = 0, \forall j \in \mathbb{N}$. Luego, con probabilidad 1 tenemos $\langle X, \eta_j \rangle = \langle \mathbb{E}[X], \eta_j \rangle = 0$ y, ya que la intersección numerable de conjuntos de probabilidad 1 también lo es, se tiene $\mathbb{P}(\langle X, \eta_j \rangle = 0, \forall j \in \mathbb{N}) = \mathbb{P}(X \in \overline{\text{Im } C}) = 1$.

2.3.1. Esperanza y operador de covarianza en el espacio L^2

Antes de proceder a la siguiente sección vamos a intentar restar abstracción a la definición de los dos objetos, media y operador de covarianza, observándolos en el caso del espacio que más trabajaremos, el espacio L^2 . Obtendremos una forma directa de computarlos, y esto además mejorará la intuición que tenemos sobre ellos.

Comencemos por el caso más familiar de la media. Para una función aleatoria en $L^2 \equiv L^2([0, 1])$ es sencillo calcular su media puntual: para cada valor de $t \in [0, 1]$ consideramos la variable aleatoria $X(t)$ y calculamos su media. Esto nos da una función $t \mapsto \mathbb{E}[X(t)]$.

El resultado clave es que esta sencilla forma de representar la media coincide en efecto con la definición dada anteriormente. Esto es,

$$\mu(t) = \mathbb{E}[X](t) = \mathbb{E}[X(t)] \quad , \quad \forall t \in [0, 1].$$

A continuación tratamos el operador de covarianza. Comencemos primero recordando cómo se define la matriz de varianzas-covarianzas de una variable aleatoria vectorial. Si $Y = (Y_1, \dots, Y_d)'$ es tal variable aleatoria, se definen las entradas de su matriz de varianzas-covarianzas como:

$$\text{Cov}(Y, Y)_{ij} = \text{Cov}(Y_i, Y_j), \quad i, j = 1, \dots, d$$

Pensando una función aleatoria como un vector con entradas continuas en lugar de discretas, podemos obtener un objeto similar sustituyendo los índices por argumentos continuos:

$$c(t, s) = \text{Cov}(X(t), X(s)), \quad t, s \in [0, 1]$$

La función $c: [0, 1]^2 \rightarrow \mathbb{R}$ así definida se denomina *función o núcleo de covarianzas*, y está íntimamente relacionada con el operador de covarianza. C es, de hecho, el operador integral asociado al núcleo c ; esto es, para cualquier $h \in L^2$:

$$C(h)(t) = \int_0^1 c(t, s)h(s)ds, \quad t \in [0, 1].$$

Las demostraciones de estos resultados giran en torno a la aplicación del teorema de Fubini.

2.3.2. Estimadores empíricos

Para la estadística es fundamental no solo conocer los objetos teóricos a nivel poblacional, sino cómo tratarlos cuando la información de la que disponemos es una muestra finita. Consideremos entonces observaciones $X_i, i = 1, \dots, n$, de una función aleatoria X en H de la que pretendemos estimar su media y su operador de covarianza.

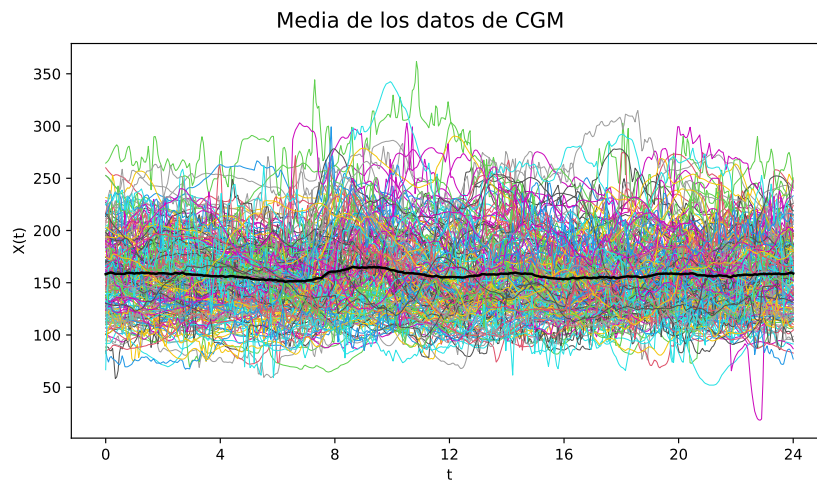


Figura 2.1: Representación de las curvas de CGM junto con su media muestral, resaltada en trazo negro grueso.

La definición de la media muestral y del operador de covarianza muestral es natural y completamente análoga al caso de variables aleatorias.

Definición 2.13. Se define la *media muestral* de X_1, \dots, X_n como

$$\bar{X} \equiv \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \in H.$$

Definición 2.14. Se define el *operador de covarianza muestral* de X_1, \dots, X_n como el operador $\hat{C}: H \rightarrow H$ que actúa de la siguiente forma:

$$\hat{C}(h) \equiv C_n(h) = \frac{1}{n-1} \sum_{i=1}^n \langle h, X_i - \bar{X} \rangle (X_i - \bar{X}), \quad \forall h \in H.$$

Observación 2.15. Notamos que en la definición del operador de covarianza muestral, el factor en el denominador se toma como $n-1$ en lugar de n . Esto convierte a \hat{C} sea un estimador insesgado de C (considerando \hat{C} como elemento aleatorio en el espacio de Hilbert $\mathcal{LB}_{HS}(H)$), al igual que ocurre con la varianza muestral en el caso escalar.

Estas definiciones son aplicables al caso de un espacio de Hilbert H arbitrario. En el caso del espacio habitual L^2 , la estructura conocida y las caracterizaciones de la sección anterior nos permiten formular estos objetos de otra manera.

La media muestral puede calcularse como $\bar{X}(t) = \overline{X(t)} = n^{-1} \sum_{i=1}^n X_i(t)$. En cuanto al operador de covarianza, computamos primero el *núcleo de covarianzas muestral*:

$$\hat{c}(t, s) = \frac{1}{n-1} \sum_{i=1}^n (X_i(t) - \bar{X}(t))(X_i(s) - \bar{X}(s)),$$

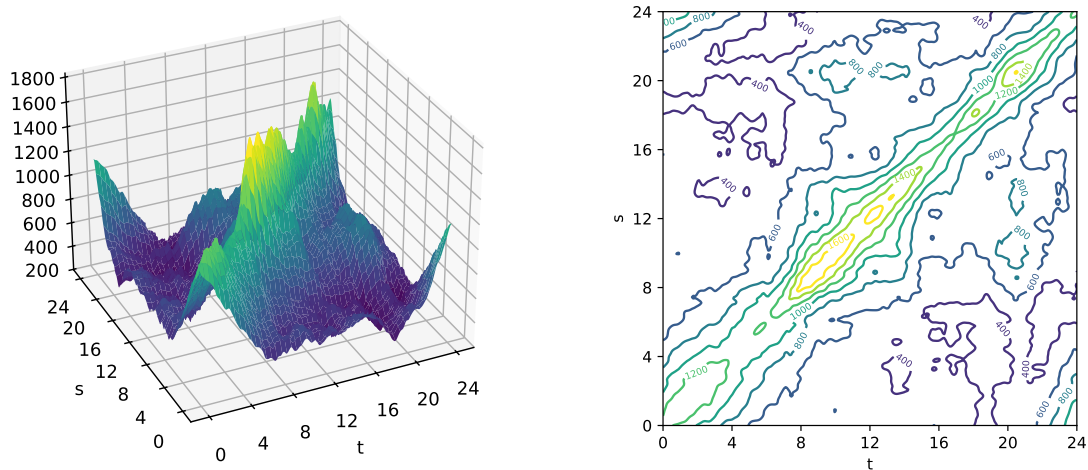


Figura 2.2: Representaciones gráficas del núcleo de covarianzas (estimado) de los datos de CGM. A la izquierda, su gráfica como superficie tridimensional; y a la derecha, el gráfico de sus curvas de nivel.

y con él expresamos el operador de covarianza muestral \hat{C} como el operador integral asociado al núcleo \hat{c} .

En la Figura 2.1 se representa la media muestral de las curvas de CGM, obtenida como el promedio puntual de las observaciones. Por otra parte, la Figura 2.2 muestra su núcleo de covarianzas muestral. Ambas figuras, así como la mayoría de las incluidas en este trabajo, fueron generadas empleando los paquetes de Python `matplotlib` [13] y `scikit-fda` [17].

La media y el operador de covarianza muestrales comparten propiedades importantes con sus contrapartes poblacionales. El proceso por el que se calcula \bar{X} es evidentemente lineal, y por ello las propiedades de linealidad de la media se sostienen también para la media muestral. En cuanto al operador de covarianza, su versión muestral no deja de ser un operador autoadjunto y semidefinido positivo. Una diferencia surge del hecho de que, aunque el operador de covarianza C puede ser definido positivo (con todo autovalor $\lambda_k > 0$), \hat{C} nunca lo será. Es evidente que $\text{Im } \hat{C} \subset \text{span}\{X_i - \bar{X}\}_{i=1}^n$ tiene dimensión finita, a lo sumo $n - 1$, por lo que siempre que H sea de dimensión infinita \hat{C} tendrá un núcleo no trivial. Como operador de rango finito, \hat{C} es trivialmente nuclear.

Como adelantamos a comienzos de este capítulo, cerraremos esta sección enunciando dos de los teoremas fundamentales de la estadística, la ley fuerte de los grandes números y el teorema central del límite, en el contexto de funciones aleatorias. Presentamos ambos en el caso general de un espacio de Hilbert separable H .

Teorema 2.16 (Ley fuerte de los grandes números). Sea X_n , $n \in \mathbb{N}$, una sucesión de funciones aleatorias en H independientes e idénticamente distribuidas a una función aleatoria cuadrado-integrable X . Entonces:

$$\lim_{n \rightarrow \infty} \bar{X}_n = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i = \mathbb{E}[X] \text{ c.s.}$$

Teorema 2.17 (Teorema central del límite). Sea X_n , $n \in \mathbb{N}$, una sucesión de funciones aleatorias en H independientes e idénticamente distribuidas a una función aleatoria cuadrado-integrable X . Entonces:

$$\frac{1}{\sqrt{n}} \left(\sum_{i=1}^n X_i - n\mathbb{E}[X] \right) \xrightarrow{d} G,$$

donde G es una función aleatoria gaussiana de media $\mathbb{E}[G] = 0$ y operador de covarianza $C_G = C_X$.

Pueden consultarse las demostraciones de ambos resultados en el capítulo 7 de [12].

2.4. Análisis de componentes principales funcionales

Ahora que ya tenemos ciertas nociones básicas sobre la teoría del FDA, y en particular el operador de covarianza, podemos adentrarnos en el análisis de componentes principales funcionales. El análisis de componentes principales es un concepto básico y muy extendido de la estadística multivariante. Permite reducir de manera óptima la dimensionalidad de los datos y se trata de una herramienta decisiva para la realización de análisis exploratorios.

En el contexto del FDA, el análisis de componentes principales funcionales tiene tanta o más importancia que su análogo multivariante. Las componentes principales funcionales (FPC, por sus siglas en inglés *Functional Principal Components*) de una función aleatoria forman la base de Hilbert óptima para aproximar dicha función, y su poder descriptivo sigue siendo extremadamente útil. Además, sus excelentes propiedades las convierten en herramientas valiosas para el desarrollo de la teoría de estimación e inferencia en regresión, como veremos más adelante en este trabajo.

La definición de las FPC es esencialmente análoga a su versión multivariante. Sea X , a lo largo de esta sección, una función aleatoria cuadrado-integrable en L^2 . Dada una función de pesos ψ con norma unidad, $\|\psi\| = 1$, podemos considerar el análogo a una combinación lineal estandarizada de un vector aleatorio, que en este contexto se obtiene mediante el producto escalar:

$$\xi = \langle \psi, X \rangle = \int \psi(t)x(t)dt.$$

La variable aleatoria ξ se denomina *score* asociado a la componente principal ψ .

La primera componente principal funcional de X es aquella función ψ_1 que maximiza la varianza del *score* asociado, es decir:

$$\psi_1 = \arg \max_{\substack{\psi \in L^2 \\ \|\psi\|=1}} \text{Var} \langle \psi, X \rangle.$$

Conocida la primera FPC ψ_1 , busquemos ψ_2 maximizando de nuevo la varianza de su *score*, pero ahora sujeta a la condición de ortogonalidad $\langle \psi_2, \psi_1 \rangle = 0$, que obliga a que ψ_2 represente un modo de variación distinto del de ψ_1 . En general, si hemos computado las $k - 1$ primeras FPCs ψ_j , $j = 1, \dots, k - 1$, la k -ésima componente principal funcional se obtiene como solución del problema de optimización:

$$\begin{aligned} & \max_{\psi \in L^2} \text{Var} \langle \psi, X \rangle \\ & \text{s. a } \|\psi\| = 1 \\ & \langle \psi, \psi_j \rangle = 0, \quad j = 1, \dots, k - 1 \end{aligned} \tag{2.1}$$

A continuación daremos una caracterización más interesante de las FPC. Observamos que por la Proposición 2.10 la varianza de los *scores* puede calcularse haciendo uso de la siguiente identidad:

$$\text{Var} \langle \psi, X \rangle = \text{Var} \langle \psi, X - \mu \rangle = \langle \psi, C(\psi) \rangle = \iint \psi(t)c(t,s)\psi(s) dt ds. \tag{2.2}$$

Recordamos ahora que el operador de covarianzas C de X es un operador Hilbert-Schmidt (luego compacto) autoadjunto y semidefinido positivo. Por lo tanto, posee autovalores reales positivos $\lambda_k > 0$, $k \in \mathbb{N}^2$, ordenados de modo que

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k \geq \dots > 0.$$

Por el teorema espectral (Teorema A.12), es posible tomar una base de autofunciones (normalizadas) $\{\psi_k\}_{k \in \mathbb{N}}$ de $\overline{\text{Im } C}$, donde cada ψ_k es la autofunción asociada al autovalor λ_k . Gracias a la identidad (2.2) se puede demostrar (Teorema 4.2.5 de [12]) que las soluciones de los problemas de optimización (2.1) son precisamente las correspondientes autofunciones ψ_k ; es decir, ψ_k es la k -ésima FPC, con $k \in \mathbb{N}$.

Observación 2.18. Es importante notar que, a priori, esta definición no garantiza la unicidad de las FPC. Podría darse, por ejemplo, el caso en que el espacio propio asociado a λ_1 fuese de dimensión 2, lo que implicaría la existencia de múltiples direcciones ψ_1 —tantas como puntos en

²Pueden darse dos casos: o bien los autovalores positivos λ_k forman una sucesión convergente a 0, o bien existe solamente una cantidad finita de autovalores no nulos. En este último caso, $\overline{\text{Im } C}$ es un espacio de dimensión finita, y por lo tanto no se trata del punto de interés de este trabajo.

la circunferencia— que serían autofunciones del mismo. En este caso, todas ellas se tratarían de soluciones óptimas para el problema de maximización de la varianza, y nos encontraríamos con ambigüedad para elegir la primera FPC.

No obstante, esto no es una posibilidad demasiado realista en la práctica, pues trabajando empíricamente los autovalores son siempre distintos. Por lo tanto, es razonable suponer adicionalmente que se tiene la cadena de desigualdades estricta: $\lambda_1 > \lambda_2 > \dots > 0$ (o bien que los autovalores son distintos hasta cierto índice p , lo cual es suficiente para la mayoría de aplicaciones). Con esto sí queda garantizada la unicidad de las (p primeras) FPC, salvo por la elección de signo.

Una vez conocidas las FPC es posible expresar X en la base $\{\psi_k\}_{k \in \mathbb{N}}$:

$$X - \mu = \sum_{k=1}^{\infty} \xi_k \psi_k, \quad (2.3)$$

donde hemos denotado $\mu = \mathbb{E}[X]$ y $\xi_k = \langle X - \mu, \psi_k \rangle$.

La Ecuación (2.3) recibe el nombre de expansión de Karhunen-Loève. Notamos que, puesto que $C_X = C_{X-\mu}$, las FPC que intervienen están asociadas tanto a X como a $X - \mu$.

Observación 2.19. La Observación 2.12 justifica la expansión de Karhunen-Loève, ya que, a pesar de que las FPC constituyen una base de $\overline{\text{Im } C}$ y no de L^2 , eliminar la componente de $X - \mu$ en $\ker C$ no supone una pérdida de información.

A continuación enunciamos ciertas propiedades interesantes relacionadas con la expansión de Karhunen-Loève y los *scores* ξ_k :

1. $\mathbb{E}[\xi_k] = 0$.
2. $\text{Var}(\xi_k) = \mathbb{E}[\xi_k^2] = \lambda_k$.
3. $\text{Cov}(\xi_k, \xi_l) = \mathbb{E}[\xi_k \xi_l] = 0$, si $k \neq l$.
4. $\mathbb{E} \|X - \mu\|^2 = \sum_{k=1}^{\infty} \lambda_k = \text{tr}(C)$.

La expansión de Karhunen-Loève es de especial interés en el caso gaussiano. Este caso, que estudiamos en el ejemplo a continuación, ilustra adicionalmente la importancia de las propiedades anteriores, por lo que merece la pena detenerse en él.

Ejemplo 2.20. Sea X una función aleatoria gaussiana y escribamos su expansión de Karhunen-Loève $X = \mu + \sum_{k=1}^{\infty} \xi_k \psi_k$, con $\xi_k = \langle X - \mu, \psi_k \rangle$. Por la definición de la distribución gaussiana de funciones aleatorias, los *scores* son variables aleatorias normales. Las propiedades 1 y 2 especifican la media y varianza de sus distribuciones: $\xi_k \sim \mathcal{N}(0, \lambda_k)$. Por último, la tercera propiedad garantiza que los *scores* son incorrelados, y por lo tanto independientes.

Por las observaciones anteriores sobre los *scores* concluimos que X es igualmente distribuida a la función aleatoria $\mu + \sum_{k=1}^{\infty} \sqrt{\lambda_k} Z_k \psi_k$, donde las Z_k , $k \in \mathbb{N}$, son variables aleatorias normales estándar independientes. En consecuencia, la distribución de una función aleatoria gaussiana X se caracteriza por su media $\mu = \mathbb{E}[X]$ y su operador de covarianza C_X (a través de su descomposición espectral), de manera análoga a lo que ocurre en los casos escalar y multivariante.

La cuarta de las propiedades anteriores puede interpretarse de la siguiente forma: la variabilidad total de la función aleatoria X se descompone en la suma de las varianzas de las proyecciones en la dirección de las distintas componentes principales funcionales. Como los autovalores se ordenan de manera decreciente, es esperable que un número pequeño de FPC pueda explicar con alta precisión el comportamiento de X .

Este es de hecho un gran foco de interés del análisis de componentes principales funcionales, pues hace que las FPC formen la base de Hilbert óptima para la expansión de X . Precisamos en la siguiente proposición lo que se entiende por esta optimalidad.

Proposición 2.21. *Sea X una función aleatoria cuadrado-integrable en L^2 y sean $\{\psi_k\}_{k \in \mathbb{N}}$ sus componentes principales. Entonces, la elección de un sistema ortonormal con p elementos $\{u_k\}_{k=1}^p$ que minimiza la distancia cuadrática media*

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

es precisamente $\{\psi_k\}_{k=1}^p$.

2.4.1. Componentes principales funcionales estimadas

Las FPC son una herramienta teórica muy potente, pero para que los métodos derivados de ellas puedan ser aplicados en un contexto práctico es necesario proporcionar métodos para su cálculo o estimación a partir de una muestra.

La forma natural de definir las componentes principales funcionales estimadas (EFPC, *Estimated Functional Principal Components*) es a partir de la descomposición espectral del operador de covarianza muestral \hat{C} . Recordamos que \hat{C} también es un operador autoadjunto y semidefinido positivo, por lo que todos sus autovalores, que denotaremos por $\hat{\lambda}_k$, son reales y no negativos. Supondremos que están ordenados de modo que $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq 0$. Así, definimos la k -ésima EFPC $\hat{\psi}_k$ como la autofunción normalizada asociada al autovalor $\hat{\lambda}_k$ de \hat{C} .

En la Figura 2.3 pueden observarse las cuatro primeras componentes principales funcionales estimadas de la variable funcional X de los datos de CGM.

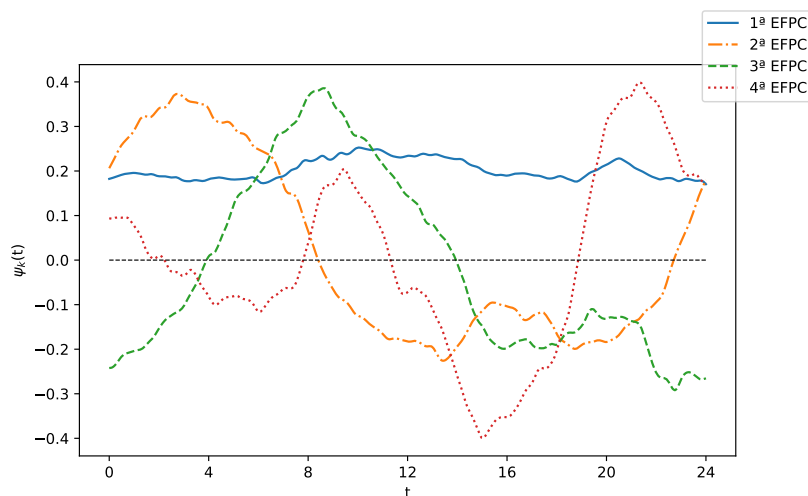


Figura 2.3: Gráfica de las cuatro primeras EFPC de los datos de CGM.

Una primera diferencia notable con respecto a las FPC es que nunca hay más que un número finito de EFPC. El operador de covarianza C puede admitir una cantidad infinita numerable de autofunciones asociadas a distintos autovalores positivos. Sin embargo, por la observación realizada en la Sección 2.3.2, sabemos que si \hat{C} se obtiene a partir de una muestra de n observaciones, a lo sumo se tendrán $n - 1$ EFPC $\hat{\psi}_k$ asociadas a autovalores no nulos. Esto refleja el hecho de que una muestra finita solamente puede aportar información limitada acerca de un objeto de complejidad infinita.

2.5. De datos discretos a datos funcionales: breve revisión de las técnicas de presuavizado

Hasta ahora hemos trabajado a nivel puramente teórico, pero si pretendemos que la teoría del FDA se pueda aplicar al estudio estadístico de datos reales tenemos que abordar una cuestión que hemos ignorado hasta ahora.

La forma más habitual en la que se presentan los datos funcionales es la siguiente: se tienen n realizaciones de un experimento, y para cada una de ellas se toman medidas de forma que los puntos obtenidos trazan una curva. Esto es, contamos con una muestra de la forma:

$$x_i(t_{i,j}) \in \mathbb{R}, t_{i,j} \in [a, b], j = 1, \dots, J_i, i = 1, \dots, n$$

Tenemos entonces n observaciones de curvas $x_i(t), t \in [a, b]$, de las cuales sus valores son conocidos únicamente en una cantidad finita de puntos $t_{i,j}$ (que pueden ser diferentes para distintas observaciones), los cuales asumimos que son lo suficientemente densos como para representar

fielmente la curva.

El primer paso en cualquier estudio de FDA es transformar estas observaciones “incompletas” en objetos funcionales con los que sí podemos trabajar, proceso que se conoce por el nombre de *presuavizado*. Si este paso se obvia, el tratamiento que podemos hacer de los datos es muy limitado. Por ejemplo, en muchos casos es interesante analizar las derivadas de la función observada, lo cual no es posible a menos que trabajemos con la propia función en lugar del conjunto de puntos original.

Como en esta sección no nos preocuparemos de cuestiones muestrales, pasamos a trabajar únicamente con una realización de nuestra función aleatoria. Supondremos que se han observado valores $y_j = x(t_j)$ para ciertos instantes $t_j \in [0, 1]$, $j = 1, \dots, J$, y que deseamos obtener una función $x(t)$ que esté definida en todo el intervalo $[0, 1]$.

Hay muchas formas de representar el conjunto de puntos mediante una función. Una posibilidad, siempre que no necesitemos imponer condiciones de diferenciabilidad, es simplemente unir los puntos con rectas. Esto es, interpolar linealmente el conjunto de puntos para extender la definición de la función a todo el intervalo.

Este enfoque hace que la función ajustada pase por todos los puntos. Sin embargo, esto no tiene por qué ser deseable. Es posible que en el proceso de toma de datos exista un error de medición o algo similar que introduzca ruido en los puntos observados, lo que podría modelarse como $y_j = x(t_j) + \delta_j$, donde δ_j es el término de error aleatorio (del cual se asume que tiene media nula). Por lo tanto, forzar a que la función pase por cada punto no refleja necesariamente el comportamiento real del suceso que estudiamos. Es por esto que de manera habitual se hace uso de métodos que suavizan los valores observados, con lo que la función x ya no presenta tantos picos ni cambios erráticos. Con esta filosofía el objetivo pasa a ser aproximar los puntos en lugar de interpolarlos.

Existen otras consideraciones importantes que tener en cuenta a la hora de convertir los datos observados en objetos funcionales. Por ejemplo, si el suceso estudiado es no negativo por definición (mediciones de precipitación a lo largo de un año, por ejemplo) no es admisible que la aproximación del conjunto de puntos cruce el eje de abscisas. Una situación similar ocurre si la función subyacente debe ser necesariamente monótona (la curva de crecimiento de un niño, por ejemplo). También conlleva una modificación considerable de la metodología el caso de las funciones dispersamente observadas (lo que se conoce como *sparsity*). Todos estos casos son abordados, por ejemplo, en [18] y [14], respectivamente.

Nosotros no nos preocuparemos por ninguna de estas cuestiones y nos centraremos en el caso más sencillo de curvas densamente observadas sin restricciones. En las dos secciones a continuación presentamos dos de los métodos de suavizado más extendidos.

2.5.1. Expansión en base. Suavizado por mínimos cuadrados

La manera más directa de suavizar datos funcionales es por medio de una expansión en base. A lo largo de este texto, la noción de base con la que trabajaremos es la siguiente. Llamamos *base* de L^2 a una familia de funciones $\mathcal{B} = \{\phi_\alpha\}_{\alpha \in \mathcal{A}}$ tal que toda función $f \in L^2$ puede ser aproximada (en norma L^2) con precisión arbitraria por una combinación lineal de elementos de \mathcal{B} . Con esta definición abarcamos todas las técnicas de aproximación que emplearemos en este trabajo.

En particular, bajo este concepto de base se encuentran las *bases ortonormales* o *de Hilbert*. Recordamos que una base de Hilbert es una base numerable $\{\phi_k\}_{k \in \mathbb{N}}$ de funciones de norma unidad y ortogonales dos a dos ($\langle \phi_k, \phi_l \rangle = 0$ si $k \neq l$). Esta es la noción canónica de base en un espacio de Hilbert, y tendrá enorme importancia a lo largo del texto. El ejemplo más clásico de base ortonormal es el sistema trigonométrico o base de Fourier, formada por senos y cosenos de frecuencias crecientes. Por otro lado, las FPC de toda función aleatoria (cuadrado-integrable) forman también una base de Hilbert.

Además de las bases ortonormales existen otros ejemplos de bases extremadamente habituales. Por ejemplo, los monomios $\{x^k\}_{k \in \mathbb{N}}$ forman una base (no ortonormal) de L^2 . Para todas las bases mencionadas hasta ahora, siempre es posible expresar una función arbitraria $f \in L^2$ como una serie en \mathcal{B} :

$$f = \sum_{k=1}^{\infty} c_k \phi_k,$$

donde la convergencia se tiene en la norma de L^2 .

Sin embargo, este no es el caso para algunos de las bases de uso más extendido en el área, y estas son el motivo de considerar una definición tan amplia. Entre otras, nos referimos a las bases de B-splines y de *wavelets*.

La expansión en base de un objeto funcional simplemente consiste en seleccionar una base $\{\phi_\alpha\}_{\alpha \in \mathcal{A}}$ y un número K de términos en la expansión, y con ello ajustar la función como la combinación lineal de K funciones básicas que mejor aproxime los puntos observados:

$$x(t) \approx \sum_{k=1}^K c_k \phi_k(t). \quad (2.4)$$

El valor de K debe ser lo suficientemente grande como para poder recoger apropiadamente el comportamiento de las observaciones, pero a su vez relativamente pequeño, especialmente en comparación con el número J de valores observados (nunca superando el mismo). Si K es muy grande la aproximación pasará por todos los puntos (o extremadamente cerca) y al hacerlo podrían introducirse comportamientos indeseados. Por lo tanto, la elección de K refleja de manera

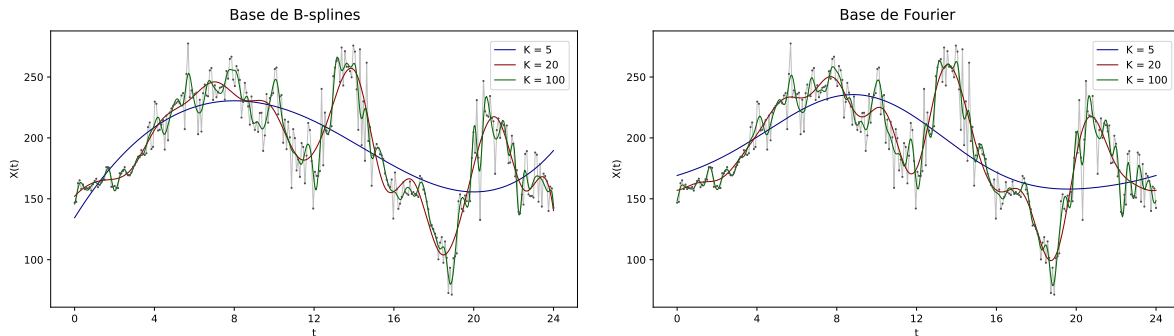


Figura 2.4: Gráficas de una única curva de CGM tras su presuavizado por medio de una expansión en base de B-splines (izquierda) y de Fourier (derecha). La línea y los puntos grises representan los datos discretos sin presuavizado.

implícita la opinión del investigador acerca de lo suave que debe ser la aproximación, y con este planteamiento es tratado como un parámetro discreto de suavizado.

Por otro lado, la elección de la base tiene una clara importancia en esta metodología, pues distintas bases pueden dar lugar a resultados dispares. Es recomendable utilizar funciones básicas con propiedades que reflejen las propiedades naturales del suceso a estudiar. Por ejemplo, la base de Fourier es útil para aproximar funciones periódicas pero resulta poco apropiada si no se espera que la función presente repeticiones en el tiempo. En cambio, las bases de B-splines y *wavelets* ofrecen una gran versatilidad. La primera es muy eficaz para representar funciones suaves, mientras que la segunda se usa particularmente cuando las funciones presentan cambios bruscos e incluso discontinuidades.

En las gráficas izquierda y derecha de la Figura 2.4 puede observarse el presuavizado de una observación de la variable funcional X de los datos de CGM por medio de una expansión en las bases de B-splines y de Fourier, respectivamente. Para cada una de ellas se representan distintos valores del índice de truncamiento K , en particular, $K = 1, 20$ y 100 . Una comparación visual de las curvas en ambas figuras hace evidente los efectos de la elección de K y el tipo de base.

Una vez seleccionado el índice de corte K , los coeficientes $\mathbf{c} = (c_1, \dots, c_K)'$ de la expansión (2.4) que mejor aproximan los valores observados y_j se pueden computar mediante el criterio de mínimos cuadrados. Esto es, se pretende minimizar:

$$SSE(c_1, \dots, c_K) = \sum_{j=1}^J (y_j - x(t_j))^2. \quad (2.5)$$

Este método resulta equivalente a considerar las funciones básicas ϕ_k como covariables en

una regresión lineal múltiple de respuesta x . Los coeficientes se obtienen entonces como

$$\mathbf{c} = (\Phi' \Phi)^{-1} \Phi' \mathbf{y}, \quad (2.6)$$

donde $\Phi_{ij} = \phi_i(t_j)$, $i = 1, \dots, K$, $j = 1, \dots, J$, y $\mathbf{y} = (y_1, \dots, y_J)'$.

De la misma manera que en la regresión lineal múltiple, si las observaciones presentan un error no sujeto al modelo estándar (por ejemplo, errores correlados) se puede adaptar este método de estimación a mínimos cuadrados ponderados.

Sin embargo, existen situaciones en las que los resultados obtenidos a través de las técnicas mencionadas no son satisfactorios. En particular, resultan bastante imprecisos a la hora de estimar derivadas de la función, y el ajuste de K como parámetro de suavizado no es ideal. Otros procedimientos de estimación, como la estimación local no paramétrica de Nadaraya-Watson, pueden dar también buenos resultados, pero presentan su propia serie de problemas.

2.5.2. Suavizado mediante penalización

Es razonable entonces preguntarse si existe alguna alternativa a estos métodos de suavizado que esquive estas complicaciones. Un concepto ya conocido que se presenta como una posibilidad es el de la regularización. Este fue introducido en el ámbito de la regresión desde la década de 1970 en adelante como solución a problemas mal condicionados a causa de multicolinealidad, concretamente en la forma de la regresión Ridge [5] y la regresión Lasso [19], entre otros.

En el ámbito del FDA, el suavizado regularizado mediante penalización es el método de suavizado más aceptado y extendido. A pesar de esto, los métodos anteriormente descritos son suficientes para multitud de usos y su simplicidad puede resultar atractiva.

La idea fundamental detrás del suavizado mediante penalización es controlar el grado de suavizado explícitamente a través de una medida de *roughness* de la siguiente forma:

$$R_L(x) = \int [L(x)(t)]^2 dt,$$

donde L es un operador diferencial lineal. Una elección habitual es la segunda derivada $L = D^2$, con la cual tendríamos $L(x) = x''$, que penaliza la curvatura de la gráfica de la función.

El procedimiento que se sigue es el siguiente: se parte de una expansión en base de la forma (2.4) con un índice de corte K elevado, lo suficiente como para que no juegue un papel restrictivo (incluso podría tomarse K mayor que el número J de puntos observados). A partir de esta se buscan los coeficientes c_k que minimizan la siguiente función objetivo:

$$PSSE(c_1, \dots, c_K) = \sum_{j=1}^J (y_j - x(t_j))^2 + \lambda R_L(x), \quad (2.7)$$

con $\lambda \geq 0$ un parámetro de suavizado.

La función objetivo $PSSE$ no es más que la suma de errores cuadráticos SSE vista en (2.5) a la que se añade el término de penalización $\lambda R_L(x)$. El procedimiento de estimación es también similar, y un desarrollo que omitimos conduce a la siguiente expresión para los coeficientes:

$$\mathbf{c} = (\mathbf{\Phi}'\mathbf{\Phi} + \lambda R)^{-1}\mathbf{\Phi}'\mathbf{y},$$

la cual de nuevo presenta una gran similitud con la obtenida para el criterio de mínimos cuadrados en (2.6). La matriz $R \in \mathcal{M}_{K \times K}$ contiene los valores $R(i, j) = \langle L(\phi_i), L(\phi_j) \rangle = \int L(\phi_i)(s)L(\phi_j)(s)ds$.

Este planteamiento permite al investigador expresar de manera directa, mediante la elección del operador diferencial L y el parámetro de suavizado λ , la clase de suavizado que se pretende lograr. Por ejemplo, $L = D^2$ es una elección estándar para contextos en los que simplemente se desea obtener una función sin cambios repentinos ni picos pronunciados, ya que estos conllevan altos valores de la curvatura x'' . Sin embargo, si se pretende realizar un estudio acerca de la velocidad o aceleración de la curva sería más razonable penalizar derivadas de órdenes mayores tomando $L = D^m$, $m \geq 3$, o similares.

Por otro lado, ajustar el parámetro λ da un control más preciso sobre la intensidad del suavizado: tomar $\lambda = 0$ elimina la penalización, por lo que se obtendrá una función que interpole los puntos (recordamos que el elevado valor de K permitiría que la expansión en base de la función ajuste con demasiada precisión los valores observados). En cambio, hacer tender λ a infinito hará que el término de penalización cobre más y más peso, y la función obtenida será una que haga $L(x) = 0$ (una recta en el caso de $L = D^2$, por ejemplo), ignorando en gran medida cualquier consideración sobre la calidad del ajuste. Es por lo tanto imprescindible encontrar valores de λ que engendren un equilibrio apropiado entre un buen ajuste y un grado de suavizado deseable. En la Sección 3.1.3 y en el Capítulo 4 volveremos a tratar métodos que involucran penalización y nos adentraremos más en profundidad en criterios para la selección de λ .

Para cerrar esta sección, presentamos nuevamente una figura que ilustra el efecto del suavizado mediante penalización, según distintos valores de λ . En la Figura 2.5 se muestran las curvas suavizadas correspondientes a valores de $\lambda = 0.00001$, 0.001 y 0.1 , utilizando una expansión en una base de B-splines de $K = 200$ funciones básicas y tomando $L = D^2$. La curva correspondiente a dicha expansión en base ($\lambda = 0$), representada en gris, recoge con razonable fidelidad el comportamiento original de los datos. Merece especialmente la pena prestar atención a la curva asociada a $\lambda = 0.001$ (en rojo), pues adelantamos que este será el tipo de suavizado que aplicaremos a los datos de CGM en el Capítulo 4.

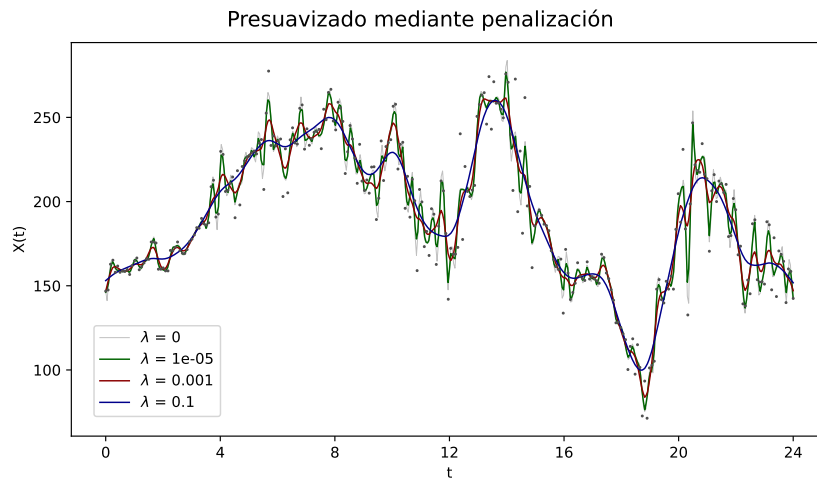


Figura 2.5: Gráfica de una única curva de CGM tras su presuavizado mediante penalización. Los puntos grises se corresponden con las mediciones discretas sin presuavizado.

Capítulo 3

Los modelos de regresión lineal funcional

En este capítulo nos adentraremos por fin en la materia que da título a este trabajo: la regresión lineal con datos funcionales. Para que un modelo de regresión sea considerado funcional, al menos una de las variables que intervienen, ya sea respuesta o explicativa, debe ser de carácter funcional. A continuación presentamos tres modelos diferenciados dependiendo de la cualidad funcional o escalar de las variables involucradas.

El primer caso, y la extensión directa del modelo lineal múltiple usual, es el modelo con respuesta escalar Y y covariable funcional $X(t)$:

$$Y = \alpha + \int \beta(s)X(s)ds + \varepsilon.$$

Por otro lado tenemos el modelo con respuesta funcional $Y(t)$ y covariable escalar X :

$$Y(t) = \alpha(t) + X\beta(t) + \varepsilon(t), \quad t \in [0, 1].$$

Y por último encontramos el modelo completamente funcional, en el que tanto respuesta como covariable son de carácter funcional:

$$Y(t) = \alpha(t) + \int \beta(t, s)X(s)ds + \varepsilon(t), \quad t \in [0, 1].$$

De estos tres modelos, nosotros estudiaremos únicamente el primero y el tercero, a los cuales nos referiremos como modelo lineal funcional (MLF) de respuesta escalar y funcional, respectivamente.

A lo largo de este capítulo dedicamos una sección a cada uno de estos modelos, en las que nos centraremos de manera casi exclusiva en la estimación de la función coeficiente. Expondremos

además un procedimiento para el contraste de significación adaptado a cada uno de los modelos. En la presentación de los contenidos, seguiremos la línea de los capítulos 4 y 5 de [14].

3.1. Modelo de regresión lineal funcional de respuesta escalar

Con la idea de presentar el MLF de respuesta escalar establecemos a continuación una analogía con los modelos lineales visitados en el Capítulo 1. El paso al modelo funcional consiste en traducir el índice vectorial discreto j del MLM (1.1) al argumento continuo s . Así, el vector de coeficientes del MLM se convierte en la llamada *función coeficiente* o *pendiente funcional* $\beta(s)$. Adaptando el resto de elementos a este contexto llegamos a la formulación del modelo lineal con covariable funcional y respuesta escalar:

$$Y = \alpha + \langle \beta, X \rangle + \varepsilon = \alpha + \int \beta(s)X(s)ds + \varepsilon,$$

donde el término de intercepto $\alpha \in \mathbb{R}$ y la función coeficiente $\beta \in L^2$ son parámetros desconocidos, y $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ es ruido blanco gaussiano independiente de X .

El término de intercepto α es útil, ya que aporta mayor generalidad al modelo y permite que las variables sean comparadas sin ninguna transformación previa. Sin embargo, con motivo de simplificar los cálculos y la notación, con lo que esperamos ganar claridad en la exposición de los contenidos, de ahora en adelante decidimos prescindir del mismo. Para ello, realizaremos una nueva simplificación suponiendo que ambas variables aleatorias, X e Y , son centradas. Esto implica:

$$0 = \mathbb{E}[Y] = \alpha + \mathbb{E}\langle \beta, X \rangle + \mathbb{E}[\varepsilon] = \alpha + \langle \beta, \mathbb{E}[X] \rangle = \alpha,$$

con lo que el valor del intercepto es necesariamente $\alpha = 0$.

Tras esta simplificación, la formulación del modelo lineal funcional de respuesta escalar con la que trabajaremos de ahora en adelante es la siguiente:

$$Y = \langle \beta, X \rangle + \varepsilon = \int \beta(s)X(s)ds + \varepsilon, \quad (3.1)$$

donde, de nuevo, $\beta \in L^2$ es la función coeficiente, ahora el único parámetro desconocido del modelo (además de la varianza del error), y $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ es ruido blanco gaussiano independiente de X .

De tratar con funciones aleatorias no centradas, y por lo tanto considerar el modelo con intercepto, este planteamiento es equivalente a sustituir las variables X e Y por $X - \bar{X}^1$ e $Y - \bar{Y}$, respectivamente, proceder a estimar la función coeficiente β con estas nuevas variables, y

¹En [11], Teorema 2.3, se establece que el orden de convergencia de $\mathbb{E}\|X - \bar{X}_n\|^2$ es de $O(n^{-1})$.

finalmente estimar el intercepto como $\hat{\alpha} = \bar{Y} - \langle \hat{\beta}, \bar{X} \rangle$. Con esto mantenemos la generalidad del modelo, por lo que de ahora en adelante procederemos bajo la hipótesis de que las variables X e Y son centradas.

Supondremos que partimos de una muestra (X_i, Y_i) , $i = 1, \dots, n$, donde las observaciones X_i e Y_i son independientes e idénticamente distribuidas a las variables aleatorias X e Y . Además, asumimos que las observaciones X_i de la covariable funcional han sido tratadas y convertidas en objetos funcionales, con lo que $X_i(t)$ es conocido en todo el intervalo.

3.1.1. Diferencias con respecto a la regresión euclidiana

El primer objetivo de la regresión es la estimación de los parámetros del modelo, en especial la función coeficiente β . Aunque el planteamiento del modelo es absolutamente análogo al de la regresión escalar, el marco funcional lleva consigo una nueva serie de dificultades propias, las cuales desglosamos brevemente a continuación.

La primera y mayor diferencia frente al caso euclídeo es que ahora la función coeficiente β vive en un espacio de dimensión infinita (L^2 o un subespacio del mismo). Esto hace que la cantidad finita de información que nos da la muestra no sea capaz de limitar dicho espacio sin aplicar restricciones adicionales. Consecuentemente, es perfectamente posible hallar una función $\hat{\beta}$ que ajuste sin error los valores muestrales observados, pero dicha estimación será demasiado variable e inestable debido a la influencia del ruido. Por lo tanto, un modelo sobreajustado de esta forma no generalizará bien a valores extraños a la muestra y tendrá poco valor predictivo y explicativo. Esto adelanta la necesidad de nuevas metodologías de estimación, que exploraremos en las secciones a continuación.

Otra dificultad propia del caso infinito-dimensional que se alza en cuanto a la estimación de β es el problema de la identificabilidad del modelo. Incluso si se considera una única covariable, la función coeficiente del modelo no es necesariamente única. Si $\beta_0 \in \ker C_X$, entonces se tiene $\text{Var} \langle \beta_0, X \rangle = \langle C_X(\beta_0), \beta_0 \rangle = 0$, luego, $\langle \beta_0, X \rangle$ es una variable aleatoria degenerada igual a $\langle \beta_0, \mathbb{E}[X] \rangle = 0$. Consecuentemente, $\beta + \beta_0$ es una función coeficiente que da lugar a un modelo equivalente para la respuesta Y . En resumen, si $\ker C_X \neq 0$ el modelo (3.1) no es identificable con $\beta \in L^2$. Para resolver esta cuestión podemos restringir la búsqueda de la función coeficiente β al subespacio $H = (\ker C_X)^\perp = \overline{\text{Im } C_X} \subset L^2$, en el cual su unicidad está garantizada. Una vez más, restringir el dominio funcional al subespacio H está justificado por la Observación 2.12.

Como última puntualización en cuanto a las diferencias entre el caso multivariante y el funcional, vamos a estudiar brevemente una identidad análoga a las conocidas ecuaciones normales de la regresión (Ecuación (1.4)). Antes de poder enunciarla, debemos definir el operador de covarianza cruzada, el cual no hemos mencionado hasta ahora. Lo definiremos en el entorno abstracto

de dos espacios de Hilbert generales, ya que su utilidad se extiende más allá del MLF de respuesta escalar.

Definición 3.1. Sean X e Y funciones aleatorias en sendos espacios de Hilbert H y W , dotados de productos interiores $\langle \cdot, \cdot \rangle_H$ y $\langle \cdot, \cdot \rangle_W$, respectivamente. Entonces, definimos el *operador de covarianza cruzada* de X e Y , el cual denotaremos por Δ , como el operador $\Delta \in \mathcal{LB}(H, W)$ dado por:

$$\Delta(h) = \mathbb{E}[\langle X - \mathbb{E}[X], h \rangle (Y - \mathbb{E}[Y])], \quad h \in H.$$

En el caso que ahora nos concierne en el que X vive en L^2 e Y en \mathbb{R} (y además ambas variables aleatorias son centradas), el operador Δ puede expresarse de la siguiente forma:

$$\Delta(h) = \mathbb{E}[\langle X, h \rangle Y] = \mathbb{E}[\langle YX, h \rangle] = \int \mathbb{E}[YX(s)] h(s) ds.$$

Esto es, $\Delta: L^2 \rightarrow \mathbb{R}$ es el funcional asociado a la función $\mathbb{E}[YX] \in L^2$, $\Delta = \langle \mathbb{E}[YX], \cdot \rangle$.

El principal interés del operador de covarianza cruzada en el contexto de la regresión es que, como adelantamos, da lugar a la versión funcional de las ecuaciones normales de la regresión.

Proposición 3.2. Sean X e Y elementos aleatorios centrados en sendos espacios de Hilbert H_1 y H_2 . Supongamos que estas funciones aleatorias se rigen bajo la relación $Y = \mathbf{B}(X) + \varepsilon$, donde $\mathbf{B} \in \mathcal{LB}(H_1, H_2)$ y ε es una función aleatoria en H_2 de media nula e independiente de X .

Entonces, si denotamos por $C = C_X$ al operador de covarianza de X y por Δ al operador de covarianza cruzada de X e Y se tiene la siguiente identidad:

$$\mathbf{B}C = \Delta. \tag{3.2}$$

Demostración. Es suficiente demostrar que $\langle \mathbf{B}C(h), g \rangle = \langle \Delta(h), g \rangle$, para $h \in H_1$ y $g \in H_2$ seleccionados arbitrariamente. Tenemos:

$$\begin{aligned} \langle \mathbf{B}C(h), g \rangle &= \langle \mathbf{B}(\mathbb{E}[\langle X, h \rangle X]), g \rangle \\ &= \mathbb{E}[\langle X, h \rangle \mathbf{B}(X), g] \\ &= \mathbb{E}[\langle X, h \rangle \langle Y - \varepsilon, g \rangle] \\ &= \mathbb{E}[\langle X, h \rangle \langle Y, g \rangle] - \mathbb{E}[\langle X, h \rangle \langle \varepsilon, g \rangle] \\ &= \langle \Delta(h), g \rangle. \end{aligned}$$

En la última igualdad hemos usado la independencia de X y ε , que a su vez implica la independencia de las variables aleatorias $\langle X, h \rangle$ y $\langle \varepsilon, g \rangle$. Estas son entonces incorreladas y de media nula, luego, su producto tiene también media nula. \square

Volviendo al MLF de respuesta escalar, podemos reescribir el modelo (3.1) en la formulación de la Proposición 3.2 sin más que considerar el funcional $\mathbf{B} = \langle \beta, \cdot \rangle$ asociado a β . Entonces, la Ecuación (3.2) se trata en efecto de la traducción natural de las ecuaciones normales de la regresión al lenguaje de operadores: la matriz de covarianzas $X'X$ se sustituye por el operador de covarianza C , la multiplicación por el vector β se convierte en composición con \mathbf{B} , y el vector $X'Y$ se traduce en el operador de covarianza cruzada Δ (que en este caso se trata de un funcional).

A la vista de la Ecuación (3.2), uno podría esperar resolver la cuestión de la estimación de β invirtiendo C y deduciendo de ello $\hat{\beta}$, de manera absolutamente análoga a la estimación por mínimos cuadrados en la regresión multivariante. Sin embargo, este enfoque no es viable por ciertos motivos.

El primero de ellos es que, como ya comentamos, puede ocurrir que $\ker C \neq 0$, con lo que el operador C no sería inyectivo, y por lo tanto no admitiría inversa de ninguna forma. Sin embargo, esta cuestión no se trata del verdadero problema. Como ya adelantamos, no podemos aspirar a estimar una función coeficiente $\beta \in L^2$ en la situación anterior, y lo que haríamos es restringirnos al subespacio $H = \overline{\text{Im } C}$. Ahora $C|_H$ sí tiene núcleo 0, y por lo tanto es biyectiva en su imagen. No obstante, no es invertible en el sentido de operadores. Como C es compacto (Teorema 2.11), no puede tener una inversa continua, sino que C^{-1} se trata de un operador lineal no acotado definido únicamente en un subespacio denso de H . No es posible extender C^{-1} continuamente a H y, en consecuencia, el problema de la inversa (3.2) no puede resolverse con el mismo enfoque que en el caso euclídeo.

El problema de la dimensión infinita tendrá una presencia constante de ahora en adelante. Un intento descuidado de esquivarlo es discretizar las observaciones funcionales. Como se explicó en la Sección 2.5, los datos suelen presentarse como puntos relativamente densos de una curva $X_i(t_j)$, $j = 1, \dots, J$. Así, una posibilidad sería considerar las J covariables escalares dadas por $X(t_j)$ y proceder con una regresión múltiple habitual. Sin embargo, rápidamente nos encontraríamos con un problema de alta dimensionalidad, ya que el número J de puntos de medición se encuentra con frecuencia en las centenas o en los miles. Además, si asumimos que las curvas observadas son suaves (lo cual es por lo general un supuesto básico) entonces se tiene $X(t_j) \approx X(t_{j+1})$, lo cual conduce a un caso grave de multicolinealidad.

Es evidente por las observaciones anteriores que se deben introducir nuevas técnicas que permitan tratar la regresión con este tipo de datos. A continuación, expondremos tres metodologías distintas para la estimación de la función coeficiente β , todas ellas con sus propias ventajas e inconvenientes.

3.1.2. Estimación por medio de una expansión en base

La primera forma de solventar el problema de la dimensionalidad infinita es restringir el espacio en el que vive β . Para ello se selecciona una base $\{\phi_k\}_{k \in \mathbb{N}}$ (habitualmente B-splines o tal vez la misma base usada para la expresión de las X_i) y un índice de corte K , y se asume que la función β puede aproximarse suficientemente bien por las K primeras funciones básicas. Entonces escribimos informalmente la igualdad:

$$\beta = \sum_{k=1}^K b_k \phi_k. \quad (3.3)$$

Con esto, el problema de estimación de la función β se reduce a la estimación de los finitos coeficientes b_k .

Sustituyendo la nueva expresión de β en la ecuación del modelo (3.1) obtenemos lo siguiente:

$$\langle \beta, X_i \rangle = \sum_{k=1}^K b_k \langle \phi_k, X_i \rangle = \sum_{k=1}^K b_k x_{ik},$$

donde hemos denotado $x_{ik} = \langle X_i, \phi_k \rangle = \int X_i(s) \phi_k(s) ds$.

Finalmente, el modelo funcional (3.1) se reduce al MLM dado por $Y_i = \sum_{k=1}^K b_k x_{ik} + \varepsilon_i$, con K covariables $\langle X, \phi_k \rangle$. Esta técnica de limitar los objetos funcionales a subespacios de dimensión finita, para posteriormente reducir el modelo que nos interesa a un modelo sobre coeficientes, será empleada reiteradamente a lo largo de este trabajo. Hacer esto es de hecho una necesidad, ya que la muestra nos aporta únicamente una cantidad finita de información, insuficiente para atar estos objetos infinito-dimensionales, por lo que todos los métodos que estudiaremos son necesariamente aproximados.

La estimación del vector de coeficientes $\mathbf{b} = (b_1, \dots, b_K)'$ se resuelve haciendo uso de la fórmula conocida del estimador de mínimos cuadrados:

$$\hat{\mathbf{b}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y},$$

donde \mathbf{X} denota la matriz de diseño $\mathbf{X} = (x_{ik}) \in \mathcal{M}_{n \times K}$ e $\mathbf{Y} = (Y_1, \dots, Y_n)'$ es el vector que contiene a las observaciones Y_i . Con esto se obtiene el siguiente estimador de β :

$$\hat{\beta}(s) = \sum_{k=1}^K \hat{b}_k \phi_k(s).$$

Este método de estimación es simple, pero en ocasiones no resulta especialmente adecuado. Primero, debemos recordar que estamos aproximando la función β por su proyección en el espacio $\{\{\phi_k\}_{k=1}^K\}$, con lo cual estamos introduciendo un sesgo. Para asegurar la consistencia de

este estimador, es necesario que K tienda a infinito junto con n , además de garantizar ciertas condiciones de convergencia de los b_k .

De todas formas, esta cuestión será común a la mayoría de estimadores que estudiaremos. El gran inconveniente viene de la elección de la base y del índice de corte, junto con la gran dependencia del estimador obtenido de dichas elecciones. Dos bases diferentes pueden dar lugar a estimadores de β con formas distintas, lo cual reduce la interpretabilidad. Tal y como ocurría con la metodología de suavizado de datos mediante una expansión en base, K se convierte en un parámetro que regula la suavidad de la función estimada: si K es demasiado grande se conseguirá un ajuste con residuos despreciables, pero con mucha variabilidad y muy sensible al error; mientras que si K se toma más pequeño el sesgo introducido es mayor.

3.1.3. Estimación mediante penalización

Una forma de resolver algunos de los inconvenientes mencionados anteriormente es la estimación penalizada, que sigue la misma filosofía que el método de presuavizado mediante penalización presentado en la Sección 2.5.2, y se trata de una especie de adaptación de la regresión Ridge en el contexto funcional. Se asume que la función coeficiente β cumple ciertas condiciones de regularidad² y se selecciona un operador diferencial lineal L (habitualmente $L = D^2$) con el que se cuantificará el nivel de suavizado de β a través de la penalización $R_L(\beta) = \int [L(\beta)(s)]^2 ds$.

De nuevo, el punto de partida es una expansión en base de la forma (3.3), pero esta vez tomando K lo suficientemente grande como para que el estimador $\hat{\beta}$ deje de ser sensible a este. A continuación, se buscan la función $\hat{\beta}$ que minimiza la siguiente función objetivo:

$$PSSR(\beta) = \sum_{i=1}^n \left(Y_i - \int \beta(s) X_i(s) ds \right)^2 + \lambda R_L(\beta), \quad (3.4)$$

donde $\lambda \geq 0$ es un parámetro de suavizado.

La selección del valor de λ no es una tarea trivial, pero antes de dar ningún método o criterio con este fin expondremos cómo computar el estimador $\hat{\beta}$ para un cierto λ fijado.

Expandiendo β podemos reescribir la función coste de la siguiente forma:

$$PSSR(b_1, \dots, b_K) = \sum_{i=1}^n \left(Y_i - \sum_{k=1}^K b_k x_{ik} \right)^2 + \sum_{k,l=1}^K b_k b_l R_{kl},$$

donde hemos vuelto a introducir la notación $x_{ik} = \langle X_i, \phi_k \rangle$ y denotado $R_{kl} = \int L(\phi_k)(s) L(\phi_l)(s) ds$.

²Formalmente, se asume que β vive en un espacio de Sobolev, y este se convierte en el espacio funcional con el que se trabaja. El espacio de Sobolev W^q es (como conjunto) el subespacio de L^2 dado por $W^q = \{f \in L^2 / D^m f \in L^2, \forall m \leq q\}$, con q el grado mayor de regularidad asumido para β .

Observación 3.3. De considerar el caso especial con $L = D^0 = \text{id}$, el segundo término de $PSSR$ penaliza la norma cuadrática de β , con lo que se convierte en una generalización funcional del criterio de optimización de la regresión Ridge.

A partir de esta expresión, un desarrollo que omitimos conduce a la siguiente fórmula para el estimador $\hat{\mathbf{b}}$ de $\mathbf{b} = (b_1, \dots, b_K)'$:

$$\hat{\mathbf{b}} = (\mathbf{X}'\mathbf{X} + \lambda R)^{-1} \mathbf{X}'\mathbf{Y},$$

donde $\mathbf{X} = (x_{ik}) \in \mathcal{M}_{n \times K}$ e \mathbf{Y} son la matriz de diseño y el vector de observaciones de Y definidos anteriormente y $R \in \mathcal{M}_{K \times K}$ es la matriz cuadrada con entradas R_{kl} , $k, l = 1, \dots, n$.

La ventaja de esta metodología de estimación frente a la estimación mediante expansión en base reside en el control más agudo del grado de suavizado de $\hat{\beta}$. Además, este estimador tiene buenas propiedades asintóticas que han sido estudiadas extensivamente.

Sin embargo, la selección del parámetro de suavizado λ no deja de ser una cuestión que debe ser atendida. Una posibilidad realista en la práctica es jugar con el valor de λ y escoger uno basándose en intuición visual acerca de la forma del estimador. De todas formas, existen múltiples métodos que permiten buscar un valor óptimo de λ de una forma más cuantitativa. Uno de ellos, y tal vez el más popular, es la validación cruzada (CV, por sus siglas en inglés *cross-validation*). Este consiste en evaluar los distintos valores de λ midiendo el error cuadrático medio cometido al predecir cada valor Y_i de la muestra cuando se calcula el estimador de β con todas salvo la i -ésima observaciones. Una alternativa menos costosa computacionalmente a la CV es la GCV (*generalized cross-validation*). Este criterio pretende aproximar el valor de la CV con un cálculo directo, en lugar de reiterar el proceso de ajustar el modelo n veces.

3.1.4. Estimación basada en componentes principales funcionales

La última metodología de estimación que consideraremos será aquella basada en las componentes principales funcionales de la covariable X . Este es, junto con la estimación mediante penalización expuesta en la sección anterior, el método de estimación más estudiado y extendido. Esto es natural, pues las FPC tienen propiedades excelentes que las vuelven herramientas poderosas en este contexto.

En la Sección 3.1.2 vimos cómo estimar de manera sencilla β valiéndonos de su expansión en una base dada. Una nueva posibilidad de estimación es considerar un procedimiento idéntico a este, pero empleando ahora como base la formada por las EFPC $\hat{\psi}_k$. La principal diferencia, por lo tanto, es que la base usada para la expansión de β es derivada de la muestra, en lugar de haber sido fijada *a priori*.

A continuación presentamos una forma alternativa, que consideramos más interesante, de llegar a un estimador fundamentado también en el análisis de componentes principales funcionales. El desarrollo es en esencia idéntico al presentado en [7].

Consideremos la ecuación del MLF de respuesta escalar (3.1) y expresemos cada elemento que interviene en función de las FPC ψ_k . Como $\{\psi_k\}_{k \in \mathbb{N}}$ es una base de Hilbert de $H = \overline{\text{Im } C}$, $\beta \in H$ admite una expansión $\beta = \sum_{k=1}^{\infty} b_k \psi_k$ con coeficientes b_k determinados unívocamente. Sustituyendo también X por su expansión de Karhunen-Loève y aplicando la ortonormalidad de las FPC obtenemos:

$$Y = \langle \beta, X \rangle + \varepsilon = \sum_{k,l=1}^{\infty} b_k \xi_l \langle \psi_k, \psi_l \rangle + \varepsilon = \sum_{k=1}^{\infty} b_k \xi_k + \varepsilon.$$

A continuación deducimos la fórmula que deben seguir los coeficientes b_k . Para ello haremos uso de las propiedades conocidas de los *scores*. Multiplicando la expresión anterior por ξ_k y tomando la media obtenemos lo siguiente:

$$\mathbb{E}[\xi_k Y] = \mathbb{E} \left[\sum_{l=1}^{\infty} b_l \xi_k \xi_l + \xi_k \varepsilon \right] = \sum_{l=1}^{\infty} b_l \mathbb{E}[\xi_k \xi_l] + \mathbb{E}[\xi_k \varepsilon].$$

Ahora bien, sabemos que $\mathbb{E}[\xi_k] = 0$ y que $\mathbb{E}[\xi_k \xi_l] = 0$ si $k \neq l$. Además, por la independencia de X y ε , y puesto que los ξ_k dependen únicamente de X , tenemos que ξ_k y ε son independientes, luego $\mathbb{E}[\xi_k \varepsilon] = 0$. Esto nos conduce a la siguiente expresión final de b_k :

$$b_k = \frac{\mathbb{E}[\xi_k Y]}{\mathbb{E}[\xi_k^2]} = \frac{\text{Cov}(Y, \xi_k)}{\lambda_k}. \quad (3.5)$$

Con esto se obtiene la expresión exacta de β en la base de FPC. Si en lugar de tomar la suma infinita esta se trunca por un índice de corte p , y se sustituyen todos los elementos que intervienen por su versión muestral, obtenemos el siguiente estimador de β :

$$\hat{\beta} = \sum_{k=1}^p \hat{b}_k \hat{\psi}_k, \quad \hat{b}_k = \frac{\hat{\sigma}_{Y,k}}{\hat{\lambda}_k}, \quad (3.6)$$

donde $\hat{\sigma}_{Y,k}$ es un estimador de $\text{Cov}(Y, \xi_k)$ apropiado, por ejemplo $\hat{\sigma}_{Y,k} = n^{-1} \sum_{i=1}^n Y_i \hat{\xi}_{ik}$.

Observación 3.4. Es posible llegar a la expresión (3.5) de los coeficientes b_k de forma más directa a partir de la identidad (3.2). Para cada $k \in \mathbb{N}$ se tiene $b_k = \langle \beta, \psi_k \rangle = \lambda_k^{-1} \Delta(\psi_k)$, y es inmediato comprobar que $\Delta(\psi_k) = \text{Cov}(Y, \xi_k)$. El estimador $\hat{\sigma}_{Y,k}$ que hemos seleccionado para $\text{Cov}(Y, \xi_k)$ es precisamente $\hat{\sigma}_{Y,k} = \Delta_n(\hat{\psi}_k)$, donde $\Delta_n = n^{-1} \sum_{i=1}^n Y_i \langle X_i, \cdot \rangle$ es el estimador empírico del operador de covarianza cruzada Δ .

Este segundo enfoque nos permite verificar que el estimador $\hat{\beta}$ presentado en esta sección coincide con el propuesto en [2]. En dicho artículo, los autores escogen trabajar en términos

de operadores y por ello no proporcionan una expresión explícita del estimador de la función coeficiente. No obstante, una sencilla comprobación directa confirma que ambos métodos son equivalentes.

En el artículo citado se establece la consistencia de este estimador bajo ciertas hipótesis de convergencia sobre los autovalores λ_k . Para ello, el índice de corte $p = p(n)$ debe tender a infinito con el tamaño muestral, y hacerlo de manera suficientemente lenta.

3.1.5. Contraste de significación

En esta sección final presentamos un procedimiento para el contraste de significación de la función coeficiente β . El estadístico que emplearemos es el introducido en [3], el cual está basado en el operador de covarianza cruzada. El contraste de significación es un test fundamental en la teoría de la regresión, ya que permite verificar si la variable X tiene verdaderamente efecto (lineal) sobre la variable Y .

Aclaremos que el propósito de este contraste no es evaluar la hipótesis de linealidad del modelo. Para ello debería aplicarse un contraste de bondad de ajuste (véase, por ejemplo, [4]). Es también importante notar que incluso si tras el contraste se concluye que β es no significativa, no podemos afirmar con seguridad que X no ejerza ningún efecto sobre Y , pues podría ocurrir que la relación entre ambas variables sea de carácter no lineal.

En resumen, en el planteamiento del contraste de significación se asume la validez del modelo lineal y se pretende contrastar la hipótesis nula

$$H_0 : \beta = 0$$

frente a

$$H_1 : \beta \neq 0.$$

Antes de comenzar, debemos hacer un comentario importante acerca del planteamiento del contraste de significación. Con este contraste se pretende comprobar si la variable X tiene un efecto no nulo sobre la variable Y . En el caso euclídeo esto es equivalente a la no nulidad del vector de coeficientes (salvo que el vector aleatorio X fuese degenerado). Sin embargo, en el contexto infinito-dimensional en el que estamos trabajando esto no es exactamente cierto.

De nuevo, esta diferencia se debe a la misma cuestión que causaba la no identificabilidad del MLF, tal y como mencionamos en la Sección 3.1.1. Insistimos una vez más en este problema.

Si suponemos que $\beta = \beta_0 \neq 0$ es un elemento de $\ker C$, donde $C = C_X$ denota el operador de covarianza de X , entonces se tiene $\mathbb{E} \langle \beta_0, X \rangle^2 = \langle C(\beta_0), \beta_0 \rangle = 0$. Como $\langle \beta_0, X \rangle$ es 0 en media

cuadrática, tenemos $\langle \beta_0, X \rangle = 0$ c.s. Concluimos que X no tiene efecto (lineal) sobre Y , a pesar de que $\beta \neq 0$.

Esta cuestión exige que se reformule el contraste de significación. Como antes para hacer que el modelo sea identificable, restringiremos el espacio en el que vive β . Para ello, notemos que por la Observación 2.12 podemos expresar β como $\beta = \beta_0 + \beta_1$, con $\beta_0 \in \ker C$ y $\beta_1 \in \overline{\text{Im } C}$, con lo que se tiene $\langle \beta, X \rangle = \langle \beta_1, X \rangle$. De esta forma, el contraste de significación equivale a contrastar $\beta_1 = 0$. Por lo tanto, el contraste debe llevarse a cabo restringiéndose al espacio $H = \overline{\text{Im } C}$. Equivalentemente, de ahora en adelante supondremos sin pérdida de generalidad que el núcleo de C es trivial, $\ker C = 0$.

En el caso en el que el espacio H es de dimensión finita, la función aleatoria X puede ser identificada con un vector aleatorio, con lo que el contraste de significación puede abordarse con técnicas de la regresión multivariante, como el F-test. Por lo tanto, de ahora en adelante limitaremos el estudio al caso en el que H es un espacio de Hilbert de dimensión infinita.

Adicionalmente, trabajaremos bajo las siguientes hipótesis: se tiene $\mathbb{E} \|X\|^4 < \infty$, y los autovalores λ_k de C son distintos (y positivos): $\lambda_1 > \lambda_2 > \dots > 0$.

Volvamos entonces al planteamiento del procedimiento de contraste. El interés del operador de covarianza cruzada en este contexto es naturalmente la Ecuación (3.2). La clave de esta identidad es que nos permite reenmarcar el contraste de la hipótesis nula $\beta = \beta_1 = 0$ como el contraste de $H_0: \Delta = 0$ frente a $H_1: \Delta \neq 0$ (donde recordamos que estamos tomando $H = \overline{\text{Im } C}$ como dominio).

Es muy sencillo verificar la equivalencia. Si $\beta = 0$ entonces $\mathbf{B} = 0$ y, por (3.2), $\Delta = 0$. Recíprocamente, si $\Delta = \mathbf{B}C = 0$, y ya que un operador es igual a 0 si y solo si lo es su adjunto, $(\mathbf{B}C)^* = C^* \mathbf{B}^* = C \mathbf{B}^* = 0$. Ahora, por la hipótesis $\ker C = 0$ (o bien restringiéndonos al espacio H), tenemos que la composición se anula si y solo si lo hace \mathbf{B}^* . Consecuentemente, llegamos a $\mathbf{B} = 0$.

Ahora podemos comenzar a desarrollar el estadístico de contraste. El punto de partida es el estimador empírico natural del operador de covarianza cruzada:

$$\Delta_n(h) = \frac{1}{n} \sum_{i=1}^n \langle X_i, h \rangle Y_i, \quad h \in H.$$

Recordamos que Δ y Δ_n son funcionales, $\Delta, \Delta_n \in H^*$, y por el teorema de representación de Riesz son caracterizados por respectivas funciones $g, g_n \in H$, siendo $g = \mathbb{E}[YX]$ y g_n la función aleatoria dada por $g_n = n^{-1} \sum_{i=1}^n Y_i X_i = \overline{(YX)_n}$. Estudiar el comportamiento del operador aleatorio Δ_n resulta equivalente a estudiar la función aleatoria g_n .

Por el teorema central del límite y bajo la hipótesis nula $H_0: \Delta = 0 \iff g = 0$, tenemos

que $\sqrt{n}g_n$ converge en distribución a una función aleatoria gaussiana G con media 0 y operador de covarianza $\sigma^2 C$, donde σ^2 es la varianza del error. Usaremos esto para derivar un estadístico con una distribución asintótica χ^2 conocida.

La expansión de Karhunen-Loève de G es especialmente sencilla por ser una función aleatoria gaussiana. Sean (λ_k, ψ_k) , $k \in \mathbb{N}$, los autovalores y autofunciones de C (y de $\sigma^2 C$). Entonces, por el Ejemplo 2.20 podemos escribir

$$G = \sigma \sum_{k=1}^{\infty} \sqrt{\lambda_k} Z_k \psi_k,$$

donde las Z_k , $k \in \mathbb{N}$, son variables aleatorias normales estándar independientes. De esta expresión se sigue que $\|G\|^2 = \sigma^2 \sum_{k=1}^{\infty} \lambda_k Z_k^2$.

La expansión de Karhunen-Loève de G , y en especial la fórmula de su norma cuadrática, motiva la manipulación de la misma para obtener un estadístico con distribución límite χ^2 . Para ello introducimos el operador $A_n \in \mathcal{LB}(H)$ definido por la matriz infinita

$$\text{diag} \left(1/\sqrt{\lambda_1}, 1/\sqrt{\lambda_2}, \dots, 1/\sqrt{\lambda_{p_n}}, 0, \dots \right)$$

en la base $\{\psi_k\}_{k \in \mathbb{N}}$, esto es, $A_n = \sum_{k=1}^{p_n} \lambda_k^{-1/2} \langle \cdot, \psi_k \rangle \psi_k$; donde $p_n \in \mathbb{N}$ es una sucesión de naturales $p_n < n$ que tiende a infinito. Es inmediato probar que $\sigma^{-2} \|A_n(G)\|^2$ sigue una distribución χ^2 de p_n grados de libertad.

Tras este desarrollo llegamos finalmente a la expresión del estadístico de contraste D_n :

$$D_n = \frac{n}{\hat{\sigma}^2} \left\| \hat{A}_n(g_n) \right\|^2, \quad (3.7)$$

donde $\hat{\sigma}^2$ es un estimador³ de σ^2 y \hat{A}_n es la versión muestral del operador A_n , que viene dada por $\hat{A}_n = \sum_{k=1}^{p_n} \hat{\lambda}_k^{-1/2} \langle \cdot, \hat{\psi}_k \rangle \hat{\psi}_k$. La definición del operador \hat{A}_n muestra la necesidad de tomar $p_n < n$, ya que recordemos que a partir de una muestra de n observaciones es posible calcular a lo sumo las n primeras EFPC.

El estadístico D_n sigue aproximadamente, para valores de n lo suficientemente grandes, una distribución $\chi_{p_n}^2$. Otro posible estadístico de contraste es el siguiente:

$$T_n = \frac{1}{\sqrt{2p_n}} \left(\frac{n}{\hat{\sigma}^2} \left\| \hat{A}_n(g_n) \right\|^2 - p_n \right), \quad (3.8)$$

el cual (bajo condiciones apropiadas sobre $\hat{\sigma}^2$ y la sucesión de p_n) converge en distribución a una normal estándar.

³Es natural tomar un estimador de σ^2 basado en la $SSR = \|\mathbf{Y} - \hat{\mathbf{Y}}\|_2^2$, donde $\hat{\mathbf{Y}}$ denota el vector de valores ajustados. En la regresión escalar, esta es habitualmente normalizada por el factor $(n - \text{df})^{-1}$. En el MLF no es trivial definir los grados de libertad del modelo, pero se suele emplear como aproximación la traza de la matriz *hat* S , siendo esta la matriz que da $\hat{\mathbf{Y}} = S\mathbf{Y}$. Entonces, el estimador usual de σ^2 es $\hat{\sigma}^2 = (n - \text{tr } S)^{-1} SSR$.

Para cerrar la sección, resumimos el procedimiento del test basado en los dos estadísticos descritos. Primero debe seleccionarse el índice de truncamiento p_n y el estadístico con el que se efectuará el contraste. Para ello puede consultarse el estudio llevado a cabo en [3]. Se fija además un nivel nominal α , habitualmente $\alpha = 0.05$ o 0.01 . Entonces, se calcula el estadístico D_n , y con él T_n , a partir de la muestra. De trabajar con D_n , su valor se compara con el valor crítico $\chi_{p_n, 1-\alpha}^2$, siendo este el cuantil $(1 - \alpha)$ de la distribución $\chi_{p_n}^2$. La hipótesis nula $H_0 : \beta = 0$ será rechazada si $D_n > \chi_{p_n, 1-\alpha}^2$, y no lo será en caso contrario. Si en su lugar se escoge emplear T_n , se rechazará la hipótesis nula si y solo si $|T_n| > z_{1-\alpha/2}$, siendo $z_{1-\alpha/2}$ el cuantil de la distribución normal estándar que deja a su derecha una probabilidad de $\alpha/2$.

3.2. Modelo de regresión lineal funcional de respuesta funcional

A lo largo de la sección anterior exploramos el modelo lineal funcional con respuesta escalar, lo que nos permitió conocer las diferencias y dificultades propias del FDA en comparación con la estadística multivariante. En esta sección elevaremos ligeramente el grado de complejidad introduciendo además el carácter funcional de la respuesta Y . Estudiaremos el MLF de respuesta funcional, el cual se distancia más del modelo lineal múltiple y se convierte en terreno relativamente desconocido. Veremos, sin embargo, que muchas de las técnicas desarrolladas para el caso escalar son aplicables, con ligeras diferencias, en el caso funcional.

Ahora, la dependencia lineal entre las variables respuesta y explicativa vendrá modelada por un operador lineal \mathbf{B} , esto es, la función de regresión es $m(f) = \mathbb{E}[Y | X = f] = \mathbf{B}(f)$. Adicionalmente supondremos que el operador \mathbf{B} pertenece a la familia de operadores Hilbert-Schmidt. En este caso, el operador admite una representación integral (Proposición A.14) mediante un núcleo $\beta \in L^2([0, 1]^2)$, esto es, podemos describir \mathbf{B} como:

$$h \in L^2 \mapsto \mathbf{B}(h)(t) = \int \beta(t, s)h(s)ds, \quad t \in [0, 1].$$

Llamaremos *núcleo* o *función coeficiente* a $\beta(t, s)$.

Con esto estamos en condiciones de presentar el modelo. El MLF de respuesta funcional puede plantearse como $Y = \alpha + \mathbf{B}(X) + \varepsilon$, donde esta vez el término de intercepto se trata de una función $\alpha(t)$ de L^2 y el error ε es, similarmente, una función aleatoria de media nula. Una vez más, el intercepto resulta incómodo a la hora de plantear los desarrollos, y lo es aún más en este caso en el que se trata de un objeto funcional. Por lo tanto, con el fin de mantener la claridad escogemos omitir el término de intercepto, asumiendo para ello la misma hipótesis de centralidad de X e Y con la que ya habíamos trabajado en el MLF de respuesta escalar. La condición $\mathbb{E}[X] = \mathbb{E}[Y] = 0$ garantiza, por un argumento absolutamente análogo al de la Sección 3.1, que el intercepto es fijado con valor $\alpha(t) = 0$ para todo $t \in [0, 1]$.

Entonces, bajo el supuesto de que tanto la respuesta como la covariable son centradas, el MLF de respuesta funcional se formula de la siguiente manera:

$$Y(t) = \int \beta(t, s)X(s)ds + \varepsilon(t), \quad t \in [0, 1], \quad (3.9)$$

donde ε es una función aleatoria (cuadrado-integrable) gaussiana de media nula e independiente de X .

De tener funciones aleatorias no centradas y trabajar con el modelo con intercepto, puede resolverse la estimación del mismo de la siguiente forma. Se reduce el modelo al caso centrado con la sustitución de X e Y por $X - \bar{X}$ e $Y - \bar{Y}$, respectivamente, se estima en este caso la función coeficiente β y finalmente se obtiene el estimador del intercepto como $\hat{\alpha}(t) = \bar{Y}(t) - \int \hat{\beta}(t, s)\bar{X}(s)ds$. Justificamos así que la hipótesis de centralidad no resulta en una pérdida de generalidad, y en adelante procedemos suponiendo $\mathbb{E}[X] = \mathbb{E}[Y] = 0$.

Nuestro primer objetivo es, una vez más, la estimación de la función coeficiente β a partir de una muestra aleatoria simple (X_i, Y_i) , $i = 1, \dots, n$, en la que los objetos funcionales han sido tratados y son totalmente observables.

La estimación de la función coeficiente β es tan solo ligeramente mayor en complejidad que la de su contraparte en el MLF de respuesta escalar (3.1), y al igual que en este caso, todas las metodologías de estimación pasan por considerar una expansión en base de la misma. Por lo tanto, merece la pena detenernos un instante para resumir y explicar las diferencias debidas al hecho de que β se trata ahora de una función bivariente.

Para ello, dejemos momentáneamente a un lado las ideas que conciernen a la función coeficiente, y supongamos que escogemos representar la relación lineal entre X e Y mediante un operador más general $\mathbf{B} \in \mathcal{LB}(L^2)$. Al tratar de pensar en cómo dar una descomposición de \mathbf{B} se hace evidente la necesidad (o conveniencia) de restringir nuestro enfoque al espacio de operadores Hilbert-Schmidt. Recordamos que $\mathcal{LB}(L^2)$ se trata de un espacio de Banach, y por lo tanto no tenemos garantía de que este espacio admita una base⁴, lo cual nos impide trabajar con una representación sencilla de un operador general. En cambio, los operadores Hilbert-Schmidt sobre L^2 forman un espacio de Hilbert (separable), por lo que la teoría expuesta en la sección anterior (en la que nuestro operador $\mathbf{B} = \langle \beta, \cdot \rangle$ pertenecía al espacio de Hilbert $(L^2)^*$) se adapta de forma natural a este nuevo espacio de operadores.

De hecho, es absolutamente trivial obtener una base de $\mathcal{LB}_{HS}(L^2)$ a partir de dos (o una) bases de L^2 . Sean $\{\phi_j\}_{j \in \mathbb{N}}$ y $\{\psi_k\}_{k \in \mathbb{N}}$ bases ortonormales de L^2 y, para $j, k \in \mathbb{N}$, consideremos

⁴La principal noción de base en espacios de Banach es aquella de base de Schauder. No todo espacio de Banach admite una base de Schauder, y estas no tienen por qué ser necesariamente numerables. De hecho, en el caso de $\mathcal{LB}(L^2)$ nunca lo serán pues este espacio es no separable.

el *producto tensor* de ϕ_j y ψ_k , que se trata del operador $(\phi_j \otimes \psi_k) \in \mathcal{LB}_{HS}(L^2)$ dado por

$$(\phi_j \otimes \psi_k)(h) := \langle \psi_k, h \rangle \phi_j.$$

Es sencillo probar (véase [12]) que $\{\phi_j \otimes \psi_k\}_{j,k \in \mathbb{N}}$ forma una base de Hilbert del espacio $\mathcal{LB}_{HS}(L^2)$.

Volvamos ahora a adoptar la perspectiva en términos de núcleos en lugar de operadores. Como adelantamos, todo operador $\mathbf{B} \in \mathcal{LB}_{HS}(L^2([0, 1]))$ es representado biunívocamente por un núcleo $\beta \in L^2([0, 1]^2)$. Si realizamos la identificación correspondiente para la base tensor anterior llegamos a una *base producto* $\{\phi_j \times \psi_k\}_{j,k \in \mathbb{N}}$ de $L^2([0, 1]^2)$, siendo estas funciones las dadas por

$$(\phi_j \times \psi_k)(t, s) := \phi_j(t)\psi_k(s), \quad t, s \in [0, 1].$$

Con esto, podemos considerar la expansión de β en la base producto:

$$\beta = \sum_{j,k=1}^{\infty} b_{jk}(\phi_j \times \psi_k). \quad (3.10)$$

La expansión (3.10) es el punto de partida de las dos metodologías de estimación que presentamos en las dos secciones a continuación. La primera de ellas es una adaptación natural de la estimación basada en componentes principales funcionales expuesta en la Sección 3.1.4. La segunda es un enfoque del planteamiento de la penalización diferente al empleado hasta ahora y más cercano a la teoría de la regresión multivariante.

3.2.1. Estimación basada en componentes principales funcionales

La estimación basada en componentes principales funcionales de la función coeficiente β es esencialmente análoga a la vista en el caso del MLF de respuesta escalar. Consideremos las bases de FPC $\{\phi_j\}_{j \in \mathbb{N}}$ y $\{\psi_k\}_{k \in \mathbb{N}}$ de Y y X , respectivamente, y la expansión (3.10) en la correspondiente base producto:

$$\beta(t, s) = \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} b_{jk} \phi_j(t) \psi_k(s), \quad t, s \in [0, 1]. \quad (3.11)$$

Observación 3.5. Por la Observación 2.12 sabemos que $X \in \overline{\text{Im } C_X}$ c.s. e $Y \in \overline{\text{Im } C_Y}$ c.s. Por lo tanto, para modelar la relación entre ambas variables aleatorias es suficiente limitarnos a considerar $\mathbf{B} \in \mathcal{LB}_{HS}(\overline{\text{Im } C_X}, \overline{\text{Im } C_Y})$, con lo que la expansión de β en las bases de FPC está justificada.

A continuación exponemos un desarrollo similar al abordado en la Sección 3.1.4, que nos permitirá expresar los coeficientes de β en términos de los *scores* de X e Y , y deducir de dicha expresión un estimador natural para β . Comenzamos considerando las expansiones de Karhunen-Loève de Y y X :

$$Y = \sum_{j=1}^{\infty} \zeta_j \phi_j, \quad X = \sum_{k=1}^{\infty} \xi_k \psi_k, \quad (3.12)$$

en las que denotamos por $\zeta_j = \langle Y, \phi_j \rangle$ y $\xi_k = \langle X, \psi_k \rangle$ los respectivos *scores* de Y y X (recordamos que trabajamos bajo el supuesto de que ambas funciones aleatorias son centradas).

Haciendo uso de las expansiones (3.11) y (3.12), podemos reescribir la expresión (3.9) del MLF de respuesta funcional como sigue:

$$\sum_{j=1}^{\infty} \zeta_j \phi_j(t) = \sum_{j', k'=1}^{\infty} b_{j'k'} \sum_{k=1}^{\infty} \xi_k \langle \psi_{k'}, \psi_k \rangle \phi_{j'}(t) + \sum_{j=1}^{\infty} e_j \phi_j(t),$$

donde escribimos $e_j = \langle \varepsilon, \phi_j \rangle$, y con ello $\varepsilon = \sum_{j=1}^{\infty} e_j \phi_j$.

Aplicando la ortonormalidad de las FPC ψ_k y notando que, por ser $\{\phi_j\}_{j \in \mathbb{N}}$ una base de Hilbert, los coeficientes de cada ϕ_j a ambos lados de la ecuación deben coincidir, llegamos a

$$\zeta_j = \sum_{k'=1}^{\infty} b_{jk'} \xi_{k'} + e_j, \quad j \in \mathbb{N}.$$

Por último, multiplicamos ambos lados de la ecuación y tomamos medias para llegar, aplicando una vez más las propiedades de los *scores* junto con la independencia de X y ε , a la última identidad:

$$\mathbb{E}[\zeta_j \xi_k] = b_{jk} \mathbb{E}[\xi_k^2] + \mathbb{E}[e_j \xi_k] = b_{jk} \lambda_k, \quad j, k \in \mathbb{N},$$

en la que λ_k es el k -ésimo autovalor de C_X (asociado a ψ_k).

De la expresión anterior se deduce la fórmula poblacional de los coeficientes b_{jk} :

$$b_{jk} = \frac{\text{Cov}(\zeta_j, \xi_k)}{\lambda_k}, \quad j, k \in \mathbb{N}. \quad (3.13)$$

Esta última fórmula es traducida al contexto muestral para dar con el siguiente estimador de β :

$$\hat{\beta}(t, s) = \sum_{j=1}^q \sum_{k=1}^p \hat{b}_{jk} \hat{\phi}_j(t) \hat{\psi}_k(s), \quad \hat{b}_{jk} = \frac{\hat{\sigma}_{jk}}{\hat{\lambda}_k}, \quad (3.14)$$

siendo $\hat{\sigma}_{jk}$ un estimador apropiado de $\text{Cov}(\zeta_j, \xi_k)$, por ejemplo $\hat{\sigma}_{jk} = n^{-1} \sum_{i=1}^n \hat{\zeta}_j \hat{\xi}_k$.

3.2.2. Estimación mediante penalización

En esta sección proponemos una metodología alternativa para estimar la función coeficiente $\beta(t, s)$, basada fundamentalmente en el concepto de penalización. No obstante, como adelantamos, este enfoque difiere del adoptado para el estimador de $\beta(s)$ en la Sección 3.1.3. Recordamos que, en el caso del estimador mencionado, la penalización es aplicada para controlar la regularidad de la función de manera explícita, regularizando por ejemplo su curvatura $\beta^{(2)}(s)$, con lo que se pretende reducir la alta variación y los picos de la función.

En el caso que ahora nos concierne de la función bivalente $\beta(t, s)$ no es inmediato llegar a un sistema de penalización análogo, pues la elección del operador diferencial lineal L a penalizar no es trivial. Un posible ejemplo, y tal vez el más natural, es el Laplaciano $L = \Delta = \frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial s^2}$. Es posible que esta elección dé lugar a resultados aceptables, pero es fundamentalmente ineficaz para penalizar funciones armónicas, las cuales pueden presentar un alto grado de variación. Por lo tanto, una medida de la forma $R_\Delta(\beta) = \iint (\Delta\beta)^2$ con valor cercano a 0 no refleja necesariamente la suavidad de la función β .

Otra posibilidad dentro del mismo enfoque sería considerar dos términos de penalización, cada uno regulando la suavidad de β en uno de los ejes. De este modo, la función objetivo a minimizar sería de la forma:

$$PSSR(\beta) = SSR(\beta) + \lambda_t \iint \left(\frac{\partial^2 \beta}{\partial t^2}(t, s) \right)^2 dt ds + \lambda_s \iint \left(\frac{\partial^2 \beta}{\partial s^2}(t, s) \right)^2 dt ds.$$

Este planteamiento sí esquiva el problema de la penalización aplicada al Laplaciano, ya que la suma de los términos de penalización se anula si y solo si la gráfica de $\beta(t, s)$ es un plano, con curvatura 0 como se pretende. Sin embargo, tiene su propia serie de problemas. Desde el punto de vista computacional, la selección de los parámetros de suavizado λ_t y λ_s crece cuadráticamente en complejidad. Esto puede tratar de solventarse sustituyendo los dos parámetros de suavizado por uno único $\lambda = \lambda_t = \lambda_s$, pero este enfoque sigue lejos de ser una solución ideal.

El propósito de estos comentarios es comunicar el hecho de que, a pesar de que es posible aplicar metodologías de estimación análogas a la expuesta en la Sección 3.1.3, estas propuestas no resultan tan elegantes cuando se deben adaptar a una función coeficiente bivalente $\beta(t, s)$. Por lo tanto, escogemos presentar como alternativa el estimador introducido en [9], el cual deja de lado estas ideas y, en su lugar, se fundamenta en la teoría de regularización en el contexto de la regresión multivariante.

Consideremos dos bases ortonormales de L^2 , $\{\phi_j\}_{j \in \mathbb{N}}$ y $\{\psi_k\}_{k \in \mathbb{N}}$, las cuales emplearemos para expandir Y y X , respectivamente. En la práctica, la base de EFPC es una elección interesante, pero también pueden serlo otras bases fijadas previamente, como la base de Fourier. Escribimos

entonces:

$$Y(t) = \sum_{j=1}^{\infty} y_j \phi_j(t), \quad X(s) = \sum_{k=1}^{\infty} x_k \psi_k(s), \quad \varepsilon(t) = \sum_{j=1}^{\infty} e_j \phi_j(t), \quad \beta(t, s) = \sum_{j,k=1}^{\infty} b_{jk} \phi_j(t) \psi_k(s).$$

Comenzaremos reduciendo el modelo funcional a un modelo escalar sobre los coeficientes. Para ello, sustituimos las expansiones anteriores en la Ecuación (3.9) del MLF de respuesta funcional, con lo que obtenemos:

$$\sum_{j=1}^{\infty} y_j \phi_j(t) = \sum_{j,k=1}^{\infty} b_{jk} \sum_{k'=1}^{\infty} x_{k'} \langle \psi_k, \psi_{k'} \rangle \phi_j(t) + \sum_{j=1}^{\infty} e_j \phi_j(t).$$

De nuevo, como $\{\phi_j\}_{j \in \mathbb{N}}$ es una base de Hilbert, los coeficientes de cada ϕ_j deben coincidir en ambos lados de la igualdad. Aplicando además la ortonormalidad de $\{\psi_k\}_{k \in \mathbb{N}}$ tenemos:

$$y_j = \sum_{k=1}^{\infty} b_{jk} x_k + e_j, \quad j \in \mathbb{N}.$$

A continuación, reescribimos este modelo poblacional en términos de la muestra $\{(X_i, Y_i)\}_{i=1}^n$. Para poder manipular dicho modelo a nivel muestral, truncamos además las bases $\{\phi_j\}_{j \in \mathbb{N}}$ y $\{\psi_k\}_{k \in \mathbb{N}}$ por sendos índices de corte q y p . Para los propósitos que conciernen a este método de estimación, los valores de q y p deben tomarse generosamente grandes. Entonces, de nuevo escribimos informalmente la igualdad:

$$y_{ij} = \sum_{k=1}^p b_{jk} x_{ik} + e_{ij}, \quad j = 1, \dots, q, \quad i = 1, \dots, n \quad (3.15)$$

donde hemos denotado $y_{ij} = \langle Y_i, \phi_j \rangle$, $x_{ik} = \langle X_i, \psi_k \rangle$ y $e_{ij} = \langle \varepsilon_i, \phi_j \rangle$.

Para facilitar la manipulación de esta ecuación, la expresamos en lenguaje matricial. Introducimos para ello la siguiente notación: $\mathbf{Y} \in \mathcal{M}_{n \times q}$, $\mathbf{X} \in \mathcal{M}_{n \times p}$ y $\mathbf{E} \in \mathcal{M}_{n \times q}$ son las matrices de coeficientes dadas por $\mathbf{Y}(i, j) = y_{ij}$, $\mathbf{X}(i, k) = x_{ik}$ y $\mathbf{E}(i, j) = e_{ij}$, respectivamente, y $\mathbf{B} \in \mathcal{M}_{p \times q}$ viene dada por $\mathbf{B}(j, k) = b_{jk}$. Con esto, podemos ahora reescribir (3.15) de la siguiente forma⁵:

$$\mathbf{Y} = \mathbf{X}\mathbf{B}' + \mathbf{E}. \quad (3.16)$$

Con esta última simplificación nos encontramos finalmente en condiciones de presentar el estimador de β . Consideramos como término de penalización la llamada regularización *elastic-net* de B :

$$R_{EN}^{\alpha}(B) = \left[\alpha \sum_{k=1}^p \|B^k\|_2 + (1 - \alpha) \|B\|_F^2 \right], \quad (3.17)$$

⁵El motivo por el que la matriz B se transpone recae sobre ciertas elecciones, más o menos arbitrarias, que hemos realizado hasta ahora. Si en lugar de considerar la función β trabajásemos con $\tilde{\beta}(t, s) = \beta(s, t)$ (esto es, si integrásemos a lo largo de la primera coordenada en lugar de la segunda, tal y como se hace en [9]) no habríamos de tomar la transpuesta.

donde denotamos por B^k la k -ésima columna de la matriz B , por $\|\cdot\|_2$ la norma euclidiana usual en \mathbb{R}^q y por $\|\cdot\|_F$ la norma matricial de Frobenius:

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n A(i,j)^2}, \quad A \in \mathcal{M}_{m \times n}.$$

La regularización *elastic-net* aspira a combinar las ya mencionadas, y altamente extendidas, regularizaciones Ridge y LASSO, a través del parámetro de interpolación $\alpha \in [0, 1]$. En efecto, estableciendo $\alpha = 0$ se tiene que el término de penalización coincide con la penalización L_1 o LASSO, y lo propio ocurre para $\alpha = 1$ con la penalización L_2 o Ridge.

Con esto, el estimador de β viene dado por

$$\hat{\beta}(t, s) = \sum_{j=1}^q \sum_{k=1}^p \hat{b}_{jk} \phi_j(t) \psi_k(s), \quad (3.18)$$

donde los coeficientes \hat{b}_{jk} son las entradas de la matriz \hat{B} que minimiza la siguiente función objetivo:

$$PSSR^\alpha(B) = \|\mathbf{Y} - \mathbf{X}B'\|_F^2 + \lambda R_{EN}^\alpha(B), \quad (3.19)$$

en la que $\lambda \geq 0$ es una vez más el parámetro que regula la intensidad de la regularización. El primer sumando de $PSSR^\alpha$ es la suma de residuos cuadrados (en norma) de las funciones truncadas, la cual puede ser expresada como tal gracias a la ortonormalidad de las ϕ_j .

Además de este estimador clásico, los autores proponen en [9] un nuevo método de estimación que aspira a combinar las ventajas de la regularización LASSO y la simplicidad de la regresión basada en FPC. En esta metodología se emplean las bases de EFPC para la expansión inicial de los objetos funcionales. Consiste en llevar a cabo una selección de variables usando la penalización LASSO, para lo cual se calcula un primer estimador \hat{B} con el método propuesto tomando $\alpha = 0$. Se espera entonces que \hat{B} contenga un número de columnas nulas. Descartando las EFPC de X correspondientes a dichas columnas, se calcula el estimador final $\tilde{\beta}$ con matriz de coeficientes

$$\tilde{B} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \mathbf{Y} \in \mathcal{M}_{\tilde{p} \times q},$$

donde $\tilde{\mathbf{X}} \in \mathcal{M}_{n \times \tilde{p}}$ es la matriz de *scores* de las X_i en las $\tilde{p} < p$ EFPC tales que $\hat{B}^k \neq 0$.

3.2.3. Contraste de significación

Como en la primera parte de este capítulo, cerraremos el estudio del MLF de respuesta funcional mostrando una propuesta de test para el contraste de significación del modelo.

En el caso del MLF de respuesta funcional, se pretende contrastar la hipótesis nula

$$H_0 : \mathbf{B} = 0$$

frente a

$$H_1 : \mathbf{B} \neq 0.$$

El procedimiento del test se basa en el estadístico introducido en [15], el cual, de forma análoga al caso del MLF de respuesta escalar, se fundamenta en el operador de covarianza cruzada:

$$\Delta(h) = \mathbb{E}[\langle X, h \rangle Y], \quad h \in L^2.$$

El desarrollo del estadístico que presentamos a continuación es, en esencia, una adaptación al caso del modelo completamente funcional del desarrollo visto para el estadístico D_n en la Sección 3.1.5. La simetría adicional de este modelo hace que la presentación sea en cierto modo más sencilla.

Tal y como ocurría en el caso del MLF de respuesta escalar, la significación del modelo no se describe de forma precisa por el hecho de que \mathbf{B} sea o no nulo. Una vez más, por la Observación 2.12 tenemos que $X \in H = \overline{\text{Im } C_X}$ c.s. (recordamos que estamos suponiendo $\mathbb{E}[X] = 0$), luego, la variable X tiene un efecto lineal nulo sobre Y si y solo si $\mathbf{B}|_H = 0$. Sin pérdida de generalidad, procederemos con el desarrollo suponiendo que $\ker C_X = 0^6$.

Adicionalmente, trabajaremos bajo las siguientes hipótesis: se asumirá que $\mathbb{E} \|X\|^4 < \infty$ y $\mathbb{E} \|\varepsilon\|^4 < \infty$, y que los autovalores λ_k y γ_j de C_X y C_Y , respectivamente, son distintos (y positivos) hasta ciertos índices p y q :

$$\lambda_1 > \dots > \lambda_p > \lambda_{p+1}, \quad \gamma_1 > \dots > \gamma_q > \gamma_{q+1}.$$

Volvamos ahora al planteamiento del contraste. Tal y como ocurría en el MLF de respuesta escalar, la identidad

$$\mathbf{B}C_X = \Delta \tag{3.20}$$

establece la relación entre el operador de covarianza cruzada y el problema de la significación de \mathbf{B} . Por la Proposición 3.2 sabemos que esta ecuación se sostiene también en el caso del MLF de respuesta funcional.

Ahora, por un argumento absolutamente análogo al de la Sección 3.1.5 se demuestra que $\mathbf{B} = 0$ si y solo si $\Delta = 0$. Por lo tanto, el contraste de significación puede reformularse como el contraste de la hipótesis nula

$$H_0 : \Delta = 0$$

frente a

$$H_1 : \Delta \neq 0.$$

⁶En adelante asumimos implícitamente que $\dim H < \infty$. El caso de dimensión finita es más sencillo, pues tomando $p = \dim H$ dejamos de perder información de la variable X y el método se vuelve más exacto.

Ahora consideramos las FPC de X , $\{\psi_k\}_{k \in \mathbb{N}}$. Estas forman una base del espacio, puesto que estamos suponiendo $\ker C_X = 0$. Consecuentemente, $\Delta = 0$ si y solo si $\Delta(\psi_k) = 0$ para todo $k \in \mathbb{N}$. Ahora, como por definición Δ toma valores en $\text{Im } C_Y$, se tiene que $\Delta(\psi_k)$ es 0 si lo son todos sus coeficientes en la base $\{\phi_j\}_{j \in \mathbb{N}}$ de FPC de Y , la cual es una base de Hilbert de $\overline{\text{Im } C_Y}$. Con esto llegamos a que la condición de nulidad de Δ equivale a

$$\langle \Delta(\psi_k), \phi_j \rangle = 0, \quad \forall j, k \in \mathbb{N}.$$

Esta condición motiva la definición del siguiente estadístico:

$$\hat{T}_n(p, q) = n \sum_{j=1}^q \sum_{k=1}^p \hat{\gamma}_j^{-1} \hat{\lambda}_k^{-1} \left\langle \Delta_n(\hat{\psi}_k), \hat{\phi}_j \right\rangle^2, \quad (3.21)$$

donde Δ_n denota el estimador empírico del operador de covarianza cruzada, que viene dado por $\Delta_n(h) = n^{-1} \sum_{i=1}^n \langle X_i, h \rangle Y_i$.

En [15] se demuestra que bajo la hipótesis nula $H_0 : \Delta = 0$ el estadístico $\hat{T}_n(p, q)$ converge en distribución a una variable aleatoria χ_{pq}^2 con pq grados de libertad.

Con esto, el procedimiento del test se resume en los siguientes pasos. Primero deben seleccionarse los índices de truncamiento p y q . A continuación, se calcula el estadístico $\hat{T}_n(p, q)$ a partir de la muestra, y su valor es comparado con el valor crítico $\chi_{pq, 1-\alpha}^2$. Este se trata del cuantil $(1 - \alpha)$ de la distribución χ_{pq}^2 , siendo α es el nivel nominal fijado para el test, habitualmente $\alpha = 0.05$. La hipótesis nula $H_0 : \mathbf{B} = 0$ será rechazada si $\hat{T}_n(p, q) > \chi_{pq, 1-\alpha}^2$, y no lo será en caso contrario.

Un lector observador puede advertir que, a pesar de las similitudes, existe una diferencia notable entre el estadístico D_n presentado en la Sección 3.1.5 para el contraste de significación en el MLF de respuesta escalar y el estadístico $\hat{T}_n(p, q)$. En este último, los índices de corte p y q son fijados, en lugar de crecer con el tamaño muestral como ocurría con p_n .

En la práctica esto no supone gran diferencia, más allá tal vez del proceso de selección de dichos índices de corte. Sin embargo, sí tiene consecuencias importantes a la hora de establecer las propiedades asintóticas del test. Para garantizar la consistencia del test basado en $\hat{T}_n(p, q)$ debe imponerse adicionalmente la hipótesis $\langle \mathbf{B}(\psi_k), \phi_j \rangle \neq 0$ para ciertos $k \leq p$ y $j \leq q$. En caso de que $\mathbf{B}(\psi_k) = 0$ para todo $k \leq p$ y \mathbf{B} solo deje de anularse pasada la p -ésima FPC (o bien si las imágenes de estas p primeras FPC no están contenidas en $\text{span}\{\phi_j\}_{j=1}^q$) el test no tiene potencia.

Como decimos, en un contexto práctico estas cuestiones son de menor importancia. Un investigador puede estar interesado en contrastar si las mayores fuentes de variabilidad de X influyen en la variable Y . Por lo tanto, siempre que los índices p y q sean razonablemente grandes (por

ejemplo, si las p y q primeras FPC explican el 90% o el 99% de la variabilidad de las respectivas funciones aleatorias) las conclusiones del contraste deberían ser fiables.

Capítulo 4

Aplicación a datos de CGM

La diabetes mellitus, o simplemente diabetes, es una enfermedad crónica cuya principal característica es la elevada concentración de glucosa en sangre, debida habitualmente a una deficiencia en la producción o acción de la insulina. Un control estricto de los niveles glucémicos es fundamental para la prevención de complicaciones renales y cardiovasculares, junto con otras enfermedades, en pacientes diabéticos.

Con este objetivo, desde comienzos de siglo se apuesta por la monitorización continua de glucosa (CGM) como una parte importante del cuidado de la diabetes. Los dispositivos de CGM miden la concentración de glucosa en líquido intersticial —con lo que se aproxima su nivel en sangre— de manera regular y frecuente. El ensayo clínico del grupo The Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group [10], fue pionero en contrastar las ventajas de la CGM como mecanismo de control de la diabetes tipo I, obteniendo resultados positivos significativos en pacientes adultos (de más de 25 años de edad).

Los datos que emplearemos en nuestro análisis fueron extraídos de dicho ensayo clínico. Nos fueron proporcionados por el Dr. M. Matabuena¹ (ver, por ejemplo, [16]), a quien agradecemos, y que además llevó a cabo un primer filtrado y preprocesado de los datos. Tras un nuevo filtrado de datos debido a observaciones mal registradas y mediciones faltantes, la base de datos con la que trabajaremos se compone de observaciones de CGM de $n = 210$ pacientes. De cada uno de ellos se cuenta con una curva diaria de CGM (en concreto, se trata de la curva diaria promedio del paciente a lo largo de los diez primeros días del estudio), conformada por $J = 288$ mediciones del nivel de glucosa en líquido intersticial (en mg/dL) espaciadas 5 minutos. Escogemos representar este dato funcional, que denotaremos por X , con dominio el intervalo $[0, 24]$. Las curvas brutas $X_i(t)$ pueden observarse en la gráfica izquierda de la Figura 4.1.

¹<https://www.marcosmatabuena.com>

En contextos clínicos, para una manipulación más simple de los datos de CGM se habitúa a tomar la media de la serie de tiempo de CGM. Este biomarcador se emplea como indicador del nivel de glucosa promedio del paciente, de forma similar a la hemoglobina glicada (aunque esta refleja el nivel medio de glucosa en un período más extenso). La base de datos cuenta con un registro de este biomarcador, que de ahora en adelante denotaremos por Y , medido en la última semana del estudio, que tuvo lugar hasta dos años después de su inicio en la mayoría de pacientes.

El objetivo de nuestro estudio será analizar si las curvas de CGM en los primeros instantes del estudio tienen poder predictivo sobre la evolución glucémica a medio-largo plazo del paciente, medida en términos del biomarcador Y . Para ello ajustaremos un modelo de regresión lineal funcional tomando las curvas X como covariable y el indicador Y como respuesta, y contrastaremos su significación. Para llevar a cabo este estudio, hemos empleado el software R y, en particular, el paquete `fda.usc` [6]. Todo el código utilizado se encuentra recogido en el Anexo B.

El primer paso del estudio es el preprocesado de los datos. Ya hemos mencionado el previo filtrado de los mismos para deshacernos de las observaciones inválidas. Con ello la base de datos se redujo a $n = 210$ observaciones, lo cual consideramos un tamaño muestral razonablemente grande. Siguiendo esto, procedimos al centrado de las dos variables, X e Y , restándoles sus respectivas medias muestrales.

Tras esto pasamos a aplicar un presuavizado a las observaciones funcionales X_i . El método de presuavizado seguido (como adelantamos en la Sección 2.5.2) fue el suavizado mediante penalización, tomando como operador diferencial lineal $L = D^2$ y penalizando con ello la segunda derivada de X . Seleccionamos una base de $K = 200$ B-splines y comparamos valores del parámetro de suavizado λ de (2.7) en distintos órdenes de magnitud, entre 10^{-5} y 1 (además de $\lambda = 0$). Para la selección de λ seguimos un criterio de GCV. La puntuación de GCV viene dada por:

$$GCV_i(\lambda) = \frac{1}{(n - \text{tr}(S_i))^2} \sum_{j=1}^J \left(X_i(t_j) - \hat{X}_i(t_j) \right)^2,$$

donde $i = 1, \dots, n$, \hat{X}_i denota la curva presuavizada con parámetro de suavizado λ correspondiente a la observación X_i y $S_i \in \mathcal{M}_{J \times J}$ se trata de la matriz *hat* que hace $SX(\mathbf{t}) = \hat{X}(\mathbf{t})$, con $\mathbf{t} = (t_1, \dots, t_J)$.

Dado que las puntuaciones de GCV se computan individualmente para cada curva, las promediamos para definir un criterio de selección global. El valor óptimo del parámetro de suavizado bajo el criterio de GCV promedio fue $\lambda = 0.001$.

En la Figura 4.1 pueden compararse las curvas de CGM antes (izquierda) y después (derecha) de la aplicación del presuavizado. Adicionalmente, en la gráfica derecha se representa el valor

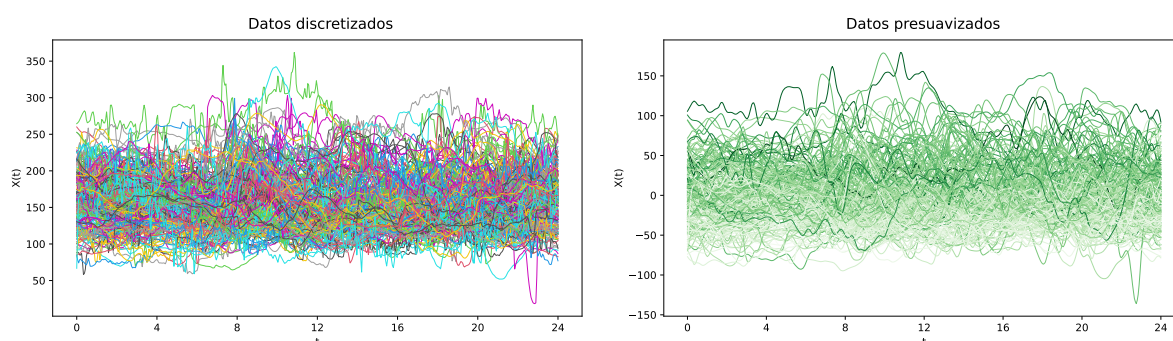


Figura 4.1: Comparativa de las curvas de CGM previamente (izquierda) y tras la aplicación del presuavizado (derecha). En la gráfica derecha, el tono verde de cada observación refleja el valor correspondiente de la respuesta (cuanto más oscuro, mayor el valor de Y).

de la respuesta Y_i correspondiente a cada curva X_i a través del color de la misma. Cuanto más oscuro el tono de verde, mayor el valor de la respuesta. La gráfica ilustra la dependencia entre la covariable y la respuesta. Por otro lado, además del evidente cambio en la forma de las curvas como consecuencia del presuavizado, cabe mencionar que en la gráfica derecha se suma también el efecto del centrado de las mismas. En particular, la escala del eje de ordenadas varía con respecto a la gráfica izquierda, y la tendencia global de las curvas cambia ligeramente, aunque este cambio no es especialmente notable. Esto se debe a que, como puede observarse en la Figura 2.1, su media es aproximadamente constante, por lo que el centrado esencialmente solo supone el desplazamiento vertical de las curvas.

Contamos con pocas herramientas para explorar los datos previamente a la regresión. En la Figura 2.2 fue representado el núcleo de covarianzas (muestral) de la variable X . En ella observamos que las variables $X(t)$ y $X(s)$ están positivamente correlacionadas para cualesquiera instantes de tiempo. A lo largo de la diagonal, se aprecia que los momentos del día con niveles de glucosa más variables se encuentran entre las 8:00 y las 16:00, lo cual resulta esperable dada la gran influencia que las principales comidas ejercen sobre dichos niveles. Por último, cabe destacar que la periodicidad inherente a los datos se refleja en los aumentos en la covarianza observados en los extremos $(0, 24)$ y $(24, 0)$.

Por otro lado, en la Figura 2.3 se muestran las cuatro primeras EFPC de X . Destacamos en particular la forma de la primera de ellas, que es aproximadamente constante. Esto indica que el mayor modo de variabilidad de los datos de CGM (con un porcentaje de varianza explicada cercano al 50 %) proviene de desplazamientos verticales de las curvas. El resto de EFPC no tienen formas fácilmente interpretables, pues están sujetas a las condiciones de ortogonalidad con las anteriores, y no son tan interesantes.

Tras este breve análisis exploratorio nos adentramos en el estudio propio de la regresión. Antes

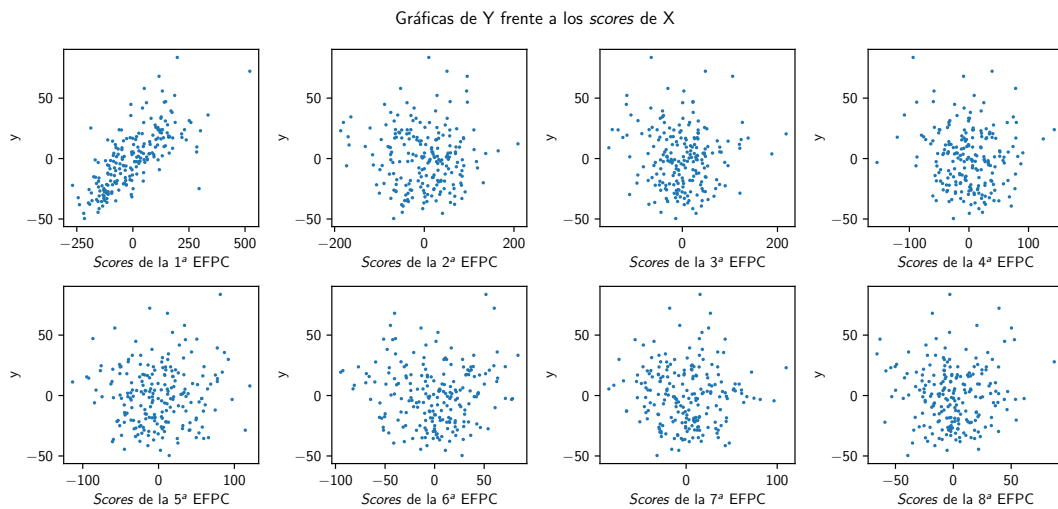


Figura 4.2: Gráficas de las observaciones Y_i frente a los *scores* de las ocho primeras EFPC de X .

de proceder al ajuste de un modelo lineal, conviene comprobar que se verifican las hipótesis del modelo. En este texto no hemos presentado ninguna herramienta para contrastar la especificación del modelo, pero podemos evaluar la hipótesis de linealidad de manera informal representando la respuesta frente a los *scores* de las primeras EFPC de X . Si el modelo es lineal, en cada una de estas gráficas debería observarse una nube de puntos con tendencia lineal. En la Figura 4.2 representamos las gráficas de Y frente a los *scores* de las ocho primeras EFPC de X , las cuales explican un 90% de la variabilidad de la función aleatoria. Observamos que, en efecto, no se distingue ningún comportamiento claramente no lineal en estas gráficas. De hecho, salvo en la primera de ellas, no se aprecia ninguna dependencia entre las variables. Esta comprobación nos da más seguridad a la hora de proceder con la regresión.

Finalmente consideramos el ajuste del modelo. Para la estimación de la función coeficiente empleamos el método de estimación mediante penalización, tomando una base de B-splines con $K = 200$ funciones básicas (la misma que fue empleada para la expansión de X). Aplicaremos la penalización a la segunda derivada de β . De nuevo, la dificultad en la aplicación de este método recae en la selección del parámetro de suavizado λ en (3.4). En esta ocasión decidimos emplear como criterio de selección la CV predictiva (*leave-one-out CV*), ya que, a pesar de ser mucho más costosa computacionalmente que la GCV, la CV se trata de un criterio más robusto.

Consideramos valores de λ abarcando de nuevo múltiples órdenes de magnitud, desde 10^{-3} (además de $\lambda = 0$) hasta 10^{10} . Entonces, buscamos el valor de λ entre los anteriores que minimiza

$$CV(\lambda) = \frac{1}{n} \sum_{i=1}^n \left(Y_i - \langle \hat{\beta}^{(-i)}, X_i \rangle \right)^2,$$

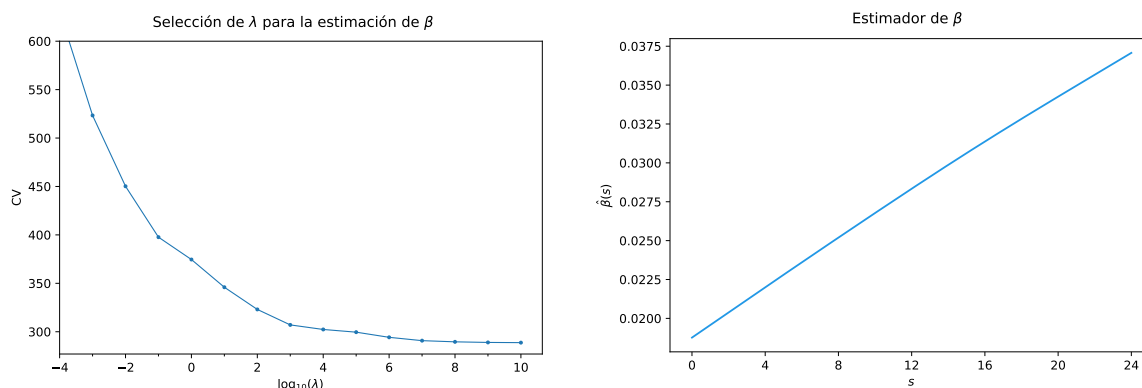


Figura 4.3: Gráfica izquierda: puntuaciones de CV frente a los valores de λ comparados (en escala logarítmica). Gráfica derecha: estimador $\hat{\beta}$ de la función coeficiente calculado con el parámetro de suavizado óptimo $\lambda = 10^{10}$.

donde $\hat{\beta}^{(-i)}$ denota el estimador de β obtenido al ajustar el modelo con todas salvo la i -ésima observación, para el parámetro de suavizado λ .

Las puntuaciones de CV obtenidas para cada valor de λ aparecen representadas en la gráfica izquierda de la Figura 4.3. A su derecha graficamos el estimador $\hat{\beta}$ calculado (con todas las observaciones) con el valor óptimo resultante de λ , el cual fue $\lambda = 10^{10}$.

Lo primero que notamos es que el valor óptimo del parámetro de suavizado es extremadamente alto. En consecuencia, la penalización cobra un gran peso y obliga a que la forma de la función objetivo sea una recta. La tendencia de las puntuaciones de CV en la gráfica izquierda de la Figura 4.3 nos comunica que valores menores de λ , que permiten estimadores menos suaves, conducen a un mayor sobreajuste del modelo.

El valor óptimo de λ es uno de los valores extremos que hemos comparado, por lo que en principio podría ocurrir que extendiendo el rango de valores hacia la derecha encontrásemos otros con una puntuación de CV menor. Sin embargo, se puede apreciar que las diferencias entre $\lambda = 10^{10}$ y los valores que lo preceden son extremadamente pequeñas, y cualquier estimador obtenido con una mayor penalización tendrá la misma forma que $\hat{\beta}$. Por estos motivos, consideramos que no es necesario comparar más valores del parámetro de suavizado.

Una vez realizado el ajuste podemos emplear una herramienta conocida de la regresión lineal multivariante para el diagnóstico del modelo. Las denominadas gráficas de residuos son una herramienta exploratoria estándar de la regresión lineal usual, y son también aplicables en el caso de la regresión lineal funcional con respuesta escalar.

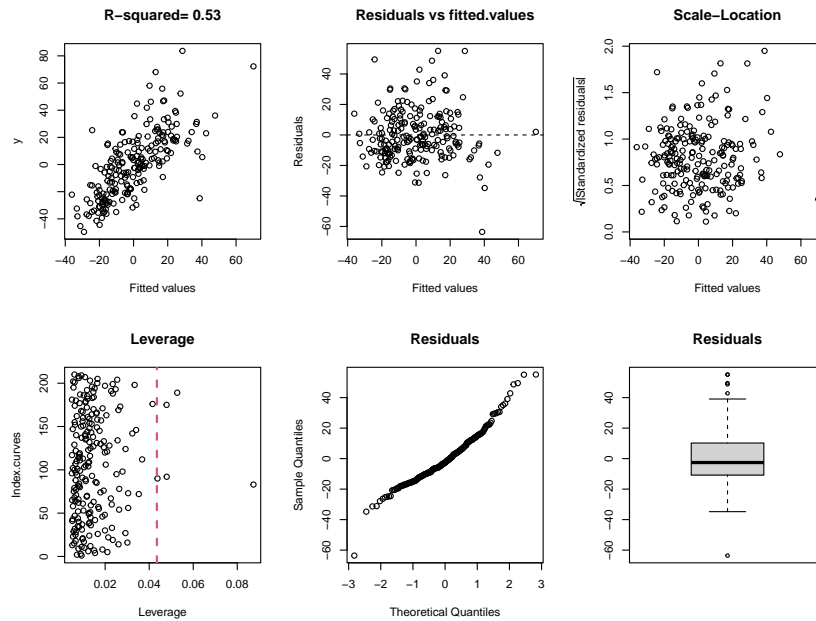


Figura 4.4: Gráficas de residuos del modelo ajustado por $\hat{\beta}$.

En la Figura 4.4 podemos observar las gráficas de residuos del modelo ajustado. No nos detendremos a explicar cada una de ellas, pero sí las usaremos para comunicar, de nuevo, que las hipótesis del modelo parecen cumplirse. La primera de las gráficas (arriba izquierda) muestra las observaciones Y_i frente a los valores ajustados, y además nos informa del valor del coeficiente de determinación $R^2 = 0.53$, que indica que el modelo explica aproximadamente el 53% de la variabilidad de la respuesta. Esto supone un ajuste razonable de los datos, aunque no es suficiente para afirmar que el modelo posee una gran capacidad predictiva. Una inspección visual de las gráficas restantes nos permite concluir informalmente que, salvo por un número de observaciones posiblemente atípicas e influyentes, las hipótesis del modelo se sostienen suficientemente bien.

Por último, resta llevar a cabo el contraste de significación de la función coeficiente β , el cual efectuaremos con el test basado en el estadístico D_n presentado en la Sección 3.1.5. Para su cálculo se hizo uso del código que figura en el Bloque de código B.4. El cálculo del estadístico requiere computar previamente un estimador de la varianza del error. Nosotros empleamos el estimador estándar descrito en [3]:

$$\hat{\sigma}^2 = \frac{1}{n - \text{tr}(S)} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2,$$

donde $\hat{Y}_i = \langle \hat{\beta}, X_i \rangle$, $i = 1, \dots, n$, son los valores ajustados y S denota la matriz *hat*, que hace $S\mathbf{Y} = \hat{\mathbf{Y}}$.

Para el test seleccionamos valores de p_n guiándonos por el estudio de simulación llevado a

cabo en [3]. Dado que los tamaños muestrales son comparables ($n = 210$ frente a $n = 200$), consideramos apropiado fijar los valores propuestos en dicho trabajo: $p_n = 5, 10$ y 20 . Al calcular el estadístico D_n para cada uno de ellos obtuvimos los valores 240.36, 240.61 y 252.73, respectivamente. En los tres casos, el p-valor resultante fue cero, por lo que existe evidencia estadísticamente significativa contra la hipótesis nula de no efecto. En consecuencia, esta se rechaza confiadamente, concluyendo que la covariable X ejerce un efecto lineal significativo sobre la respuesta Y .

Para cerrar este capítulo nos gustaría tratar de dar una interpretación de los resultados. Sin embargo, la interpretabilidad de los modelos de regresión funcionales es notoriamente complicada. En nuestro caso, la función coeficiente estimada $\hat{\beta}$ tiene una forma muy sencilla, una simple recta con pendiente positiva, pero aún así resulta difícil extraer conclusiones a partir de ella.

Por ser $\hat{\beta}$ una recta creciente que toma siempre valores positivos, podríamos concluir que los instantes hacia el final del día tienen mayor influencia que las primeras horas de la madrugada, lo cual parece razonable. De todas formas, observar la Figura 4.1 podría llevarnos a pensar que las curvas más elevadas a lo largo de todo el día son las que conducen a valores más altos de la respuesta, con lo que β estaría simplemente promediando estos valores, y actuando aproximadamente como una constante. Esto último concuerda con el hecho de que la única EFPC correlacionada con la respuesta parece ser la primera, como observamos en la Figura 4.2.

El fin de este estudio ilustrativo no es, ni mucho menos, tratar de extraer resultados de interés estadístico y/o clínico de esta base de datos. Nuestras pretensiones se limitan a ejemplificar una posible aplicación de las técnicas expuestas en el trabajo.

Referencias

- [1] Denis Bosq. *Linear Processes in Function Spaces: Theory and Applications*. Vol. 149. Lecture Notes in Statistics. New York: Springer, 2000. ISBN: 978-1-4612-1154-9. DOI: 10.1007/978-1-4612-1154-9.
- [2] Hervé Cardot, Frédéric Ferraty y Pascal Sarda. «Functional linear model». En: *Statistics & Probability Letters* 45.1 (1999), págs. 11-22. DOI: 10.1016/S0167-7152(99)00036-X.
- [3] Hervé Cardot et al. «Testing Hypotheses in the Functional Linear Model». En: *Scandinavian Journal of Statistics* 30.1 (2003), págs. 241-255. DOI: 10.1111/1467-9469.00329.
- [4] Juan A. Cuesta-Albertos et al. «Goodness-of-fit tests for the functional linear model based on randomly projected empirical processes». En: *The Annals of Statistics* 47.1 (2019), págs. 439-467. DOI: 10.1214/18-AOS1693.
- [5] Arthur E. Hoerl y Robert W. Kennard. «Ridge Regression: Biased Estimation for Nonorthogonal Problems». En: *Technometrics* 12.1 (1970), págs. 55-67. DOI: 10.1080/00401706.1970.10488634.
- [6] Manuel Febrero-Bande y Manuel Oviedo de la Fuente. «Statistical Computing in Functional Data Analysis: The R Package *fda.usc*». En: *Journal of Statistical Software* 51.4 (2012), págs. 1-28. DOI: 10.18637/jss.v051.i04.
- [7] Manuel Febrero-Bande, Pedro Galeano y Wenceslao González-Manteiga. «Functional Principal Component Regression and Functional Partial Least-squares Regression: An Overview and a Comparative Study». En: *International Statistical Review* 85.1 (2017), págs. 61-83. DOI: 10.1111/insr.12116.
- [8] Frédéric Ferraty y Philippe Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics. New York: Springer, 2006. ISBN: 978-0-387-30369-7. DOI: 10.1007/0-387-36620-2.
- [9] Eduardo García-Portugués et al. «A goodness-of-fit test for the functional linear model with functional response». En: *Scandinavian Journal of Statistics* 48.2 (2021), págs. 502-528. DOI: 10.1111/sjos.12486.

- [10] The Juvenile Diabetes Research Foundation Continuous Glucose Monitoring Study Group. «Continuous Glucose Monitoring and Intensive Treatment of Type 1 Diabetes». En: *New England Journal of Medicine* 359.14 (2008), págs. 1464-1476. DOI: 10.1056/NEJMoa0805017.
- [11] Lajos Horváth y Piotr Kokoszka. *Inference for Functional Data with Applications*. Vol. 200. Springer Series in Statistics. New York: Springer, 2012. ISBN: 978-1-4614-3654-6. DOI: 10.1007/978-1-4614-3655-3.
- [12] Tailen Hsing y Randolph Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. Chichester, UK: John Wiley & Sons, 2015. ISBN: 978-1-118-76254-7. DOI: 10.1002/9781118762547.
- [13] John D. Hunter. «Matplotlib: A 2D graphics environment». En: *Computing in Science & Engineering* 9.3 (2007), págs. 90-95. DOI: 10.1109/MCSE.2007.55.
- [14] Piotr Kokoszka y Matthew Reimherr. *Introduction to Functional Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton: Chapman & Hall/CRC, 2017. ISBN: 978-1498746342. DOI: 10.1201/9781315117416.
- [15] Piotr Kokoszka et al. «Testing for lack of dependence in the functional linear model». En: *Canadian Journal of Statistics* 36.2 (2008), págs. 207-222. DOI: 10.1002/cjs.5550360203.
- [16] Marcos Matabuena y Ciprian M. Crainiceanu. *Multilevel functional distributional models with application to continuous glucose monitoring in diabetes clinical trials*. 2024. arXiv: 2403.10514 [stat.ME].
- [17] Carlos Ramos-Carreño et al. «scikit-fda: A Python Package for Functional Data Analysis». En: *Journal of Statistical Software* 109.2 (mayo de 2024), págs. 1-37. DOI: 10.18637/jss.v109.i02.
- [18] J. O. Ramsay y B. W. Silverman. *Functional Data Analysis*. 2.^a ed. Springer Series in Statistics. New York: Springer, 2005. ISBN: 978-0-387-40080-8. DOI: 10.1007/b98888.
- [19] Robert Tibshirani. «Regression Shrinkage and Selection Via the Lasso». En: *Journal of the Royal Statistical Society: Series B (Methodological)* 58.1 (dic. de 1996), págs. 267-288. DOI: 10.1111/j.2517-6161.1996.tb02080.x.

Anexo A

Fundamentos teóricos del análisis funcional

En este anexo presentamos parte de la teoría del análisis funcional en espacios de Hilbert que, por motivos de limitación de espacio, no hemos podido incluir en el cuerpo principal del texto, pero que puede resultar de utilidad para entender con mayor precisión los contenidos. Recogemos algunos teoremas fundamentales que se invocan a lo largo del trabajo (especialmente en el Capítulo 2), pero que no enunciamos en el texto para no interrumpir la exposición.

A.1. Subespacios. Ortogonalidad

Definición A.1. Decimos que dos funciones $f, g \in H$ son *ortogonales* si $\langle f, g \rangle = 0$.

Dado un conjunto $K \subset H$, se define su *complemento ortogonal* K^\perp como

$$K^\perp = \{g \in H / \langle f, g \rangle = 0, \forall f \in K\}.$$

Proposición A.2. Para todo conjunto $K \subset H$, $K^\perp = (\overline{\text{span } K})^\perp$ es un subespacio cerrado de H , y $(K^\perp)^\perp = \overline{\text{span } K}$.

Proposición A.3. Todo subespacio cerrado de un espacio de Hilbert es a su vez un espacio de Hilbert con la estructura inducida por el espacio ambiente.

Proposición A.4. Si $K \subset H$ es un subespacio cerrado de H , entonces el espacio H puede ser descompuesto como la suma directa de K y su complemento ortogonal, esto es, $H = K \oplus K^\perp$.

Definición A.5. Una *base de Hilbert* de un espacio de Hilbert H separable es una sucesión $\{\phi_k\}_{k \in \mathbb{N}} \subset H$ que verifica:

1. $\langle \phi_i, \phi_j \rangle = \delta_{ij}, \forall i, j \in \mathbb{N}$;
2. $\overline{\text{span}\{\phi_k\}_{k \in \mathbb{N}}} = H$,

donde δ_{ij} denota la delta de Kronecker.

A.2. Propiedades básicas de operadores

Definición A.6. Una aplicación lineal $T: H \rightarrow W$ es un *operador acotado*, y se denota por $T \in \mathcal{LB}(H, W)$, si para todo conjunto acotado $B \subset H$ se tiene que $T(B)$ está acotado. Esto es equivalente a que T tenga norma de operadores finita:

$$\|T\| = \sup_{\substack{h \in H \\ \|h\|_H \leq 1}} \|T(h)\|_W < \infty.$$

Teorema A.7 (de representación de Riesz). *Sea H un espacio de Hilbert, y $L \in H^* = \mathcal{LB}(H, \mathbb{R})$ un funcional sobre H . Entonces, existe un único elemento $h \in H$ que lo caracteriza de la siguiente forma:*

$$L(f) = \langle h, f \rangle, \forall f \in H.$$

Además, la aplicación $\Phi: H \rightarrow H^*$ dada por $\Phi(h) = \langle h, \cdot \rangle$ es un isomorfismo isométrico de espacios normados, con lo que induce en H^* una estructura de espacio de Hilbert isomorfa a H .

Definición A.8. Diremos que un operador $T \in \mathcal{LB}(H)$ es *definido positivo (negativo)* si $\langle T(h), h \rangle \in \mathbb{R}$ y $\langle T(h), h \rangle > (<) 0$ para todo $h \in H$. Diremos que T es *semidefinido positivo (negativo)* si se cumple la correspondiente desigualdad no estricta.

Definición A.9. Dado un operador $T \in \mathcal{LB}(H, W)$ se define su *adjunto* como el único operador $T^* \in \mathcal{LB}(W, H)$ que verifica:

$$\langle T(h), g \rangle_W = \langle h, T^*(g) \rangle_H, \forall h \in H, g \in W.$$

Definición A.10. Un operador $T \in \mathcal{LB}(H)$ se dice *autoadjunto* si $T = T^*$.

A.3. Operadores compactos. Subclases

Definición A.11. Un operador lineal $T: H \rightarrow W$ se dice *compacto*, y se denota $T \in \mathcal{LB}_C(H, W)$, si existe un entorno $U \subset H$ del origen tal que $\overline{T(U)}$ es compacto.

Teorema A.12 (Teorema espectral para operadores compactos autoadjuntos). *Sea H un espacio de Hilbert y $T \in \mathcal{LB}_C(H)$ un operador compacto y autoadjunto en H . Entonces, existe una base de*

Hilbert $\{\phi_k\}_{k \in \mathbb{N}}$ de $\overline{\text{Im } T}$ (la cual puede estar formada por un número finito de funciones básicas en el caso de que T posea solamente finitos autovalores no nulos) formada por autofunciones de T con autovalores asociados $\lambda_k \in \mathbb{R}$ de modo que $|\lambda_1| \geq |\lambda_2| \geq \dots \geq 0$ y $\lambda_k \xrightarrow[k \rightarrow \infty]{} 0$.

Definición A.13. Un operador lineal $T: H \rightarrow W$ se dice *Hilbert-Schmidt*, y se denota $T \in \mathcal{LB}_{HS}(H, W)$, si para toda base de Hilbert $\{\phi_j\}_{j \in \mathbb{N}}$ de H verifica

$$\|T\|_{HS}^2 := \sum_{j=1}^{\infty} \|T(\phi_j)\|_W^2 < \infty.$$

El espacio de operadores $\mathcal{LB}_{HS}(H, W)$ es un espacio de Hilbert al dotarlo del producto interior

$$\langle T, S \rangle_{HS} := \sum_{k=1}^{\infty} \langle T(\phi_k), S(\phi_k) \rangle_W$$

La norma asociada a dicho producto interior es precisamente $\|\cdot\|_{HS}$. Ambas definiciones son independientes de la elección de base.

Proposición A.14. Consideremos el espacio de Hilbert $L^2(\mathcal{D})$, con $\mathcal{D} \equiv (\mathcal{D}, \Sigma, \nu)$ un espacio de medida, y el espacio de operadores Hilbert-Schmidt en el mismo, $\mathcal{LB}_{HS}(L^2(\mathcal{D}))$.

Existe una isometría entre los espacios de Hilbert $L^2(\mathcal{D}^2)$ y $\mathcal{LB}_{HS}(L^2(\mathcal{D}))$ que asocia a cada núcleo $k \in L^2(\mathcal{D}^2)$ el operador $K \in \mathcal{LB}_{HS}(L^2(\mathcal{D}))$ dado por $K(f) = \int_{\mathcal{D}} k(x, y) f(y) d\nu(y)$, con $f \in L^2(\mathcal{D})$.

En particular, todo operador Hilbert-Schmidt en L^2 puede ser representado como un operador integral.

Definición A.15. Un operador compacto y autoadjunto $T \in \mathcal{LB}(H)$ se dice de *clase traza* o *nuclear*, y se denota $T \in \mathcal{LB}_{TR}(H)$, si, denotando por $\{\lambda_k\}_{k \in \mathbb{N}}$ sus autovalores (los cuales son reales), se cumple

$$\|T\|_{TR} := \sum_{k=1}^{\infty} |\lambda_k| < \infty.$$

Entonces, la serie $\sum_{k=1}^{\infty} \lambda_k$ está bien definida y es finita, y se conoce por la *traza* de T :

$$\text{tr}(T) := \sum_{k=1}^{\infty} \lambda_k.$$

Observación A.16. Es posible definir $\|\cdot\|_{TR}$ en el caso de un operador compacto $T \in \mathcal{LB}_C(H, W)$, y con ello definir $\mathcal{LB}_{TR}(H, W)$. Esto requiere un poco más de trabajo y decidimos omitirlo ya que no será de importancia para este trabajo

Proposición A.17. Se tiene la siguiente cadena de inclusiones:

$$\mathcal{LB}_{TR}(H, W) \subset \mathcal{LB}_{HS}(H, W) \subset \mathcal{LB}_C(H, W) \subset \mathcal{LB}(H, W),$$

y además, para todo operador T se tiene $\|T\| \leq \|T\|_{HS} \leq \|T\|_{TR}$.

Definición A.18. Sean $f \in H$ y $g \in W$ elementos de espacios de Hilbert H y W . Se define su *producto tensor* como el operador $(g \otimes f): H \rightarrow W$ dado por:

$$(g \otimes f)(h) = \langle f, h \rangle g, \quad h \in H.$$

Proposición A.19. Si $\{\psi_k\}_{k \in \mathbb{N}}$ y $\{\phi_j\}_{j \in \mathbb{N}}$ son bases de Hilbert de los espacios H y W , respectivamente, entonces $\{(\phi_j \otimes \psi_k)\}_{j,k=1}^{\infty}$ es un conjunto linealmente independiente y que genera un subespacio denso en el espacio de operadores de cualquiera de las subclases consideradas anteriormente, es decir, $\mathcal{LB}_C(H, W)$, $\mathcal{LB}_{HS}(H, W)$ y $\mathcal{LB}_{TR}(H, W)$ (como espacios normados con las apropiadas normas $\|\cdot\|$, $\|\cdot\|_{HS}$ y $\|\cdot\|_{TR}$).

De hecho, en el caso del espacio de Hilbert de los operadores Hilbert-Schmidt, $\{(\phi_j \otimes \psi_k)\}_{j,k=1}^{\infty}$ es un conjunto ortonormal, luego, forma una base de Hilbert del espacio.

Anexo B

Código

En este anexo recogemos, en varios bloques de código, el código de R empleado para llevar a cabo el estudio del Capítulo 4 sobre los datos de CGM.

```
1 library(fda.usc)
2 # 1. LECTURA DE DATOS
3 # 1.1. Lectura del archivo .csv
4 data = read.csv('clean_cgm_data.csv')
5 y = data$biomarcador1
6 X.data = as.matrix(data[, 4:ncol(data)])
7 n = nrow(X.data); J = ncol(X.data)
8 n; J # == 210; 288
9
10 # 2. PREPROCESADO DE LOS DATOS
11 # 2.1. Centrado de las variables
12 X.data = sweep(X.data, 2, colMeans(X.data), FUN = "-")
13 y = y - mean(y)
14 max(abs(colMeans(X.data))); mean(y) # == 0
15
16 # 2.2. Representacion como datos funcionales discretizados
17 argvals = seq(0, 24, length=J)
18 X.discr = fdata(mdata = X.data, argvals = argvals)
19
20 # 2.3. Suavizado por medio de una expansion en base mediante penalizacion
21 search = optim.basis(fdataobj = X.discr,
22                     lambda = c(0, 1e-5, 1e-4, 0.001, 0.01, 0.1, 1),
23                     numbasis = 200, # base de B-splines por defecto
24                     Lfdobj = int2Lfd(2))
25 search$lambda.opt # == 1e-3
26 X = search$fdata.est
27 basis = search$base.opt
```

Bloque de código B.1: Código empleado para la lectura y preprocesado de los datos de CGM.

```

1 # 3. REGRESION
2 # 3.1. Búsqueda del parametro de suavizado optimo en base a CV
3 loglambda.grid = seq(-3, 10, length = 14)
4 lambda.grid = c(0, 10**loglambda.grid)
5 l = length(lambda.grid)
6 CV = numeric(l)
7
8 compute_CV_score = function(lambda) {
9   print(paste0('Testing value lambda = ', lambda))
10  score = 0
11  for (i in 1:n) {
12    X_noi = X[-i]
13    y_noi = y[-i]
14    model_noi = fregre.basis(fdataobj = X_noi, y = y_noi,
15                           basis.x = basis, basis.b = basis,
16                           lambda = lambda,
17                           Lfdobj = int2Lfd(2))
18    score = score + (y[i] - predict(model_noi, X[i]))**2
19    progress_bar(i, n)
20  }
21  return(score/n)
22 }
23
24 for (lambda.ind in 1:l) {
25   CV[lambda.ind] = compute_CV_score(lambda.grid[lambda.ind])
26 }
27
28 # 3.2. Ajuste del modelo
29 best.lambda = lambda.grid[which(CV == min(CV))] # == 1e10
30 model = fregre.basis(fdataobj = X, y = y,
31                    basis.x = basis, basis.b = basis,
32                    lambda = best.lambda,
33                    Lfdobj = int2Lfd(2))
34 beta.hat = model$beta.est

```

Bloque de código B.2: Código empleado para hallar el valor óptimo de λ en base al criterio de CV, y posteriormente ajustar el modelo.

```

1 # 4. CONTRASTE DE SIGNIFICACION
2 # 4.1. Calculo del estimador de la varianza del error
3 residuals = model$residuals
4 df = sum(diag(model$H))
5 sigma2.hat = sum(residuals**2) / (n - df)
6
7 # 4.2. Contraste
8 d_n5 = D_n(X, y, p_n = 5, sigma2.hat = sigma2.hat)
9 # [1] "The p-value associated to D_n=240.364303403471 with 5 df is: 6.41338840349712e-50"

```

```

10 d_n10 = D_n(X, y, p_n = 10, sigma2.hat = sigma2.hat)
11 # [1] "The p-value associated to D_n=240.6156542427 with 10 df is: 5.0874201655224e-46"
12 d_n20 = D_n(X, y, p_n = 20, sigma2.hat = sigma2.hat)
13 # [1] "The p-value associated to D_n=252.73265455742 with 20 df is: 3.2101834293436e-42"

```

Bloque de código B.3: Código empleado para efectuar el contraste de significación. La definición de la función `D_n` puede encontrarse en el Bloque de código B.4.

```

1 # 4.0. Definicion de la funcion D_n
2 D_n = function(X, y, p_n, sigma2.hat = NULL, print.pvalue = TRUE) {
3   # Calculates the D_n test statistic for no effect proposed in Cardot et al., 2003.
4   # Arguments:
5   # - X: functional covariate (class 'fd' or 'fdata')
6   # - y: scalar response (numeric)
7   # - p_n: degrees of freedom for the chisq test
8   # - sigma2.hat: estimator of error variance
9   # Returns:
10  # - D_n statistic value
11
12  # Check X arg
13  if (class(X) == 'fdata') {
14    argvals = X$argvals
15    X.fdata = X
16    X.fd = fdata2fd(X)
17  }
18  else if (class(X) == 'fd') {
19    argvals = seq(X$basis$rangeval[1], X$basis$rangeval[2], length=J)
20    X.fdata = fdata(mdata = t(eval.fd(evalarg = argvals, fdoj = X)),
21                  argvals = argvals)
22    X.fd = X
23  }
24  else { stop('X object must be of class 'fd' or 'fdata') }
25
26  # Check length of X and y
27  n = length(y)
28  if (nrow(X.fdata$data) != n) { stop('y and X are not of the same length.') }
29
30  # Check p_n
31  if (p_n < 1) { stop('p_n must be greater than 0.') }
32  if (p_n >= n - 1) { stop('Max value for p_n is length(y)-1.')}
33
34  # Check if variables are centered
35  tol = 1e-6
36  if (mean(y) > tol) {
37    y = y - mean(y)
38  }

```

```

39 mu.X = mean.fd(X.fd)
40 if (inprod(mu.X, mu.X) > tol) {
41   new.coefs = X.fd$coefs - matrix(rep(mu.X$coefs, n), ncol = n)
42   X.fd = fd(new.coefs, X.fd$basis)
43   X.fdata = fdata(mdata = t(eval.fd(evalarg = argvals, fdoobj = X.fd)),
44                 argvals = argvals)
45 }
46
47 # Calculate sigma2.hat if NULL
48 if (is.null(sigma2.hat)) {
49   loglambda.grid = seq(-5, 5, length = 10)
50   lambda.grid = 10 ** loglambda.grid
51   search = fregre.basis.cv(fdataobj = X.fdata, y = y,
52                          lambda = lambda.grid, Lfdoobj = int2Lfd(2))
53   sigma2.hat = search$fregre.basis$sr2
54 }
55
56 # Calculate g_n
57 K = X.fd$basis$nbasis
58 g_n = fd(coef = matrix(0, nrow = K, ncol = 1), basisobj = X.fd$basis)
59 for (i in 1:n) {
60   g_n = g_n + y[i] * X.fd[i]
61 }
62 g_n = 1/n * g_n
63
64 # Compute A_n.g_n
65 fpca = pca.fd(X.fd, nharm = p_n)
66 efpc = fpca$harmonics
67 lambdas = fpca$values
68 A_n.g_n = fd(coef = matrix(0, nrow = K, ncol = 1), basisobj = X.fd$basis)
69 for (k in 1:p_n) {
70   A_n.g_n = A_n.g_n + inprod(g_n, efpc[k])[1,1] / sqrt(lambdas[k]) * efpc[k]
71 }
72
73 # Return
74 d_n = n / sigma2.hat * inprod(A_n.g_n, A_n.g_n)[1,1]
75 if (print.pvalue) {
76   print(paste0("The p-value associated to D_n=", d_n, " with ", p_n,
77              " df is: ", pchisq(d_n, df = p_n, lower.tail = FALSE)))
78 }
79 return(d_n)
80 }

```

Bloque de código B.4: Implementación del cálculo del estadístico D_n para el contraste de significación del MLF de respuesta escalar.