

Capítulo 1

Introducción

1.1 Datos Funcionales

Existe un importante número de situaciones provenientes de las ciencias aplicadas (química, biométrica, medicina, economía, etc) en donde los datos recolectados resultan ser curvas. En efecto, el desarrollo de las herramientas computacionales actuales nos permiten trabajar con conjuntos grandes de datos, en donde “grande” se debe entender por un lado con respecto a la cantidad de datos y por otro lado por el tamaño de cada uno de ellos. Es decir, para un fenómeno particular podríamos llegar a observar un gran número de variables, teniendo así una alta dimensionalidad de la información. Por ejemplo, si tenemos un proceso estocástico indexado en tiempo, tendríamos entonces $\{X(t_j)\}_{j=1,\dots,J}$. La grilla de puntos en donde se observan los datos podría ser cada vez más fina y los puntos consecutivos cada vez más cercanos entre sí. Una manera de considerar este problema es pensar que nuestro dato es una observación discretizada del proceso $\{X(t), t \in (t_{\min}, t_{\max})\}$. En general, la idea es que este elemento X presenta alguna información de regularidad que puede ser explotado de forma funcional, por ejemplo, que sea continua o diferenciable. Como en Ferraty y Vieu (2006) diremos que un elemento aleatorio $X : \Omega \rightarrow \mathcal{H}$ es una variable funcional si \mathcal{H} , el espacio de posibles valores de X , tiene dimensión infinita. Una observación de X se dirá un dato funcional

Ejemplos de datos funcionales podrían ser la temperatura registrada a lo largo de un período por una estación climática posicionada en un determinado lugar. Si consideramos todas las estaciones climáticas, tendríamos entonces un conjunto de datos funcionales. Ramsay y Silverman (2005) proveen un análisis de este tipo de observaciones. Cabe aclarar que el hecho que el dato sea funcional no implica necesariamente que tenga que ser una curva en el sentido estricto de la palabra. Podríamos estar considerando también imágenes o superficies, aspectos quizás más complejos. Lo importante en cualquier caso es que el dato es un elemento de un espacio infinito-dimensional, y cuenta, preferentemente, con algún tipo de regularidad que se pueda aprovechar.

1.2 Ejemplos

Estaciones climáticas

Ramsay y Silverman (2005) consideran el seguimiento anual de las temperaturas de 35 estaciones climáticas de Canadá; dichas temperaturas se muestran en la Figura 1.1.

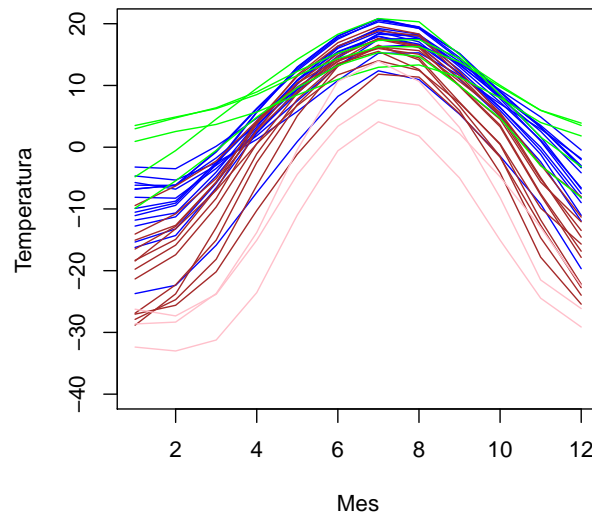


Figura 1.1: Temperatura de 35 estaciones climáticas en Canadá

Movimientos del labio

Otro ejemplo lo constituye el siguiente conjunto de datos considerado entre otros por Gervini (2008) y Bali y Boente (2014). Se les solicita a un individuo que diga la palabra *bob* 32 veces y se registra la posición del labio inferior en cada instante. El movimiento de los labios fue estudiado originalmente por Malfait and Ramsay (2003). En la Figura 1.2 las curvas graficadas corresponden a las 32 trayectorias del labio inferior versus el tiempo. Tres de esas curvas (graficadas con líneas gruesas en la Figura 1.2) parecen tener un patrón diferente, con un retraso en el segundo pico.

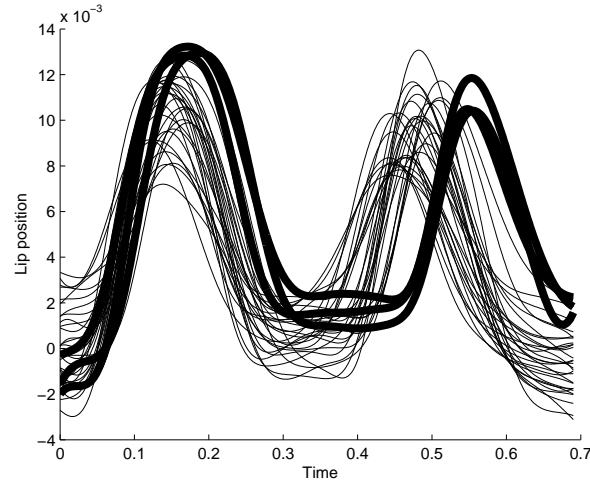


Figura 1.2: Datos del movimiento del labio. Trayectorias de un individuo que pronuncia la palabra *bob* 32 veces. Las trayectorias 24, 25 y 27 se indican con líneas gruesas.

Datos espectrométricos

El conjunto de datos espectrométricos fue descrito y analizado en Ferraty y Vieu (2006), Aneiros-Pérez y Vieu (2006) y Shang (2014). Los datos pueden obtenerse en <http://lib.stat.cmu.edu/datasets/tecator> y están disponibles en la librería `fda.usc` en el dataset `tecator`. Cada muestra contiene carne triturada con distintos porcentajes de grasa, proteína y contenido de humedad. Para cada muestra se observa una curva espectrométrica, indicada X_i , que corresponde a la absorbancia medida en una grilla de 100 longitudes de onda desde 850nm a 1050nm. Los porcentajes de grasa, proteínas y contenido de humedad se determinan por un procedimiento analítico. La tabla siguiente da las primeras 3 filas del conjunto de datos y las primeras 100 columnas corresponden a los 100 canales del espectro.

						Protein	Fat	Moisture
Row	Col 1	Col 2	Col 3	...	Col 100	Col 101	Col 102	Col 103
1	2.6177	2.6181	2.6185	...	0.8996	60.5	22.5	16.7
2	2.8558	2.8606	2.8656	...	0.6203	46	40.1	13.5
3	2.5828	2.5845	2.5862	...	0.6905	71	8.4	20.5

La Figura 1.3(a) muestra todas las observaciones correspondientes a la absorbancia. Dicha figura ilustra claramente que

- Las unidades aparecen claramente como una curva discretizada.

- Debido de la precisión de la grilla podemos considerar a cada individuo como una curva continua correspondiente a una función de la longitud de onda.

Este hecho fue observado por Leurgans, Moyeed y Silverman (1993) que mencionan que para estos datos *the spectra observed are to all intents and purposes a functional observations*.

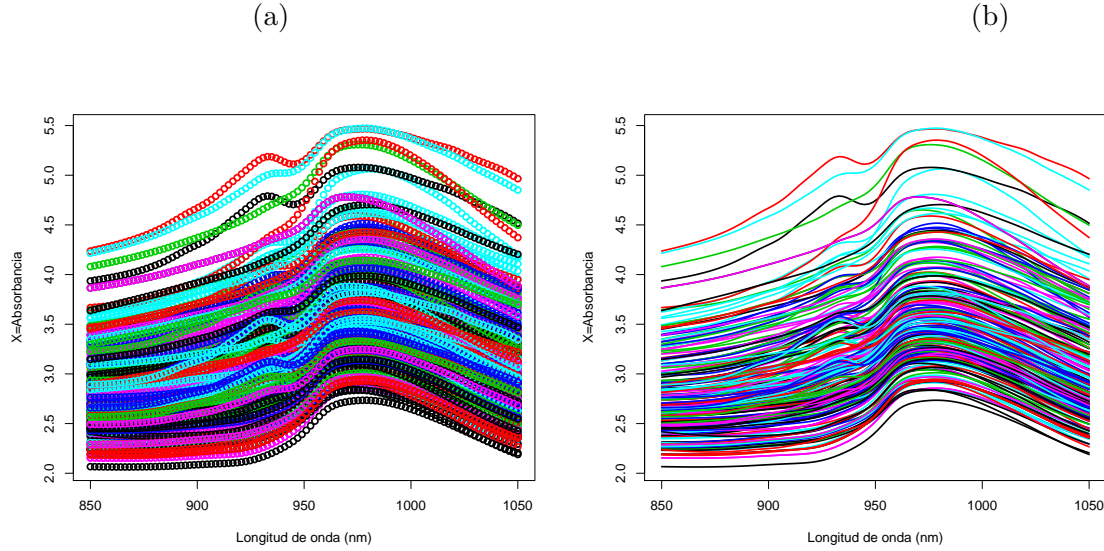


Figura 1.3: Datos TECATOR. (a) Puntos observados. (b) Curvas.

Otra característica de estos datos es que las curvas espectrométricas parecen similares excepto por un corrimiento vertical. Por esta razón, en el análisis de estos datos suelen tomarse las derivadas del espectro que se presentan en la Figura 1.4.

Cabe mencionar que obtener la curva espectrométrica es más económico en tiempo y costo que el procedimiento analítico para determinar el contenido graso. Por esta razón, resulta interesante obtener un modelo que permita predecir el contenido graso de una muestra a partir de su contenido de proteína y agua así como del espectro de su absorbancia. Entre los modelos considerados está el modelo parcialmente lineal funcional. Más precisamente, intentamos predecir el contenido de grasa, que será nuestra respuesta y , utilizando como variables explicativas el contenido de agua, que indicaremos v , el de proteínas (z) y alguna característica de la curva espectrométrica que indicaremos por X .

Aneiros-Pérez y Vieu (2006) recomiendan usar un modelo donde (z, v) aparezca como covariable en la regresión lineal, mientras que la curva espectrométrica \mathcal{X} aparece modelada con una componente no paramétrica $y = \beta_{10} z + \beta_{20} v + \eta_0(\mathcal{X}) + \sigma_0 \epsilon$ y la cercanía entre curvas se mide a través de la distancia L^2 de las derivadas segundas.

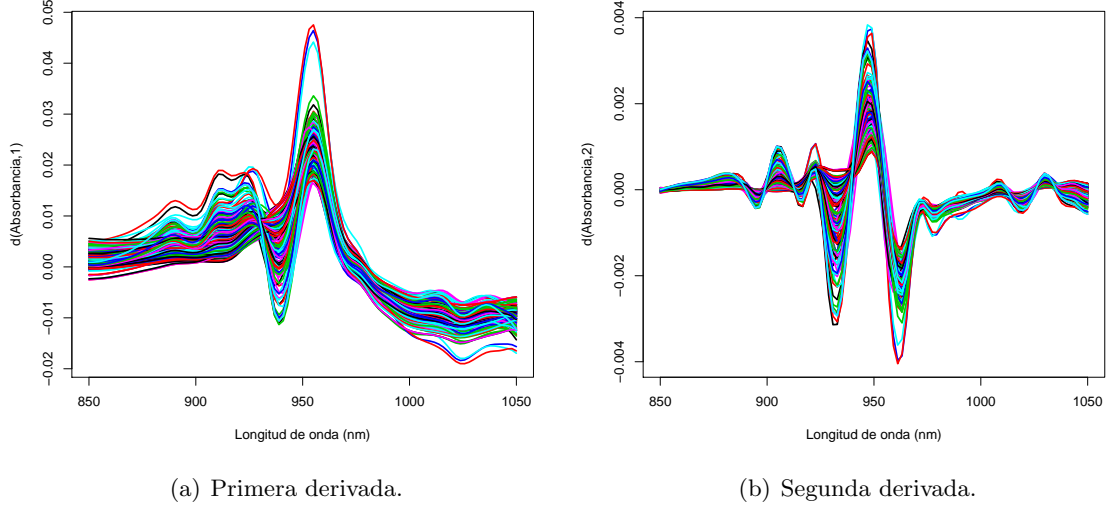


Figura 1.4: Conjunto de datos reales TECATOR: derivadas de la absorbancia \mathcal{X} .

Huang *et al.* (2015) estudiaron varios modelos donde X es la derivada segunda de la curva espectrométrica \mathcal{X} . En todos sus modelos, X aparece como covariable en una componente de regresión lineal funcional mientras que las componentes de (z, v) aparecen mediante un término no paramétrico aditivo $y = \langle X, \beta_0 \rangle + \eta_{0,1}(z) + \eta_{0,2}(v) + \sigma_0 \epsilon$, o bien, mediante un término de interacción entre contenido de los contenidos de proteína y agua $y = \gamma_0 + \langle X, \beta_0 \rangle + v \eta_0(z) + \sigma_0 \epsilon$.

Notch shape data

El conjunto de datos siguiente fue estudiado en Ramsay y Silverman (2002) quienes hicieron un análisis de componentes principales sobre el conjunto total de data y luego un análisis discriminante.

Los datos representan la forma de las rodillas en distintos individuos que son clasificados como *sanos* o como *enfermos* por sufrir una condición de artrosis. Para cada individuo se cuenta con información sobre la forma del *intercondylar notch* utilizada por los traumatólogos para proveer evidencia de daño causado por artritis(osteoarthritis).

Los datos consisten en $N = 96$ formas de la muesca (*notch*), sobre los que tenemos alguna información concomitante que da evidencia de daño producido por artritis. Por simplicidad, los índices 1 y 2 indicarán al grupo de personas *enfermas* y *sanas*, respectivamente. Del total de datos, $n_1 = 21$ femuers corresponden a individuos con artrosis y $n_2 = 75$ a individuos sin signos de cambios producidos por artrosis.

Los datos se parametrizan por la longitud de arco, de modo que el dato funcional corresponde



Figura 1.5: Forma de la rodilla.

a una función bidimensional $(X(t), Y(t))^T$, where $t \in (0, 1)$. La Figura 1.6 muestra, en el plano $X - Y$. Claramente, hay un dato en la población enferma (indicado en rojo en la Figura 1.7) que tiene un patrón distinto al resto mostrando una meseta en la parte central.

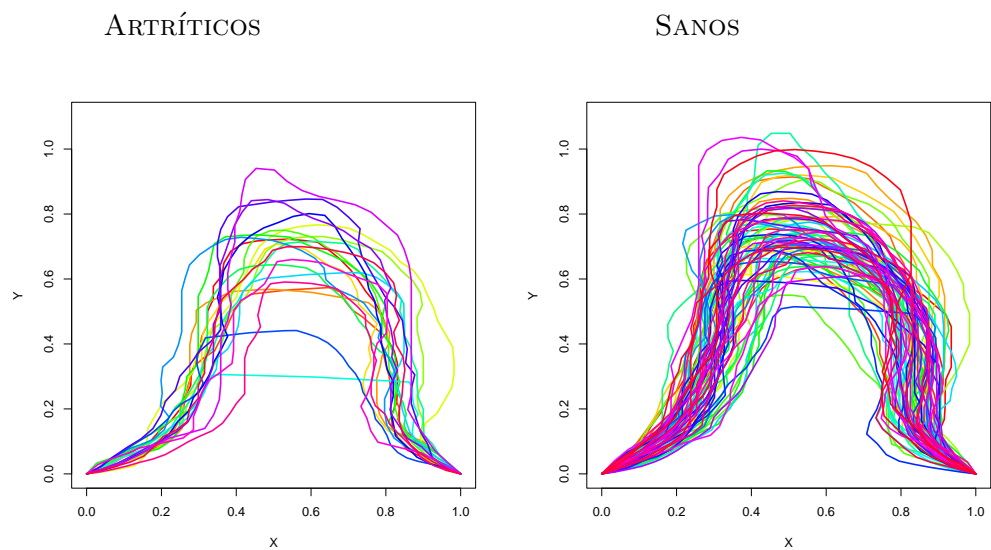


Figura 1.6: Forma de la muesca de 96 individuos.

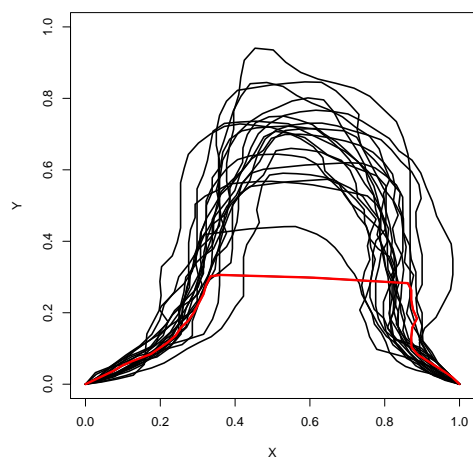


Figura 1.7: Forma de la muesca de 21 individuos *enfermos*.

Capítulo 2

Nociones previas

2.1 Análisis Funcional

Los resultados que se describen pueden verse en Conway (1990). En todo lo que sigue \mathbb{F} será el cuerpo de los números reales o complejos.

Definición 2.1. Diremos que un espacio normado \mathcal{X} es un espacio de Banach si resulta ser completo con la métrica inducida por la norma, es decir, un espacio de Banach es un espacio vectorial \mathcal{X} sobre \mathbb{F} con una norma $\|\cdot\|$ tal que toda sucesión de Cauchy (con respecto a la métrica $d(x, y) = \|x - y\|$) en \mathcal{X} tiene un límite en \mathcal{X} .

Definición 2.2. Si \mathcal{X} y \mathcal{Y} son espacios de Banach y $A : \mathcal{X} \rightarrow \mathcal{Y}$ es una transformación lineal, entonces son equivalentes:

- A es continua.
- A es continua en el 0.
- A es continua en algún punto.
- A está acotado, es decir, existe una constante positiva c tal que $\|Ax\| \leq c\|x\|$.

Definición 2.3. Dado \mathcal{X} , un funcional lineal acotado $f : \mathcal{X} \rightarrow \mathbb{F}$ es una aplicación lineal continua. El conjunto de funcionales lineales acotados de \mathcal{X} será notado como \mathcal{X}^* .

Teorema 2.1. (Teorema de Hahn-Banach) Sea \mathcal{X} un espacio normado y $M \subset \mathcal{X}$ un subespacio. Consideremos $f : M \rightarrow \mathbb{F}$ un funcional lineal y acotado, luego existe $F \in \mathcal{X}^*$ tal que $F|_M = f$ y $\|F\| = \|f\|$.

Corolario 2.2. Sea \mathcal{X} un espacio normado y $x \in \mathcal{X}$. Entonces,

$$\|x\| = \sup\{\|f(x)\| : f \in \mathcal{X}^*, \|f\| \leq 1\}.$$

Más aún, el supremo se alcanza.

Definición 2.4. Un espacio de Hilbert es un espacio vectorial \mathcal{H} sobre un cuerpo \mathbb{F} junto con un producto interno $\langle \cdot, \cdot \rangle$ de manera tal que la métrica inducida por el mismo es completa.

Definición 2.5. Si \mathcal{H} es un espacio de Hilbert, luego diremos que $f \in \mathcal{H}$ y $g \in \mathcal{H}$ son ortogonales si $\langle f, g \rangle = 0$.

Teorema 2.3. Sea M un subespacio cerrado de un espacio de Hilbert \mathcal{H} y sea $h \in \mathcal{H}$. Luego, existe un único elemento de M , que llamaremos $P_M h$ tal que $h - P_M h$ es ortogonal a todo M . Así definido, $P_M : \mathcal{H} \rightarrow \mathcal{H}$ resultará ser una aplicación lineal y continua, de normal acotada por 1 (y que vale 1 a menos que M sea trivial) y tal que es además un proyector ($P_M^2 = P_M$), con $\text{Ker}(P_M) = M^\perp$ e $\text{Im}(P_M) = M$.

Definición 2.6. Bajo las mismas condiciones que las del teorema anterior, P_M será la proyección ortogonal sobre el subespacio M .

Teorema 2.4. (Teorema de representación de Riesz) Sea $f \in \mathcal{H}^*$, luego existe un único vector $h_0 \in \mathcal{H}$ tal que $f(h) = \langle h, h_0 \rangle$. Más aún, $\|f\| = \|h_0\|$.

Definición 2.7. Un subconjunto ortonormal de un espacio de Hilbert \mathcal{H} es un subconjunto ε tal que

- $\forall e \in \varepsilon, \|e\| = 1$
- Si $e_1, e_2 \in \varepsilon$ y $e_1 \neq e_2$, luego $e_1 \perp e_2$.

Una base ortonormal será un conjunto ortonormal maximal.

Proposición 2.5. (Desigualdad de Bessel) Si $\{e_n : n \in \mathbb{N}\}$ es un conjunto ortonormal y $h \in \mathcal{H}$, luego

$$\sum_{i=1}^{\infty} |\langle h, e_n \rangle|^2 \leq \|h\|^2.$$

Teorema 2.6. Sea ε un conjunto ortonormal de \mathcal{H} . Son equivalentes

- ε es una base de \mathcal{H} .
- Si $h \in \mathcal{H}$ y $h \perp \varepsilon$, luego $h = 0$.

- Todo $h \in \mathcal{H}$ se escribe como $h = \sum_{e \in \varepsilon} \langle h, e \rangle h$.
- $\|h\|^2 = \sum_{e \in \varepsilon} |\langle h, e \rangle|^2$ (identidad de Parseval).

Proposición 2.7. Si \mathcal{H} es un espacio de Hilbert de dimensión infinita, luego \mathcal{H} es separable si y solamente si admite una base numerable.

Teorema 2.8. Sean $\mathcal{H}_1, \mathcal{H}_2$ espacios de Hilbert. Sea $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ lineal y continuo. Luego, existe un único $B : \mathcal{H}_2 \rightarrow \mathcal{H}_1$ tal que $\forall h_1 \in \mathcal{H}_1, h_2 \in \mathcal{H}_2, \langle Ah_1, h_2 \rangle = \langle h_1, Bh_2 \rangle$. Notaremos $A^* = B$ y diremos que es el adjunto de A .

Definición 2.8. Si $A : \mathcal{H} \rightarrow \mathcal{H}$ es un operador lineal y continuo, luego diremos que es autoadjunto si $A^* = A$.

Proposición 2.9. Si A es autoadjunto, luego vale que

$$\|A\| = \sup\{|\langle Ah, h \rangle| : \|h\| = 1\}.$$

Definición 2.9. A se dice una isometría si $\|Ah\| = \|h\| \forall h \in \mathcal{H}$. Si además A es sobreyectiva diremos que es unitario.

Proposición 2.10. Si $A : \mathcal{H} \rightarrow \mathcal{H}$ lineal y continuo, luego son equivalentes

- $A^*A = AA^* = Id$.
- A es unitario.

Definición 2.10. Una transformación lineal $T : \mathcal{H} \rightarrow \mathcal{H}$ es compacto si $T(B(0,1))$ tiene clausura compacta. Denotaremos por $\mathcal{B}_0(H)$ al conjunto de operadores compactos de \mathcal{H} en \mathcal{H} .

Definición 2.11. Sea $A : \mathcal{H} \rightarrow \mathcal{H}$ un operador lineal y continuo. Un escalar α se dice autovalor de A si $\ker(A - \alpha) \neq \{0\}$. Notaremos como $\sigma_p(A)$ al conjunto de autovalores de A .

Teorema 2.11. Sea T un operador compacto y autoadjunto en \mathcal{H} . Entonces vale que T tiene una cantidad numerable o finita de autovalores distintos. Si $\{\lambda_1, \lambda_2, \dots\}$ son los autovalores no nulos de T y P_n es la proyección ortogonal en el autoespacio $\ker(T - \lambda_n)$, entonces $P_n P_m = P_m P_n$ si $n \neq m$, los λ_n son reales y además

$$T = \sum_{n=1}^{\infty} \lambda_n P_n,$$

donde la serie converge a T con la norma de operadores.

Proposición 2.12. Si $T : \mathcal{H} \rightarrow \mathcal{H}$ compacto y M un subespacio invariante para T , luego $T|_M$ es compacto.

Definición 2.12. Un operador $T : \mathcal{H} \rightarrow \mathcal{H}$ se dice que es de Hilbert-Schmidt si es lineal, acotado y que tiene además una norma Hilbert-Schmidt finita. Es decir, existe una base ortonormal $\{e_i, i \in I\}$ de \mathcal{H} cumpliendo que

$$\sum_{i \in I} \|Te_i\|^2 < \infty.$$

Teorema 2.13. Teorema de Mercer. Sea $\mathcal{I} \subset \mathbb{R}$ un intervalo acotado y $\gamma(a, b) : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$ una función simétrica, definida no-negativa y continua. Entonces, existe una base ortonormal completa de funciones $\{\phi_j\}_{j \geq 1}$ en $L^2(\mathcal{I})$, y una sucesión no-creciente de números reales $\lambda_j \geq 0$, tales que

$$\gamma(a, b) = \sum_{j \geq 1} \lambda_j \phi_j(a) \phi_j(b), \quad a, b \in \mathcal{I}, \quad (2.1)$$

y la convergencia es absoluta y uniforme en $\mathcal{I} \times \mathcal{I}$. Más aún, si definimos el operador $\Gamma : L^2(\mathcal{I}) \rightarrow L^2(\mathcal{I})$ como

$$\Gamma f = \int \gamma(s, \cdot) f(s) ds,$$

entonces $\{\phi_j\}_{j \geq 1}$ son las autofunciones de Γ , o sea, $\Gamma \phi_j = \lambda_j \phi_j$, $j \geq 1$.

2.2 Teoría básica de datos funcionales

2.2.1 Elemento Aleatorio

La idea en este capítulo es extender las definiciones clásicas de variable aleatoria al contexto funcional, recurriendo en la medida de lo posible a las definiciones que ya se tienen para el caso real. En todo lo que sigue se supondrá, en un principio, que trabajaremos en un espacio de Banach y en particular, consideraremos después el caso de un espacio de Hilbert separable. El ejemplo más importante con el que trabajaremos será $\mathcal{H} = L^2[0, 1]$.

Definimos ahora la extensión de la noción de variable aleatoria para un espacio de Banach.

Definición 2.13. Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sea \mathcal{X} un espacio de Banach con su σ -álgebra de Borel $\mathcal{B}(\mathcal{X})$ y \mathcal{X}^* el dual topológico de \mathcal{X} . Sea $V : \Omega \rightarrow \mathcal{X}$. Diremos que V es un elemento aleatorio en sentido débil si para todo $f \in \mathcal{X}^*$, $f(V) : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria real.

Definición 2.14. Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad y sea \mathcal{X} un espacio de Banach con su σ -álgebra de Borel. Sea $V : \Omega \rightarrow \mathcal{X}$. Diremos que V es un elemento aleatorio en sentido fuerte si para todo conjunto A boreliano en \mathcal{X} vale que $V^{-1}(A) \in \mathcal{A}$.

Es claro que un elemento aleatorio en sentido fuerte lo será en sentido débil, los elementos del dual son continuos, por ende medibles, y la medibilidad se preserva bajo la composición. Pero se tiene además el siguiente teorema.

Teorema 2.14. *Si \mathcal{X} es separable, luego todo elemento aleatorio en sentido débil lo será en el sentido fuerte.*

Demostración. Sea V un elemento aleatorio en sentido débil, queremos ver que lo es en el sentido fuerte. Sea $\mathcal{B}(\mathcal{C})$ la σ -álgebra generada por los conjuntos \mathcal{C} (los cilindros unidimensionales) de la forma

$$\{x \in \mathcal{X}, f(x) \in B\}_{f \in \mathcal{X}^*, B \in \mathcal{B}(\mathbb{R})}$$

siendo $\mathcal{B}(\mathbb{R})$ la σ -álgebra de Borel en \mathbb{R} . O sea, estamos considerando conjuntos con una coordenada restringida a moverse en un boreliano real determinado y el resto libre. Vamos a probar que $\mathcal{B}(\mathcal{C}) = \mathcal{B}(\mathcal{X})$, la σ -álgebra de Borel en \mathcal{X} .

i) Demostremos primero $\mathcal{B}(\mathcal{C}) \subseteq \mathcal{B}(\mathcal{X})$.

Sea A un conjunto generador de $\mathcal{B}(\mathcal{C})$, o sea un cilindro unidimensional, determinado por un B en \mathbb{R} y la función del dual f , que al ser continua será medible. Notemos que $A = f^{-1}(B)$, como B es medible entonces la preimagen por una función medible también lo será. Por lo tanto, A será medible en $\mathcal{B}(\mathcal{X})$ y tenemos entonces esta primera inclusión.

ii) Veamos que $\mathcal{B}(\mathcal{X}) \subseteq \mathcal{B}(\mathcal{C})$.

Como el espacio es separable, consideremos $\{x_n, n \geq 1\}$ un denso numerable de \mathcal{X} . Asociado a cada x_n definimos un elemento del dual tal que $\|f_n\| = 1, f_n(x_n) = \|x_n\|$. Esto es posible en virtud al teorema de Hahn-Banach, primero se define cada f_n en x_n y luego se procede a efectuar una extensión continua de la misma utilizando ese resultado.

Sea $r > 0$, definamos

$$C_1 = \{\|x\| \leq r\}, C_2 = \bigcap_{i=1}^{\infty} \{x, f_n(x) \leq r\}.$$

Notamos que cada $\{x, f_n(x) \leq r\}$ es un conjunto cilíndrico finito dimensional, luego estarán en $\mathcal{B}(\mathcal{C})$, y como $\mathcal{B}(\mathcal{C})$ es cerrado por intersecciones numerables entonces C_2 estará en $\mathcal{B}(\mathcal{C})$. Además, como cada f_n tiene norma 1, entonces si x es tal que $\|x\| \leq r$, entonces $f_n(x) \leq r$, por lo tanto $C_1 \subseteq C_2$. Vamos a probar que ambos conjuntos son iguales. Equivalentemente, demostraremos que $C_1^c \subseteq C_2^c$. Sea $x \in C_1^c$. Luego, $\|x\| > r$, y como $\{x_n\}$ es denso en \mathcal{X} , existirá un x_k tal que $\|x - x_k\|(\|x\| - r)/2$. Por lo tanto,

$$\begin{aligned} \|x_k\| &= \|x - x + x_k\| \geq \|x\| - \|x - x_k\| > \|x\| - \frac{\|x\| - r}{2} = \frac{\|x\| + r}{2} \\ |f_k(x) - \|x_k\|| &= |f_k(x) - f_k(x_k)| = |f_k(x - x_k)| \leq \|x - x_k\| < \frac{\|x\| - r}{2}. \end{aligned}$$

De donde,

$$f_k(x) = \|x_k\| - (\|x_k\| - f_k(x)) \geq \|x_k\| - \frac{\|x\| - r}{2} > \frac{\|x\| + r}{2} - \frac{\|x\| - r}{2} = r,$$

con lo cual, $x \in C_2^c$. Luego, $C_1 = C_2$, y por lo tanto, $C_1 \in \mathcal{B}(\mathcal{C})$. Es decir que las bolas con centro en el origen pertenecen a $\mathcal{B}(\mathcal{C})$. Pero como $\mathcal{B}(\mathcal{C})$ es invariante por traslaciones, todas las bolas, con cualquier centro, pertenecen a $\mathcal{B}(\mathcal{C})$. Teniendo en cuenta que el espacio es separable, todo abierto se puede escribir como unión numerable de estas bolas con lo que entonces los abiertos, que son los generadores de $\mathcal{B}(\mathcal{X})$, están en $\mathcal{B}(\mathcal{C})$, de donde se concluye $\mathcal{B}(\mathcal{X}) \subseteq \mathcal{B}(\mathcal{C})$, que era lo que queríamos probar.

Veamos que V es medible si se toma como σ -álgebra $\mathcal{B}(\mathcal{C})$ en lugar de $\mathcal{B}(\mathcal{X})$. Dado un elemento A de $\mathcal{B}(\mathcal{C})$, deberíamos ver que $V^{-1}(A) \in \mathcal{A}$. Basta probarlo para un generador de $\mathcal{B}(\mathcal{C})$, es decir, si $f \in \mathcal{X}^*$ y $B \in \mathcal{B}(\mathbb{R})$ para el conjunto $A = \{x \in \mathcal{X}, f(x) \in B\} = f^{-1}(B)$. Pero $V \circ f$ es medible, por hipótesis, luego $V^{-1}(A) = V^{-1}(f^{-1}(B)) = (V \circ f)^{-1}(B)$ será un elemento de \mathcal{A} .

Luego, como $\mathcal{B}(\mathcal{C}) = \mathcal{B}(\mathcal{X})$, V será medible considerando $\mathcal{B}(\mathcal{X})$ y por lo tanto será un elemento aleatorio en sentido fuerte. ■

Será conveniente tener presente la σ -álgebra $\mathcal{B}(\mathcal{C})$, pues la volveremos a utilizar. Al ser equivalente a $\mathcal{B}(\mathcal{X})$ (bajo separabilidad) a veces será conveniente considerarla en lugar de $\mathcal{B}(\mathcal{X})$ puesto que permite, en algún sentido, trabajar directamente sobre \mathbb{R} , mediante la proyección.

Nuestro contexto de trabajo serán los espacios separables (de Banach o de Hilbert), por lo tanto diremos que V es un elemento aleatorio a secas, sin especificar el sentido puesto que ambos coinciden. Bajo separabilidad recuperaremos la propiedad de que la suma de elementos aleatorios es a su vez aleatorio.

Corolario 2.15. *Si \mathcal{X} es separable, luego la suma de dos elementos aleatorios es un elemento aleatorio.*

Demostración. Sean V_1 y V_2 dos elementos aleatorios, consideramos la suma $V_1 + V_2$. Si $f \in \mathcal{X}^*$, luego $f(V_1 + V_2) = f(V_1) + f(V_2)$, y cada sumando es una variable aleatoria real, luego la suma también lo será. Entonces, $f(V_1 + V_2)$ es una variable aleatoria real para todo elemento f en el dual, luego $V_1 + V_2$ es un elemento aleatorio. ■

2.2.2 Distribuciones de probabilidad e independencia

Precisamos definir primero un concepto auxiliar que se usará en esta sección.

Definición 2.15. *Una familia $U \subseteq \mathcal{B}(\mathcal{X})$ se dice una clase determinante de probabilidad si dadas P y Q probabilidades sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ se tiene*

$$P(D) = Q(D) \quad \forall D \in U \Rightarrow P = Q \quad \text{en } \mathcal{B}(\mathcal{X})$$

Notemos que bajo separabilidad $\mathcal{B}(\mathcal{C}) = \mathcal{B}(\mathcal{X})$, luego los cilindros unidimensionales que generan $\mathcal{B}(\mathcal{C})$ serán una clase determinante.

Definición 2.16. Dado V un elemento aleatorio en un espacio de Banach separable \mathcal{X} , se define la distribución de probabilidad de V , P_V , como una probabilidad sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ definida por $P_V(B) = \mathbb{P}(V^{-1}(B)) = \mathbb{P}(V \in B)$. O sea, $P_V = \mathbb{P} \circ V^{-1}$.

Dos elementos aleatorios V_1 y V_2 se dicen idénticamente distribuidos si $P_{V_1} = P_{V_2}$.

Por los enunciados anteriores, basta definir la distribución en una clase determinante (como los cilindros unidimensionales) para poder tenerla automáticamente definida sobre todo $\mathcal{B}(\mathcal{X})$.

Proposición 2.16. Sean V_1 y V_2 dos elementos aleatorios en \mathcal{X} idénticamente distribuidos, sea $\phi : \mathcal{X} \rightarrow \mathcal{Y}$ medible, con \mathcal{X} e \mathcal{Y} espacios de Banach. Luego, $\phi(V_1)$ y $\phi(V_2)$ serán elementos aleatorios en \mathcal{Y} idénticamente distribuidos.

Demostración. Sea $B \in \mathcal{B}(\mathcal{Y})$, luego $\mathbb{P}(\phi(V_1) \in B) = \mathbb{P}(V_1 \in \phi^{-1}(B)) = \mathbb{P}(V_2 \in \phi^{-1}(B)) = \mathbb{P}(\phi(V_2) \in B)$, lo que demuestra el resultado. ■

Proposición 2.17. Sea \mathcal{X} un espacio de Banach separable. Entonces los elementos aleatorios V_1 y V_2 en \mathcal{X} son idénticamente distribuidos si y sólo si $f(V_1)$ y $f(V_2)$ son variables aleatorias idénticamente distribuidas para cada $f \in \mathcal{X}^*$.

Demostración. Primero, supongamos que V_1 y V_2 son idénticamente distribuidas. Como $f \in \mathcal{X}^*$ es una función continua, es medible. Luego, por la Proposición 2.16, $f(V_1)$ y $f(V_2)$ serán idénticamente distribuidas.

Veamos la recíproca. Supongamos que $f(V_1)$ y $f(V_2)$ son variables aleatorias idénticamente distribuidas para cada $f \in \mathcal{X}^*$. Es decir, $\mathbb{P} \circ V_1^{-1} \circ f^{-1} = \mathbb{P} \circ V_2^{-1} \circ f^{-1}$, $\forall f \in \mathcal{X}^*$. Nos basta ver que $P_{V_1} = P_{V_2}$ en una clase determinante, como pueden ser los cilindros unidimensionales que generan $\mathcal{B}(\mathcal{C})$. Sea $B \in \mathcal{B}(\mathbb{R})$ y $D = \{x \in \mathcal{X} : f(x) \in B\} = f^{-1}(B)$. Luego,

$$P_{V_1}(D) = P_{V_1}(f^{-1}(B)) = \mathbb{P}(V_1^{-1} \circ f^{-1}(B)) = \mathbb{P}(V_2^{-1} \circ f^{-1}(B)) = P_{V_2}(f^{-1}(B)) = P_{V_2}(D)$$

lo que concluye la demostración. ■

Definamos ahora la noción de independencia.

Definición 2.17. Una familia $\{V_t : \Omega \rightarrow \mathcal{X}_t, t \in A\}$ de elementos aleatorios se dicen independientes si $\forall F \subseteq A$ finito, $F = \{t_1, \dots, t_k\}$, vale que

$$\mathbb{P} \left(\bigcap_{t \in F} \{V_t \in E_t\} \right) = \prod_{t \in F} \mathbb{P}(V_t \in E_t), \quad \forall E_{t_1} \in \mathcal{B}(\mathcal{X}_{t_1}), \dots, E_{t_r} \in \mathcal{B}(\mathcal{X}_{t_r}).$$

Tenemos resultados análogos a los de idéntica distribución.

Proposición 2.18. Sean V_1 y V_2 dos elementos aleatorios en \mathcal{X}_1 y \mathcal{X}_2 respectivamente, ambos independientes. Sean $\phi_1 : \mathcal{X}_1 \rightarrow \mathcal{Y}_1$, $\phi_2 : \mathcal{X}_2 \rightarrow \mathcal{Y}_2$ medibles. \mathcal{X}_1 , \mathcal{X}_2 , \mathcal{Y}_1 y \mathcal{Y}_2 espacios de Banach. Luego, $\phi_1(V_1)$ y $\phi_2(V_2)$ serán elementos aleatorios independientes.

Demostración. Sea B_1 en $\mathcal{B}(\mathcal{Y}_1)$ y B_2 en $\mathcal{B}(\mathcal{Y}_2)$, se cumple

$$\mathbb{P}(\phi_1(V_1) \in B_1, \phi_2(V_2) \in B_2) = \mathbb{P}\left(\{V_1 \in \phi_1^{-1}(B_1)\} \cap \{V_2 \in \phi_2^{-1}(B_2)\}\right).$$

Por la independencia de V_1 y V_2 , el miembro derecho de la igualdad es igual a

$$\mathbb{P}(\{V_1 \in \phi_1^{-1}(B_1)\}) \mathbb{P}(\{V_2 \in \phi_2^{-1}(B_2)\}) = \mathbb{P}(\phi_1(V_1) \in B_1) \mathbb{P}(\phi_2(V_2) \in B_2).$$

lo que concluye la demostración. ■

Enunciaremos un último resultado de independencia que se terminará de demostrar después, una vez se dispongan de los instrumentos necesarios de funcionales característicos que se definirán en la próxima sección.

Proposición 2.19. Sean \mathcal{X}_1 , \mathcal{X}_2 espacios de Banach separables. Entonces los elementos aleatorios V_1 y V_2 en \mathcal{X}_1 y \mathcal{X}_2 son independientes si y sólo si $f(V_1)$ y $g(V_2)$ son variables aleatorias independientes para toda $f \in \mathcal{X}_1^*$, $g \in \mathcal{X}_2^*$.

Demostración. Primero, supongamos que V_1 y V_2 son independientes. Como f y g son funciones continuas, serán medibles y por lo tanto, por la Proposición 2.18, $g(V_1)$ y $g(V_2)$ serán independientes.

La recíproca la probaremos más adelante después de introducir la noción de funcionales característicos. ■

2.2.3 Esperanza

Extenderemos la noción de esperanza al caso funcional.

Definición 2.18. Dado \mathcal{X} un espacio de Banach separable y V un elemento aleatorio, diremos que V tiene esperanza $\mathbb{E}(V) \in \mathcal{X}$ si

1. $\mathbb{E}(|f(V)|) < \infty \forall f \in \mathcal{X}^*$
2. $\mathbb{E}(V)$ es el elemento de \mathcal{X} que verifica que $f(\mathbb{E}(V)) = \mathbb{E}(f(V)) \forall f \in \mathcal{X}^*$

Es fácil ver que si existe el valor esperado de una variable aleatoria entonces es único. En efecto, supongamos que E_1 y E_2 son ambos valores esperados del elemento aleatorio V . Entonces,

tendremos que $f(E_1) = \mathbb{E}(f(V)) = f(E_2)$, para todo $f \in \mathcal{X}^*$. Luego $f(E_1) = f(E_2)$ para todo elemento del dual, o sea $f(E_1 - E_2) = 0$ para todo $f \in \mathcal{X}^*$. Pero, por el Corolario 2.2 tenemos

$$\|E_1 - E_2\| = \sup\{|f(E_1 - E_2)|, f \in \mathcal{X}^*, \|f\|_{\mathcal{X}^*} = 1\} = 0$$

con lo cual se deduce la igualdad.

La idea de esta definición es la de extender la noción de vector de esperanzas del caso multivariado al caso funcional. Pretenderemos entonces una extensión de los resultados que ya se tienen para ese caso, por ejemplo, el de linealidad de la esperanza y en particular, la propiedad $E(AX) = AE(X)$ con A un operador lineal y acotado. Efectivamente, la proposición siguiente muestra que este último resultado se puede generalizar.

Proposición 2.20. *Sean \mathcal{X}, \mathcal{Y} espacios de Banach separables y $h : \mathcal{X} \rightarrow \mathcal{Y}$ un operador lineal acotado. Sea además V un elemento aleatorio en \mathcal{X} con valor esperado $\mathbb{E}(V)$. Entonces, vale que $\mathbb{E}(h(V))$ está bien definido y además $\mathbb{E}(h(V)) = h(\mathbb{E}(V))$.*

Demostración. Debemos ver que se verifican las dos condiciones de un valor esperado. Primero, debemos ver que $\mathbb{E}(|f(h(V))|) < \infty$ para todo $f \in \mathcal{Y}^*$. Esto vale puesto que $f \circ h = g \in \mathcal{X}^*$ y por hipótesis $\mathbb{E}(V)$ existe luego vale que $\mathbb{E}(|f(h(V))|) = \mathbb{E}(|g(V)|) < \infty$.

Veamos ahora el segundo requerimiento, la conmutatividad con todo funcional. Hay que ver que $f(h(\mathbb{E}(V))) = \mathbb{E}(f(h(V)))$ para todo $f \in \mathcal{Y}^*$.

Observemos que $\mathbb{E}(f(h(V))) = \mathbb{E}(g(V))$ con $g = f \circ h \in \mathcal{X}^*$ y por la definición de $\mathbb{E}(V)$ esto último será $g(\mathbb{E}(V)) = f(h(\mathbb{E}(V)))$, lo que concluye la demostración. ■

Así como se tiene la esperanza, es posible también definir la varianza de un elemento aleatorio.

Definición 2.19. *Sea \mathcal{X} un espacio de Banach separable y V un elemento aleatorio en \mathcal{X} para el que exista $\mathbb{E}(V)$. Se define la varianza de V como*

$$Var(V) = \mathbb{E}(\|V - \mathbb{E}(V)\|^2) = \int_{\Omega} \|V - \mathbb{E}(V)\|^2 d\mathbb{P},$$

siempre que $Var(V) < \infty$.

2.2.4 Operador de Covarianza

Nos centraremos en lo que sigue en \mathcal{H} un espacio de Hilbert separable, pues precisaremos la estructura ofrecida por el producto interno. Supondremos además que el elemento aleatorio tiene segundo momento finito, cosa de tener bien definida la noción de varianza. Vamos a querer extender la idea de matriz de covarianza para el caso funcional. En realidad, nos interesa el caso de elementos aleatorios que toman valores en un espacio de Hilbert separable, siendo el caso particular y el que se estudiará con más profundidad $\mathcal{H} = L^2[0, 1]$.

Definición 2.20. Sea \mathcal{H} un espacio de Hilbert separable con producto interno $\langle \cdot, \cdot \rangle$ y sea V un elemento aleatorio definido en \mathcal{H} con momento de orden 2 finito. Definimos $a_V : \mathcal{H} \times \mathcal{H} \rightarrow \mathbb{R}$ el operador bilineal de covarianza puntual como

$$a_V(u, v) = \text{Cov}(\langle u, V \rangle, \langle v, V \rangle).$$

Es claro que a_V es bilineal, lo heredar  de la covarianza real. La continuidad se deduce de aplicar la desigualdad de Cauchy-Schwartz y de usar que V tiene segundo momento finito. En efecto,

$$\begin{aligned} |a_V(u, v)| &= |\text{Cov}(\langle u, V \rangle, \langle v, V \rangle)| \leq \sqrt{\text{Var}(\langle u, V \rangle)} \sqrt{\text{Var}(\langle v, V \rangle)} \\ &\leq \sqrt{\mathbb{E}(\langle u, V \rangle)^2} \sqrt{\mathbb{E}(\langle v, V \rangle)^2} \leq \|u\| \|v\| \mathbb{E}\|V\|^2 \end{aligned}$$

Debido a que $a_V(u, v)$ es bilineal podremos entonces, aplicando el Teorema de Representaci n de Riesz y la continuidad del operador bilineal, definir  nicos operadores A y B acotados tales que $a_V(u, v) = \langle Au, v \rangle_{\mathcal{H}} = \langle u, Bv \rangle_{\mathcal{H}}$ para todo $u, v \in \mathcal{H}$.

Definici n 2.21. Sea a_V un operador de covarianza puntual y sea A tal que $a_V(u, v) = \langle Au, v \rangle \forall u, v \in \mathcal{H}$. Diremos entonces que A es el operador de covarianza de V y lo notaremos como Γ_V .

Proposici n 2.21. El operador Γ_V es autoadjunto (sim trico).

Demostraci n. Se deduce inmediatamente de la simetr a del operador bilineal a_V . $\langle \Gamma_V u, v \rangle = a_V(u, v) = \text{Cov}(\langle u, V \rangle, \langle v, V \rangle) = \text{Cov}(\langle v, V \rangle, \langle u, V \rangle) = \langle \Gamma_V v, u \rangle = \langle u, \Gamma_V v \rangle$. ■

Proposici n 2.22. El operador Γ_V es semidefinido positivo.

Demostraci n. Basta observar que $\langle \Gamma_V u, u \rangle = \text{Cov}(\langle u, V \rangle, \langle u, V \rangle) = \text{Var}(\langle u, V \rangle) \geq 0$. ■

An logamente podemos definir un operador de covarianza cruzada, requeriremos ah  dos elementos aleatorios U y V no necesariamente en el mismo espacio de Hilbert.

Definici n 2.22. Sean \mathcal{H}_1 y \mathcal{H}_2 espacios de Hilbert separables con productos internos $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ y $\langle \cdot, \cdot \rangle_{\mathcal{H}_2}$, respectivamente. Sean U y V dos elementos aleatorios definidos en \mathcal{H}_1 y \mathcal{H}_2 , respectivamente con momento de orden 2 finito. Definimos $a_{UV} : \mathcal{H}_1 \times \mathcal{H}_2 \rightarrow \mathbb{R}$ el operador bilineal de covarianza cruzada puntual como

$$a_{UV}(u, v) = \text{Cov}(\langle u, U \rangle_{\mathcal{H}_1}, \langle v, V \rangle_{\mathcal{H}_2}).$$

Igual que antes, el operador de covarianza cruzada ser  el operador $\Gamma_{UV} : \mathcal{H}_2 \rightarrow \mathcal{H}_1$ tal que

$$\langle u, \Gamma_{UV} v \rangle_{\mathcal{H}_1} = a_{UV}(u, v).$$

Un resultado muy importante en an lisis multivariado es el que permite obtener la matriz de covarianza de un vector aleatorio compuesto con una transformaci n lineal, obteniendo lo siguiente: $\Sigma_{AX} = A \Sigma_X A^T$. En este contexto, se tiene un resultado semejante.

Proposición 2.23. Sean \mathcal{H}_1 y \mathcal{H}_2 espacios de Hilbert separables. Sea V un elemento aleatorio en \mathcal{H}_1 , con segundo momento finito. Sea $A : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ un operador acotado. Se cumple que $\Gamma_{AV} : \mathcal{H}_2 \rightarrow \mathcal{H}_2$ es igual a $\Gamma_{AV} = A\Gamma_V A^*$, siendo $A^* : \mathcal{H}_2 \rightarrow \mathcal{H}_1$ el adjunto de A .

Demostración. Queremos ver que

$$\langle \Gamma_{AV} u, v \rangle_{\mathcal{H}_2} = \langle A\Gamma_V A^* u, v \rangle_{\mathcal{H}_2}$$

para todo $u \in \mathcal{H}_1$, $v \in \mathcal{H}_2$. Ahora bien, tenemos que

$$\begin{aligned} \langle A\Gamma_V A^* u, v \rangle_{\mathcal{H}_2} &= \langle \Gamma_V A^* u, A^* v \rangle_{\mathcal{H}_2} = \text{Cov}(\langle A^* u, V \rangle_{\mathcal{H}_2}, \langle A^* v, V \rangle_{\mathcal{H}_2}) \\ &= \text{Cov}(\langle u, AV \rangle_{\mathcal{H}_1}, \langle v, AV \rangle_{\mathcal{H}_1}) = a_{AV}(u, v) = \langle \Gamma_{AV} u, v \rangle_{\mathcal{H}_2}, \end{aligned}$$

lo que concluye la demostración. ■

Tenemos un resultado análogo para el operador de covarianza cruzada.

Proposición 2.24. Sean \mathcal{H}_1 , \mathcal{H}_2 , \mathcal{H}_3 y \mathcal{H}_4 espacios de Hilbert separables. Sean V y W elementos aleatorios en \mathcal{H}_1 y \mathcal{H}_2 respectivamente, espacios de Hilbert, con segundo momento finito. Sea $A : \mathcal{H}_1 \rightarrow \mathcal{H}_3$ y sea $B : \mathcal{H}_2 \rightarrow \mathcal{H}_4$ operadores acotados. Se cumple que

- i) $\Gamma_{AV,W} = A\Gamma_{V,W}$
- ii) $\Gamma_{V,BW} = \Gamma_{V,W}B^*$
- iii) $\Gamma_{AV,BW} = A\Gamma_{V,W}B^*$

Es posible relacionar el operador de covarianza con la varianza, definida anteriormente. En efecto, tenemos el siguiente resultado.

Lema 2.25. Sea \mathcal{H} un espacio de Hilbert separable y $\{e_n\}_{n \in \mathbb{N}}$ una base ortonormal de \mathcal{H} . Sea V un elemento aleatorio con varianza finita. Luego $\text{Var}(V) = \sum_{n=1}^{\infty} \langle \Gamma_V e_n, e_n \rangle$.

Demostración. Observemos que

$$\begin{aligned} \sum_{n=1}^{\infty} \langle \Gamma_V e_n, e_n \rangle &= \sum_{n=1}^{\infty} a_V(e_n, e_n) = \sum_{n=1}^{\infty} \text{Cov}(\langle V, e_n \rangle, \langle V, e_n \rangle) \\ &= \sum_{n=1}^{\infty} \text{Var}(\langle V, e_n \rangle) = \sum_{n=1}^{\infty} \mathbb{E}(\langle V, e_n \rangle - \mathbb{E}(\langle V, e_n \rangle))^2. \end{aligned}$$

Como $\langle \cdot, e_n \rangle$ es lineal y continuo, es posible permutarlo con la esperanza, es decir, $\mathbb{E}\langle V, e_n \rangle = \langle \mu, e_n \rangle$, donde $\mu = \mathbb{E}(V)$ por simplicidad. Luego, obtenemos

$$\begin{aligned}
\sum_{n=1}^{\infty} \langle \Gamma_V e_n, e_n \rangle &= \sum_{n=1}^{\infty} \mathbb{E} (\langle V, e_n \rangle - \langle \mathbb{E}(V), e_n \rangle)^2 \\
&= \sum_{n=1}^{\infty} \mathbb{E} (\langle V, e_n \rangle - \langle \mu, e_n \rangle)^2 \\
&= \sum_{n=1}^{\infty} \mathbb{E} (\langle V - \mu, e_n \rangle)^2 = \mathbb{E} \left\{ \sum_{i=1}^{\infty} (\langle V - \mu, e_n \rangle)^2 \right\} \\
&= \mathbb{E}(\|V - \mu\|^2) = \text{Var}(V)
\end{aligned}$$

donde la última igualdad, se deduce de aplicar la identidad de Parseval y del hecho que $\{e_n\}_{n \in \mathbb{N}}$ es una base ortonormal de \mathcal{H} . ■

2.2.5 Funcionales Característicos

Definición 2.23. Sea \mathcal{X} un espacio de Banach separable y P una probabilidad sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$. El funcional característico de P se define como una función $\hat{P} : \mathcal{X}^* \rightarrow \mathbb{R}$ tal que

$$\hat{P}(f) = \int_{\mathcal{X}} e^{if(x)} dP(x).$$

Si V es un elemento aleatorio en \mathcal{X} y P_V es su probabilidad inducida sobre $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, el funcional característico de V se define como aquel asociado a P_V , o sea, como una función $\hat{P}_V : \mathcal{X}^* \rightarrow \mathbb{R}$ tal que

$$\hat{P}_V(f) = \int_{\mathcal{X}} e^{if(x)} dP_V(x) = \mathbb{E}(e^{if(V)}).$$

Alternativamente, utilizaremos φ_V para denotar el funcional característico, en especial cuando consideremos el caso multivariado $\mathcal{X} = \mathbb{R}^d$.

Proposición 2.26. Valen las siguientes propiedades

1. $\hat{P}(0) = 1$.
2. $|\hat{P}(f)| \leq 1, \forall f \in \mathcal{X}^*$.
3. $\hat{P}(f) = \overline{\hat{P}(-f)}$, (conjugación compleja).
4. Para cada conjunto finito de números complejos z_1, \dots, z_N y para todo $f_1, \dots, f_N \in \mathcal{X}^*$ se cumple

$$\sum_{j=1}^N \sum_{k=1}^N z_j \overline{z_k} \hat{P}(f_j - f_k) \geq 0.$$

5. Sean V y W elementos aleatorios de un mismo espacio \mathcal{X} independientes y sea $Z = V + W$. Luego $\widehat{P}_Z(f) = \widehat{P}_V(f) \cdot \widehat{P}_W(f)$, $\forall f \in \mathcal{X}^*$.

La demostración de todas estas propiedades es inmediata y se omite.

Proposición 2.27. \widehat{P} es uniformemente continua en \mathcal{X}^* .

Demostración. Sean f_1, f_2 elementos de \mathcal{X}^* . Luego,

$$\widehat{P}(f_1) - \widehat{P}(f_2) = \int_{\mathcal{X}} \left[e^{if_1(x)} - e^{if_2(x)} \right] dP(x).$$

Teniendo en cuenta, la siguiente igualdad trigonométrica $|1 - e^{ia}| = (1 - \cos(a))^2 + \sin^2(a) = 2(1 - \cos(a)) = 4 \sin^2(a/2)$, se deduce

$$\left| \widehat{P}(f_1) - \widehat{P}(f_2) \right| \leq \int_{\mathcal{X}} \left| 1 - e^{i(f_1 - f_2)(x)} \right| dP(x) = 4 \int_{\mathcal{X}} \left| \sin^2 \left(\frac{(f_1 - f_2)(x)}{2} \right) \right| dP(x).$$

Sea $\epsilon > 0$ arbitrario y $\nu = \nu(\epsilon)$ tal que

$$P(\{x \in \mathcal{X} : \|x\| \geq \nu\}) \leq \epsilon/8.$$

Sea $\delta = \epsilon/(4\nu)$. Luego, dadas $f_1, f_2 \in \mathcal{X}^*$ tales que $\|f_1 - f_2\| < \delta$, tenemos que

$$\left| \widehat{P}(f_1) - \widehat{P}(f_2) \right| \leq 4 \int_{\|x\| \geq \nu} \left| \sin^2 \left(\frac{(f_1 - f_2)(x)}{2} \right) \right| dP(x) + 4 \int_{\|x\| < \nu} \left| \sin^2 \left(\frac{(f_1 - f_2)(x)}{2} \right) \right| dP(x).$$

El integrando del primer sumando se puede acotar por 1, y luego toda la integral quedará la probabilidad del dominio de integración, que es menor que ϵ . El integrando del segundo sumando se acota directamente por el argumento del seno, con lo cual

$$\left| \widehat{P}(f_1) - \widehat{P}(f_2) \right| \leq \frac{\epsilon}{2} + 4 \int_{\|x\| < \nu} \left| \frac{(f_1 - f_2)(x)}{2} \right| dP(x) \leq \frac{\epsilon}{2} + 2\|f_1 - f_2\|\nu.$$

Como $\|f_1 - f_2\| < \delta = \epsilon/(4\nu)$ se obtiene $|\widehat{P}(f_1) - \widehat{P}(f_2)| < \epsilon$, lo que concluye la demostración. ■

Proposición 2.28. $\widehat{P}_1 = \widehat{P}_2$ si y sólo si $P_1 = P_2$.

Demostración. Es claro que si $P_1 = P_2$ entonces $\widehat{P}_1 = \widehat{P}_2$. Bastará probar la otra implicación. Supongamos $\widehat{P}_1 = \widehat{P}_2$. Por la unicidad de la transformada de Fourier en \mathbb{R} tendremos que $P_1 = P_2$ en el álgebra de cilindros unidimensionales que es una clase determinante, luego coincidirán en todo la σ -álgebra $\mathcal{B}(\mathcal{X})$. ■

Vamos a probar ahora el resultado que había quedado pendiente de la Sección 2.2.2, la Proposición 2.19, es decir, queremos mostrar

Proposición 2.29. Sean $\mathcal{X}_1, \mathcal{X}_2$ espacios de Banach separables. Si $f(V_1)$ y $g(V_2)$ son variables aleatorias independientes para toda $f \in \mathcal{X}_1^*, g \in \mathcal{X}_2^*$, entonces los elementos aleatorios V_1 y V_2 en \mathcal{X}_1 y \mathcal{X}_2 son independientes.

Demostración. Tomemos $W = (V_1, V_2)$ que es un elemento aleatorio en $\mathcal{X}_1 \times \mathcal{X}_2$ con la σ -álgebra producto y con función de distribución P_W inducida por la medida producto de P_{V_1} y P_{V_2} , es decir, la medida de probabilidad que se obtendría de suponer la independencia entre V_1 y V_2 . Consideramos $V = (V_1, V_2)$, también en $\mathcal{X}_1 \times \mathcal{X}_2$ y con la misma σ -álgebra, pero ahora la medida P_V será la que se infiera “genuinamente” de la distribución conjunta de V_1 y V_2 , tal como se definió en la Sección 2.2.2.

Queremos ver que ambas medidas coinciden, para ello, bastará probar que ambas medidas inducen la misma función característica.

Llamemos \hat{P}_W la función característica de P_W y \hat{P}_V a la de P_V .

Observemos en primer lugar que hay una relación biunívoca entre los elementos de $(\mathcal{X}_1 \times \mathcal{X}_2)^*$ y los de $\mathcal{X}_1^* \times \mathcal{X}_2^*$. En efecto, dada $f \in (\mathcal{X}_1 \times \mathcal{X}_2)^*$ podemos definir las proyecciones $f_1 \in \mathcal{X}_1^*$ como $f_1(x_1) = f(x_1, 0)$ y $f_2 \in \mathcal{X}_2^*$ como $f_2(x_2) = f(0, x_2)$. Resultarán ambas continuas pues f lo era. Análogamente, si tenemos $f_1 \in \mathcal{X}_1^*$ y $f_2 \in \mathcal{X}_2^*$ podemos definir $f \in (\mathcal{X}_1 \times \mathcal{X}_2)^*$ como $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$.

Entonces, se cumple

$$\hat{P}_V(f) = \mathbb{E} \left(e^{if(W)} \right) = \mathbb{E} \left(e^{i(f_1(V_1) + f_2(V_2))} \right) = \mathbb{E} \left(e^{i f_1(V_1)} e^{i f_2(V_2)} \right).$$

Ahora bien, $f_1(V_1)$ y $f_2(V_2)$ son variables aleatorias reales que son independientes por hipótesis, luego

$$\hat{P}_V(f) = \mathbb{E} \left(e^{i f_1(V_1)} e^{i f_2(V_2)} \right) = \mathbb{E} \left(e^{i f_1(V_1)} \right) \mathbb{E} \left(e^{i f_2(V_2)} \right) = \hat{P}_{V_1}(f_1) \hat{P}_{V_2}(f_2) = \hat{P}_W(f_1 \times f_2).$$

Por lo tanto, coinciden ambas funciones características de donde se deducen que las medidas de probabilidad son iguales, es decir, la medida P_V coincide con la medida producto entre V_1 y V_2 que corresponde a la medida conjunta de ambos elementos aleatorios bajo la independencia. Por lo tanto, V_1 y V_2 son independientes. ■

2.2.6 Caso $\mathcal{H} = L^2[0, 1]$

En el caso $\mathcal{H} = L^2[0, 1]$ estaremos con un espacio separable, la base de Fourier tiene cardinal numerable y por lo tanto, al tener una base ortonormal numerable, el espacio de Hilbert \mathcal{H} resulta separable.

En este caso, se suele definir el operador de covarianza como el operador $\Gamma_V : \mathcal{H} \rightarrow \mathcal{H}$ tal que evaluado en toda función $f \in L^2[0, 1]$ devuelve un elemento de $L^2[0, 1]$ tal que

$$(\Gamma_V f)(s) = \int_0^1 \text{Cov}(V(s), V(t)) f(t) dt,$$

y a la cantidad $\gamma_V(s, t) = \text{Cov}(V(s), V(t))$ se la llama núcleo de covarianza.

Vamos a ver que esta definición coincide con introducida anteriormente.

Por la unicidad del operador caracterizado por la forma bilineal $a_V(\cdot, \cdot)$, nos basta ver que

$$\langle \Gamma_V f, g \rangle = a_V(f, g) = \text{Cov}(\langle f, V \rangle, \langle g, V \rangle). \quad (2.2)$$

Desarrollemos el término de la izquierda,

$$\langle \Gamma_V f, g \rangle = \int_0^1 (\Gamma_V f)(s) g(s) ds = \int_0^1 g(s) \left(\int_0^1 \text{Cov}(V(s), V(t)) f(t) dt \right) ds. \quad (2.3)$$

Por otra parte, si $E_f = \mathbb{E} \left(\int_0^1 f(s) V(s) ds \right)$, el término derecho de (2.2) resulta igual a

$$\text{Cov} \left(\int_0^1 f(s) V(s) ds, \int_0^1 g(s) V(s) ds \right) = E \left[\left(\int_0^1 f(s) V(s) ds - E_f \right) \left(\int_0^1 g(s) V(s) ds - E_g \right) \right] \quad (2.4)$$

Ahora bien, tenemos que $E_f = \mathbb{E} \langle f, V \rangle$. Como, fijado f , $\langle f, \cdot \rangle$ es un funcional lineal en \mathcal{H} , podemos conmutar con la esperanza, por su definición en el caso funcional. Obtenemos entonces $E_f = \langle f, \mathbb{E}(V) \rangle = \langle f, \mu \rangle$, siendo $\mu \in \mathcal{H}$ la esperanza de V (observemos que $\mu(s) = \mathbb{E}V(s)$). Reemplazando en (2.4) obtenemos

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^1 f(s) V(s) ds - E_f \right) \left(\int_0^1 g(s) V(s) ds - E_g \right) \right] &= \mathbb{E} ((\langle f, V \rangle - E_f) (\langle g, V \rangle - E_g)) \\ &= \mathbb{E} (\langle f, V - \mu \rangle \langle g, V - \mu \rangle) = \int_{\Omega} \langle f, V - \mu \rangle \langle g, V - \mu \rangle d\mathbb{P} \\ &= \int_{\Omega} d\mathbb{P} \int_0^1 f(s) (V(s) - \mu(s)) ds \int_0^1 g(t) (V(t) - \mu(t)) dt \\ &= \int_{\Omega} d\mathbb{P} \int_0^1 f(s) (V(s) - \mu(s)) ds \int_0^1 g(t) (V(t) - \mu(t)) dt \\ &= \int_0^1 \int_0^1 f(s) g(t) \int_{\Omega} (V(s) - \mu(s)) (V(t) - \mu(t)) d\mathbb{P} ds dt = \int_0^1 \int_0^1 f(s) g(t) \text{Cov}(V(s), V(t)) ds dt \end{aligned}$$

y esto es el término derecho de (2.3) teniendo así lo deseado.

2.2.7 Esperanza y Distribución Condicional

Definiremos ahora distribuciones condicionales. Puesto que no tendremos en principio la noción de probabilidad puntual o de función de densidad, será necesario tratarlo desde un ámbito más general. Empezaremos considerando la esperanza condicional y a partir de la misma efectuamos la construcción clásica de probabilidad condicional con el uso de la función indicadora.

Definición 2.24. Sea $(\Omega, \mathcal{A}, \mathbb{P})$ un espacio de probabilidad, \mathcal{H}_j , $j = 1, 2$, dos espacios de Hilbert con producto interno $\langle h(X), Y - g(X) \rangle_{\mathcal{H}_j}$. Dados $X : \Omega \rightarrow \mathcal{H}_1$ e $Y : \Omega \rightarrow \mathcal{H}_2$, dos elementos aleatorios de $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$, definimos $g(X) = \mathbb{E}(Y|X)$ como la única función $g : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ medible respecto de $\mathcal{B}(\mathcal{H}_1)$ tal que

$$\mathbb{E}(\langle h(X), Y - g(X) \rangle_{\mathcal{H}_2}) = 0,$$

para toda $h : \mathcal{H}_1 \rightarrow \mathcal{H}_2$ medible.

En particular, se cumple que

$$\mathbb{E}(\langle g(X), Y - g(X) \rangle_{\mathcal{H}_2}) = 0.$$

Por otra parte, dado $v \in \mathcal{H}_2$ fijo, tomando $h \equiv v$ obtenemos

$$0 = \mathbb{E}(\langle v, Y - g(X) \rangle_{\mathcal{H}_2}) = \langle v, \mathbb{E}(Y - g(X)) \rangle_{\mathcal{H}_2},$$

para todo $v \in \mathcal{H}_2$. Utilizando la Proposición 2.20, deducimos que

$$\mathbb{E}(Y) = \mathbb{E}(g(X))$$

La existencia de la esperanza condicional se puede demostrar utilizando el teorema de Radon-Nikodym. Por otra parte, g es la esperanza condicional ($g(X) = \mathbb{E}(Y|X)$) si y sólo si para toda $f \in \mathcal{H}_2^*$,

$$f(g(X)) = \mathbb{E}\left(f(Y) \middle| X\right).$$

A partir de la definición de esperanza condicional, podemos construir la medida condicional como

$$\mathbb{P}_{Y|X=x}(A) = \mathbb{P}(Y \in A | X = x) = \mathbb{E}(\mathbb{I}_Y(A) | X = x)$$

Bibliografía

- [1] Aneiros-Pérez, G. and Vieu, P. (2006). Semi-functional partial linear model. *Statistics & Probability Letters*, **11**, 1102-1110.
- [2] Bali, and Boente, G. (2014). Robust Functional Principal Components. *New advances in statistical modelling and applications*. Antonio Pacheco, Rui Santos, Maria do Rosario Oliveira and Carlos Daniel Paulino (Ed.). Springer. pp. 41-54. ISBN: 978-3-319-05322-6.
- [3] Conway J.B. (2000). A Course in Functional Analysis. Springer.
- [4] Cuevas, A., Febrero, M. and Fraiman, R. (2007). Robust estimation and classification for functional data via projection-based depth notions. *Comp. Statist.*, **22**, 481 - 96.
- [5] Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multivar. Anal.*, **12**, 136-154.
- [6] Ferraty, F. and Romain, Y. (2010). *The Oxford handbook of functional data analysis*, Oxford University Press.
- [7] Ferraty, F., Laksaci, A. and Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference Stochastic Process*, **9**, 47-76.
- [8] Ferraty, F., Laksaci, A., Tadj, A. and View, P. (2010). Rate of uniform consistency for non-parametric estimates with functional variables. *Journal of Statistical Planning and Inference*, **140**, 335-352.
- [9] Ferraty, F. and Vieu, Ph. (2006). *Nonparametric functional data analysis: Theory and practice*. Springer, New York.
- [10] Fraiman, R. and Muñiz, G. (2001). Trimmed means for functional data. *Test*, **10**, 419-40.
- [11] Gervini, D. (2006). Free-knot spline smoothing for functional data. *J. Roy. Statist. Soc. Ser. B*, **68**, 671-87.

- [12] Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.
- [13] Hall, P. and Hosseini-Nasab, M. (2006). On properties of functional principal components analysis. *J. Roy. Statist. Soc. Ser. B*, **68**, 109-26.
- [14] Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Ann. Statist.*, **34**, 1493-517.
- [15] Hyndman, R. J. and S. Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Comp. Statist. Data Anal.*, **51**, 4942-4956.
- [16] Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L. (1999). Robust principal components for functional data (with Discussion). *Test*, **8**, 1-73.
- [17] López-Pintado, S. and Romo, J. (2007). Depth-based inference for functional data. *Comp. Statist. Data Anal.*, **51**, 4957-68.
- [18] Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Can. J. Statist.*, **31**, 115-128.
- [19] Pezzulli, S. D. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Comput. Statist.*, **8**, 1-16.
- [20] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag.
- [21] Rice, J. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, **53**, 233-243.
- [22] Sawant, P., Billor, N. and Shin, H. (2011) Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, **27**, 83-102.
- [23] Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Ann. Statist.*, **24**, 1-24.
- [24] Shang, H.L. (2014). Bayesian bandwidth estimation for a semi-functional partial linear regression model with unknown error density. *Computational Statistics*, **29**, 829-848.
- [25] Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *J. R. Statist. Soc. Ser. B*, **68**, 3-25.