UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Algunos resultados sobre puntos principales y distribuciones elípticas en el caso funcional

Juan Lucas Bali

Directores: Dra. Graciela L. Boente Boente y Lic. Daniela Rodriguez

Lugar de trabajo: Instituto de Cálculo, FCEyN, UBA.

15 de Marzo de 2008

Índice

1. Introducción

1.1. Objetivos y Motivación

Dentro de la estadística existen numerosas situaciones en donde los datos recolectados no se pueden representar bien a través de los esquemas clásicos como números o vectores numéricos y en donde lo más adecuado será una representación funcional de los mismos. Por ejemplo, los resultados de un electrocardiograma o el estudio de la variación de la temperatura en una estación meteorológica se prestan mejor para ser representados a través de una función antes que una serie discretizada de datos en donde bien podrían diluirse aspectos funcionales como la continuidad o la diferenciablidad.

Para paliar esta problemática se consideraron en las últimas décadas métodos de estimación para datos funcionales. De manera informal, podríamos decir que un dato funcional es una variable aleatoria cuyo resultado, en vez de ser un número o un vector de dimensión finita de números, es una función. En el presente trabajo estudiaremos fundamentalmente las cuestiones más teóricas de este desarrollo, que resultan en esencia una amalgama de la estadística con el análisis funcional sobre espacios de Banach y de Hilbert. Definiremos formalmente el concepto de elemento aleatorio (funcional) y extrapolaremos las definiciones clásicas de esperanza, varianza y covarianza al contexto funcional. Más adelante detallaremos con precisión el cometido final, pero baste adelantar que la idea es poder reconciliar, dentro de una familia bastante general de distribuciones, algunas nociones de componentes principales y puntos principales. Estos trabajos fueron llevados a cabo principalmente por Bernhard Flury y Tarpey a comienzos de la década del 90 para el caso multivariado. La idea de este trabajo será adaptar los resultados concernientes a los puntos principales al caso funcional.

Tal vez el objetivo final de este trabajo no recaiga tanto en el hecho específico de lo que se pretende demostrar. Más bien la idea es demostrar, digamos mediante un "proof of concept" sobre la factibilidad de poder hacer, con ciertas dificultades pero que no son mayores, una generalización interesante de los resultados clásicos del análisis multivariado y poder así alcanzar mayor comprensión del fenómeno, como cada vez que se efectúa una labor de generalización o abstracción de un concepto matemático.

1.2. Descripción de la Tesis

En el **Capítulo 2**, introduciremos la noción de elemento aleatorio, nuestro elemento de trabajo, en un espacio de Banach/Hilbert y demostraremos una serie de propiedades sobre el mismo. Se introduce además la noción de esperanza y covarianza funcional de manera tal de poder recurrir a las definiciones clásicas del caso multivariado, a través de proyecciones. Esta será una línea de trabajo que se utilizará ampliamente en esta tesis.

En el **Capítulo 3**, se dará un descripción somera sobre la noción de componentes principales, primero en el caso multivariado y posteriormente para el funcional. Nos centraremos fundamentalmente en el caso poblacional puesto que será éste el que se desarrollará posteriormente. Para el caso multivariado optamos por seguir la línea de un texto clásico, como el Seber (1984).

En el **Capítulo 4**, definiremos las familias esféricas y elípticas de distribuciones. Tal como se hizo en el capítulo anterior, empezaremos primero en el caso multivariado y posteriormente se extenderán las definiciones para el caso funcional. Como siempre, se trata en la medida de lo posible de recurrir a lo antes planteado para la situación clásica.

El **Capítulo 5** es el núcleo fuerte de la tesis. Se definen los puntos principales y, en sentido más general, los autoconsistentes, para el caso funcional. Se extienden los resultados de los trabajo de Flury (1990, 1993) y Tarpey y Flury (1995, 1996) para el caso de considerar elementos en un espacio de Hilbert separable. Se concluye ofreciendo una caracterización que, bajo una hipótesis de elipticidad de la familia, nos permite efectuar una vinculación importante entre componentes principales y puntos autoconsistentes.

2. Teoría básica de datos funcionales

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 2.1. Sea $(\Omega, \mathcal{A}, P)$ 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 $V : \Omega \to \mathcal{X}$. Diremos que V es un elemento aleatorio en sentido débil si para todo $f \in \mathcal{X}^*$, $f(V) : \Omega \to I\!R$ es una variable aleatoria real.

Definición 2.2. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y sea $\mathcal{X}$ un espacio de Banach con su σ-álgebra de Borel. Sea $V : \Omega \to \mathcal{X}$. 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 2.1. *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

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

siendo $\mathcal{B}(I\!R)$ la σ-álgebra de Borel en $I\!R$. 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 $I\!R$ 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 $\mathcal{B}(\mathcal{C})$. 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 $B \in \mathcal{B}(I\!R)$ 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. $\square$

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 $I\!R$, 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 2.1. *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. Distribuciones de probabilidad e independencia

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

Definición 2.3. 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

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

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 2.4. 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 $P_V(B) = P(V^{-1}(B)) = P(V \in B)$. O sea, $P_V = P \circ V^{-1}$.

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 2.1. *Sean V_1 y V_2 dos elementos aleatorios en $\mathcal{X}$ idénticamente distribuidos, sea $\phi : \mathcal{X} \to \mathcal{Y}$ 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 $P(\phi(V_1) \in B) = P(V_1 \in \phi^{-1}(B)) = P(V_2 \in \phi^{-1}(B)) = P(\phi(V_2) \in B)$, lo que demuestra el resultado. □

Proposición 2.2. *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 2.1, $f(V_1)$ 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, $P \circ V_1^{-1} \circ f^{-1} = P \circ V_2^{-1} \circ f^{-1}$, $\forall f \in \mathcal{X}^*$. 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 $B \in \mathcal{B}(I\!R)$ y $D = \{x \in \mathcal{X} : f(x) \in B\} = f^{-1}(B)$. Luego,

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

lo que concluye la demostración. □

Definamos ahora la noción de independencia.

Definición 2.5. Una familia $\{V_t : \Omega \to \mathcal{X}_t, t \in A\}$ de elementos aleatorios se dicen independientes si $\forall F \subseteq A$ finito, $F = \{t_1, \ldots, t_k\}$, vale que

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

Tenemos resultados análogos a los de idéntica distribución.

Proposición 2.3. *Sean V_1 y V_2 dos elementos aleatorios en $\mathcal{X}_1$ y $\mathcal{X}_2$ respectivamente, ambos independientes. Sean $\phi_1 : \mathcal{X}_1 \to \mathcal{Y}_1$, $\phi_2 : \mathcal{X}_2 \to \mathcal{Y}_2$ 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

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

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

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

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 2.4. *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 2.3, $g(V_1)$ y $g(V_2)$ serán independientes.

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

2.3. Esperanza

Extenderemos la noción de esperanza al caso funcional.

Definición 2.6. Dado $\mathcal{X}$ un espacio de Banach separable y V un elemento aleatorio, diremos que V tiene esperanza $E(V) \in \mathcal{X}$ si

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

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 $f(E_1) = E(f(V)) = f(E_2) \forall f \in \mathcal{X}^*$. Luego $f(E_1) = f(E_2)$ para todo elemento del dual, o sea $f(E_1 - E_2) = 0 \forall f \in \mathcal{X}^*$. Pero, por 6.1 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 2.5. *Sean $\mathcal{X}, \mathcal{Y}$ espacios de Banach separables y $h : \mathcal{X} \to \mathcal{Y}$ un operador lineal acotado. Sea además V un elemento aleatorio en $\mathcal{X}$ con valor esperado $E(V)$. Entonces, vale que $E(h(V))$ está bien definido y además $E(h(V)) = h(E(V))$.*

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

Veamos ahora el segundo requerimiento, la conmutatividad con todo funcional. Hay que ver que $f(h(E(V))) = E(f(h(V))) \ \forall f \in \mathcal{Y}^*$. $E(f(h(V))) = E(g(V))$ con $g = f \circ h \in \mathcal{X}^*$ y por la definición de $E(V)$ esto último será $g(E(V)) = f(h(E(V)))$, lo que concluye la demostración. $\square$

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

Definición 2.7. Sea $\mathcal{X}$ un espacio de Banach separable y V un elemento aleatorio en $\mathcal{X}$ para el que exista $E(V)$. Se define la varianza de V como

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

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

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 $\mathcal{H} = L^2[0,1]$. Ya se dispone en dicho espacio de una definición para el operador de covarianza, probaremos más adelante que la definición propuesta en este trabajo efectivamente la extiende.

Definición 2.8. Sea $\mathcal{H}$ un espacio de Hilbert separable y sea V un elemento aleatorio definido en $\mathcal{H}$ con momento de orden 2 finito. Definimos $a_V : \mathcal{H} \times \mathcal{H} \to R$ el operador bilineal de covarianza puntual como $a_V(u,v) = Cov(< u, V >, < v, V >)$.

Es claro que a_V es bilineal, lo heredará 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,

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

Debido a que $a_V(u,v)$ es bilineal podremos entonces, aplicando el Teorema de Representación de Riesz (6.9) y la continuidad del operador bilineal, definir únicos operadores A y B acotados tales que $a_V(u,v) =< Au, v >_{\mathcal{H}} =< u, Bv >_{\mathcal{H}} \;\; \forall u, v \in \mathcal{H}$.

Definición 2.9. Sea a_V un operador de covarianza puntual y sea A tal que $a_V(u,v) =< Au, v >$ $\forall u, v \in \mathcal{H}$. Diremos entonces que A es el operador de covarianza de V y lo notaremos como Γ_V.

Proposición 2.6. *El operador Γ_V es autoadjunto (simétrico).*

Demostración. Se deduce inmediatamente de la simetría del operador bilineal a_V. $< \Gamma_V u, v >= a_V(u,v) = Cov(< u, V >, < v, V >) = Cov(< v, V >, < u, V >) =< \Gamma_V v, u >=< u, \Gamma_V v >$. □

Proposición 2.7. *El operador Γ_V es semidefinido positivo.*

Demostración. Efectivamente, tenemos $< \Gamma_V u, u >= a_V(u,u) = Cov(< u, V >, < u, V >) = Var(< u, V >) \geq 0$. □

Análogamente podemos definir un operador de covarianza cruzada, requeriremos ahí dos elementos aleatorios U y V no necesariamente en el mismo espacio de Hilbert.

Definición 2.10. Sean $\mathcal{H}_1$ y $\mathcal{H}_2$ espacios de Hilbert separables, sean U y V dos elementos aleatorios definido en $\mathcal{H}_1$ y $\mathcal{H}_2$ respectivamente con momento de orden 2 finito. Definimos $a_{UV} : \mathcal{H}_1 \times \mathcal{H}_2 \to R$ el operador bilineal de covarianza cruzada puntual como $a_{UV}(u,v) = Cov(< u, U >, < v, V >)$.

Igual que antes, el operador de covarianza cruzada será Γ_{UV} tal que $< u, \Gamma_{UV} v >= a_{UV}(u,v)$.

Un resultado muy importante en análisis multivariado es el que permite obtener la matriz de covarianza de un vector aleatorio compuesto con una transformación lineal, obteniendo lo siguiente: $\mathbf{\Sigma}_{AX} = A\mathbf{\Sigma}_X A'$. Tendremos acá un resultado equivalente.

Proposición 2.8. *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 $A : \mathcal{H}_1 \to \mathcal{H}_2$ un operador acotado. Se cumple que $\Gamma_{AV} : \mathcal{H}_2 \to \mathcal{H}_2$ es igual a $\Gamma_{AV} = A\Gamma_V A^*$, siendo $A^* : \mathcal{H}_2 \to \mathcal{H}_1$ el adjunto de A.*

Demostración. Queremos ver que $< \Gamma_{AV} u, v >=< A\Gamma_V A^* u, v > \quad \forall u, v$. Ahora bien, tenemos $< A\Gamma_V A^* u, v >=< \Gamma_V A^*, A^* v >= Cov(< A^* u, V >, < A^* v, V >) = Cov(< u, AV >, < v, AV >) = a_{AV}(u,v) =< \Gamma_{AV} u, v >$ lo que concluye la demostración. □

Tenemos un resultado análogo para el operador de covarianza cruzada.

Proposición 2.9. *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 $A : \mathcal{H}_1 \to \mathcal{H}_3$ y sea $B : \mathcal{H}_2 \to \mathcal{H}_4$ 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, según se definió en la sección anterior. En efecto, vale el siguiente resultado:

Lema 2.1. *Sea $\{e_n\}_{n\in\mathbb{N}}$ una base ortonormal de $\mathcal{H}$, espacio de Hilbert separable. Sea V un elemento aleatorio con varianza definida. Luego $Var(V) = \sum_{n=1}^{\infty} < \Gamma_V e_n, e_n >$.*

Demostración. Observemos que

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

Como $< \cdot, e_n >$ es lineal y continuo, es posible permutarlo con la esperanza, obteniendo:

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

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 $\mathcal{H}$. $\square$

2.5. Funcionales Característicos

Definición 2.11. Sea $\mathcal{X}$ un espacio de Banach separable y P una probabilidad sobre $(\mathcal{X},\mathcal{B}(\mathcal{Y}))$. El funcional característico de P se define como una función $\widehat{P} : \mathcal{X}^* \to \mathbb{R}$ tal que

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

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

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

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

Proposición 2.10. *Valen las siguientes propiedades*

1. $\widehat{P}(0) = 1$.
2. $|\widehat{P}(f)| \leq 1$, $\forall f \in \mathcal{X}^*$.
3. $\widehat{P}(f) = \overline{\widehat{P}(-f)}$, *(conjugación compleja).*
4. *Para cada conjunto finito de números complejos* $z_1, \ldots, z_N$ *y para todo* $f_1, \ldots, f_N \in \mathcal{X}^*$ *se cumple*
$$\sum_{j=1}^{N}\sum_{k=1}^{N} z_j \overline{z_k} \widehat{P}(f_j - f_k) \geq 0.$$
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.

Proposición 2.11. $\widehat{P}$ *es uniformemente continua en* $\mathcal{X}^*$.

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

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

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

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

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

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

Luego,

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

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 (suponiendo que la diferencia es pequeña), con lo cual

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

Ahora, si $\|f_1 - f_2\| < \epsilon/(4\nu)$ se obtiene $|\widehat{P}(f_1) - \widehat{P}(f_2)| < \epsilon$, lo que concluye la demostración. $\square$

Proposición 2.12. *$\widehat{P}_1 = \widehat{P}_2$ 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 $\mathcal{B}(\mathcal{X})$. $\square$

Vamos a probar ahora el resultado que había quedado pendiente de la Sección 2.2, la Proposición 2.4.

Teníamos dos elementos aleatorios $V_1 : \Omega \to \mathcal{X}_1$ y $V_2 : \Omega \to \mathcal{X}_2$ en los espacios de Banach $\mathcal{X}_1$ y $\mathcal{X}_2$, respectivamente. 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 ha definido en la Sección 2.2. Veremos que ambas medidas coinciden, para ello, veremos que ambas medidas inducen la misma función característica.

Llamemos $\widehat{P}_W$ la función característica de P_W y $\widehat{P}_V$ a la de P_V. Antes de hacer cuentas con estas funciones, notemos 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

$$\widehat{P}_V(f) = E(e^{if(W)}) = E(e^{i(f_1(V_1)+f_2(V_2))}) = E(e^{i(f_1(V_1)}\ e^{if_2(V_2))}).$$

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

$$\widehat{P}_V(f) = E(e^{i(f_1(V_1)}\ e^{if_2(V_2))}) = E(e^{if_1(V_1)})E(e^{if_2(V_2)}) = \widehat{P}_{V_1}(f_1)\widehat{P}_{V_2}(f_2) = \widehat{P}_W(f_1 \times f_2)\ ,$$

que era a lo que se quería llegar. Luego, como coinciden ambas funciones características coincidirán entonces las medidas de probabilidad. 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, con lo cual V_1 y V_2 son independientes.

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 un Γ_V tal que evaluado en toda función $f \in L^2[0,1]$ devuelve otra elemento de $L^2[0,1]$ tal que $(\Gamma_V f)(s) = \int_0^1 Cov(V(s), V(t))f(t)dt$.

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

$$< \Gamma_V f, g >= a_V(f, g) = Cov(< f, V >, < g, V >)\ . \tag{1}$$

Desarrollemos el término de la izquierda,

$$< \Gamma_V f, g >_{L^2}= \int_0^1 (\Gamma_V f)(s)g(s)ds = \int_0^1 g(s)(\int_0^1\ Cov(V(s), V(t))\ f(s)dt)\ ds. \tag{2}$$

Por otra parte, si $E_f = E(\int_0^1 f(s)V(s)ds)$, luego, el término derecho de (1) resulta igual a

$$Cov\left(\int_0^1 f(s)V(s)ds, \int_0^1 g(s)V(s)ds\right) = E\left[\left(\int_0^1 f(s)V(s)ds - E_f\right)\left(\int_0^1 g(s)V(s)ds - E_g\right)\right] \tag{3}$$

Ahora bien, tenemos que $E_f = E(< f, V >_{L^2})$. Como, fijado f, $< f, \cdot >_{L^2}$ es un funcional lineal en $\mathcal{H}$, podemos conmutar con la esperanza, por su definición en el caso funcional. Obtenemos entonces $E_f =< f, E(V) >_{L^2}=< f, \mu >$, siendo $\mu \in \mathcal{H}$ la esperanza de V y donde por simplicidad hemos eliminado el subíndice $_{L^2}$. Reemplazando en (3) obtenemos

$$E\left[\left(\int_0^1 f(s)V(s)ds - E_f\right)\left(\int_0^1 g(s)V(s)ds - E_g\right)\right] = E\left((< f, V > -E_f)\left(< g, V > -E_g\right)\right)$$

$$= E(< f, V - \mu >< g, V - \mu >) = \int_\Omega < f, V - \mu >< g, V - \mu > dP$$
$$= \int_\Omega dP \int_0^1 f(s)(V(s) - \mu(s))ds \int_0^1 g(t)(V(t) - \mu(t))dt$$
$$= \int_\Omega dP \int_0^1 f(s)(V(s) - \mu(s))ds \int_0^1 g(t)(V(t) - \mu(t))dt$$
$$= \int_0^1 \int_0^1 f(s)g(t) \int_\Omega (V(s) - \mu(s))(V(t) - \mu(t))dPdsdt = \int_0^1 \int_0^1 f(s)g(t)Cov(V(s), V(t))dsdt$$

y esto es el término derecho de (2) teniendo así lo deseado.

2.7. Esperanza y Distribución Condicional

Trabajaremos ahora con distribuciones condicionales. Puesto que no tendremos en principio la noción de probabilidad puntual o de función de densidad, será necesario tratarlo desde lo 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 2.12. Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad. Dados $X : \Omega \to \mathcal{H}_1$ e $Y : \Omega \to \mathcal{H}_2$, dos elementos aleatorios de $\mathcal{H} = \mathcal{H}_1 \times \mathcal{H}_2$, definimos $g(X) = E(Y|X)$ como la única función $g : \mathcal{H}_1 \to \mathcal{H}_2$ medible respecto de $\mathcal{B}(\mathcal{H}_1)$ tal que $E(< h(X), Y - g(X) >_{\mathcal{H}_2}) = 0$ para toda $h : \mathcal{H}_1 \to \mathcal{H}_2$ medible.

En particular, se cumple que $E(< g(X), Y - g(X) >_{\mathcal{H}_2}) = 0$. Por otra parte, dado $v \in \mathcal{H}_2$ fijo, tomando $h \equiv v$ obtenemos $0 = E(< v, Y - g(X) >_{\mathcal{H}_2}) =< v, E(Y - g(X)) >_{\mathcal{H}_2}$ para todo $v \in \mathcal{H}_2$, con lo cual $E(Y) = E(g(X))$ por la Proposición 2.5. La existencia de la esperanza condicional se puede demostrar utilizando el teorema de Radon-Nikodym. Para una demostración se sugiere consultar Fava y Zó (1996), página 278. Por otra parte, g es la esperanza condicional si y sólo si para toda $f \in \mathcal{H}_2^*$, $f(g(X)) = E(f(Y)|X)$.

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

$$P_{Y|X=x}(A) = P(Y \in A|X = x) = E(I_Y(A)|X = x)$$

3. Componentes Principales

3.1. Caso multivariado

3.1.1. Motivación

El Análisis de Componentes Principales (Principal Components Analysis, o PCA) es una técnica utilizada para poder reducir la dimensión de un conjunto de datos multidimensionales. La idea de la misma es detectar las direcciones que concentran en algún sentido la mayor cantidad posible de información sobre la muestra.

Formalmente, se la define como un nuevo sistema de coordenadas (dada por una transformación ortogonal) de manera tal que la mayor cantidad de variabilidad de cualquier proyección de los datos se encuentre en la primer coordenada (la primera componente principal), la segunda cantidad estará en la segunda coordenada y así hasta agotar la dimensión del espacio. Querríamos entonces quedarnos con un número reducido de coordenadas/dimensiones de manera tal que en las mismas se concentre una parte importante de la variabilidad. De esa forma, se espera poder rescatar los aspectos más importantes de los datos en cuestión. Dependiendo del caso esto podría o no ser así.

Supondremos, inicialmente, que se conoce exactamente la esperanza y la matriz de covarianza del vector aleatorio, por lo que no será necesario contar con una muestra para estimar dichos parámetros. Posteriormente, consideraremos el caso más realista (e interesante) en el que ambos parámetros son desconocidos y se tiene una muestra de tamaño n del vector aleatorio. El segundo caso recae, en los aspectos teóricos, fuertemente en el primero por lo que pretendemos justificar así el esquema con el cual se encararán los temas.

3.1.2. Primer enfoque: Caso Poblacional

Consideremos $\mathbf{X} \in I\!R^d$ un vector aleatorio, con esperanza $E(\mathbf{X}) = \boldsymbol{\mu}$ y matriz de covarianza $\boldsymbol{\Sigma}_X = \boldsymbol{\Sigma}$.

Puesto que $\boldsymbol{\Sigma}$ es una matriz simétrica y semidefinida positiva podemos entonces considerar una descomposición espectral de la misma con autovalores no negativos. Es decir, podemos escribir

$$\boldsymbol{\Sigma} = \mathbf{T}\boldsymbol{\Lambda}\mathbf{T}' \tag{4}$$

con $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_d)$ una matriz diagonal formada por los autovalores de $\boldsymbol{\Sigma}$ ordenados de mayor a menor y $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_d)$ una matriz ortogonal en donde $\mathbf{t}_i$ es el autovector asociado al autovalor λ_i.

Sea $\mathbf{Y}_j = \mathbf{t}_j'(\mathbf{X} - \boldsymbol{\mu})$ la proyección ortogonal de $\mathbf{X} - \mu$ en la dirección $\mathbf{t}_j$. Llamaremos a Y_j la *j-ésima componente principal de* $\mathbf{X}$.

Trabajaremos también con $Z_j = \lambda_j^{-1/2} Y_j$, la j-ésima componente principal estandarizada. Notemos entonces que $\mathbf{Y} = \mathbf{T}'(\mathbf{X} - \boldsymbol{\mu})$. Tenemos además los siguientes resultados con respecto a la varianza.

- $\boldsymbol{\Sigma}_Y = \boldsymbol{\Sigma}_{\mathbf{T}'(\mathbf{X}-\boldsymbol{\mu})} = \mathbf{T}'\boldsymbol{\Sigma}_{\mathbf{X}}\mathbf{T} = \mathbf{T}'\boldsymbol{\Sigma}\mathbf{T} = \boldsymbol{\Lambda}$

- $\mathbf{\Sigma_Z} = \mathbf{\Sigma_{\Gamma Y}} = \mathbf{\Gamma\Sigma_Y\Gamma'} = \mathbf{\Gamma\Lambda\Gamma'} = \mathbf{I}_d$

con $\mathbf{\Gamma} = \text{diag}(\lambda_1^{-1/2}, \ldots, \lambda_d^{-1/2})$.

El siguiente teorema justifica la elección hecha con respecto a $\mathbf{Y}$.

Teorema 3.1. *Sea* $\mathbf{Y} = \mathbf{T}'(\mathbf{X} - \boldsymbol{\mu})$ *entonces se verifica*

a) $\forall \mathbf{a}_1$ *tal que* $\|\mathbf{a}_1\| = 1$, $Var(\mathbf{a}_1'\mathbf{X}) \leq Var(\mathbf{t}_1'\mathbf{X}) = \lambda_1$.

b) $\forall \mathbf{a}_j$ *tal que* $\|\mathbf{a}_j\| = 1$ *y* $\mathbf{a}_j \perp \ < \mathbf{t}_1, \ldots, \mathbf{t}_{j-1} >$, $Var(a_j'\mathbf{X}) \leq Var(\mathbf{t}_j'\mathbf{X}) = \lambda_j$.

c) $\sum_{j=1}^d Var(Y_j) = \sum_{j=1}^d Var(X_j) = tr(\mathbf{\Sigma})$.

Demostración. a) Es inmediato a partir del hecho que $Var(\mathbf{a}_1'\mathbf{X}) = \mathbf{a}_1'\mathbf{\Sigma}\mathbf{a}_1$ y de la caracterización del máximo autovalor.

b) Sea $\mathbf{a}_j \perp \ < \mathbf{t}_1, \ldots, \mathbf{t}_{j-1} >$, luego $\mathbf{a}_j = \sum_{i=j}^d c_i \mathbf{t}_i$ donde los coeficientes $c_r =< \mathbf{a}_r, \mathbf{t}_r >$ verifican $\sum_{i=j}^d c_i^2 = \|\mathbf{a}_j\|^2 = 1$. Usando nuevamente que $Var(\mathbf{a}_j'\mathbf{X}) = \mathbf{a}_j'\mathbf{\Sigma}\mathbf{a}_j$, la expresión para $\mathbf{a}_j$ y el hecho que $\mathbf{t}_i$ son autovectores de $\mathbf{\Sigma}$, obtenemos

$$Var(\mathbf{a}_j'\mathbf{X}) = \mathbf{a}_j' \sum_{i=j}^d c_i \mathbf{\Sigma} \mathbf{t}_i = a_j' \sum_{i=j}^d c_i \lambda_i \mathbf{t}_i = \left(\sum_{i=j}^d c_i \mathbf{t}_i \right)' \left(\sum_{i=j}^d c_i \lambda_i \mathbf{t}_i \right) = \sum_{i=j}^d c_i^2 \lambda_i \leq \lambda_j \sum_{i=j}^d c_i^2 = \lambda_j \ .$$

Por otra parte, la cota superior se alcanza cuando $c_j = 1$ y $c_i = 0$ para $i \neq j$, es decir, si $\mathbf{a}_j = \mathbf{t}_j$.

c) $\sum_{i=1}^d Var(Y_j) = \sum_{i=1}^d \lambda_j = \text{tr}(\mathbf{\Lambda}) = \text{tr}(\mathbf{T\Lambda T}') = \text{tr}(\mathbf{\Sigma})$, lo que concluye la demostración. □

En virtud de este teorema podemos entonces justificar la elección de los Y_j y $\mathbf{t}_j$ antes hecha. Efectivamente, en esas direcciones se estará maximizando la varianza. Por otra parte, la propiedad c) da un método para elegir la cantidad k de componentes principales a tomar. El cociente $\lambda_j / \sum_{i=1}^d \lambda_i = Var(Y_j)/\text{tr}(\mathbf{\Sigma})$ mide la contribución de la $j-$ésima componente a la variación total de $\mathbf{X}$. Luego, uno podría tomar un número de componentes $Y_1, \ldots, Y_k$ tal que $\sum_{i=1}^k \lambda_i/\text{tr}(\mathbf{\Sigma})$ sea suficientemente próximo a 1.

3.1.3. Enfoque Geométrico

Existe otra manera de caracterizar las componentes principales, consistente en un enfoque "geométrico". Dado un vector aleatorio $\mathbf{X} \in I\!R^d$ buscamos un subespacio $\mathcal{L}$ de dimensión $k \leq d$ de manera tal que la proyección de $\mathbf{X}$ sobre $\mathcal{L}$ se encuentre en promedio lo más cercano posible al vector $\mathbf{X}$. Sin pérdida de generalidad podemos suponer que $\mathbf{X}$ es un vector aleatorio de media 0. Observemos que el Teorema 3.1 permite deducir que si $\lambda_j = 0$ para $j \geq k+1$, entonces Y_j tendrá varianza 0 para $j \geq k+1$, de donde $P(\mathbf{t}_j\mathbf{X} = 0\,,\ k+1 \leq j \leq d) = 1$, es decir $\mathbf{X}$ yace en un subespacio de dimensión k. En este caso, es claro que el subespacio óptimo es el generado por las primeros k autovectores de $\mathbf{\Sigma}$.

En el caso general, el objetivo es minimizar $E(\|\mathbf{X} - P_{\mathcal{L}}\mathbf{X}\|^2)$, siendo $P_{\mathcal{L}}$ la proyección ortogonal sobre el espacio $\mathcal{L}$. El siguiente resultado muestra que el subespacio óptimo es nuevamente el generado por las q componentes principales.

Teorema 3.2. *Sea* $\mathbf{X} \in \mathbb{R}^d$ *un vector aleatorio tal que* $E(\mathbf{X}) = 0$ *y sea* $\mathbf{\Sigma_X} = \mathbf{\Sigma}$ *que satisface (4). Supongamos que* $\lambda_k > \lambda_{k+1}$ *y sea* $\mathcal{L}_0 =< \mathbf{t}_1, \ldots, \mathbf{t}_k >$*. Entonces se tiene que* $E(\|\mathbf{X} - P_{\mathcal{L}_0}\mathbf{X}\|^2) \leq E(\|\mathbf{X} - P_{\mathcal{L}}\mathbf{X}\|^2)$ *para todo subespacio* $\mathcal{L}$ *de dimensión* k.

Demostración. Sea $\mathbf{v}_1, \ldots, \mathbf{v}_k >$ una base ortonormal de $\mathcal{L}$. Entonces, $P_{\mathcal{L}}\mathbf{X} = \sum_{i=1}^k < \mathbf{v}_i, \mathbf{X} > \mathbf{v}_i$. Completando a una base $\{\mathbf{v}_i\}_{1\leq i\leq d}$ de $\mathbb{R}^d$, tendremos que $\mathbf{X} = \sum_{i=1}^d < \mathbf{v}_i, \mathbf{X} > \mathbf{v}_i$. Como $E(\|\mathbf{X}\|^2) = E(\|P_{\mathcal{L}}\mathbf{X}\|^2) + E(\|\mathbf{X} - P_{\mathcal{L}}\mathbf{X}\|^2)$ y el lado izquierdo de la igualdad es independiente de $\mathcal{L}$, obtenemos que minimizar el segundo sumando es equivalente a maximizar el primero. Observemos que

$$E(\|P_{\mathcal{L}}\mathbf{X}\|^2) = E\left(\sum_{i=1}^k | < \mathbf{v}_i, \mathbf{X} > |^2\right) = \sum_{i=1}^k E(\mathbf{v}_i'\mathbf{X}\mathbf{X}'\mathbf{v}_i) = \sum_{i=1}^k \mathbf{v}_i'\mathbf{\Sigma}\mathbf{v}_i$$

donde la última igualdad se deduce del hecho que $E(\mathbf{X}) = 0$. Bastará ver que $\sum_{i=1}^k \mathbf{v}_i'\mathbf{\Sigma}\mathbf{v}_i$ se maximiza tomando $\mathbf{v}_i = \mathbf{t}_i$ el $i-$ésimo autovector de $\mathbf{\Sigma}$. Sea $\mathbf{v}_i = \sum_{j=1}^d < \mathbf{v}_i, \mathbf{t}_j > \mathbf{t}_j = \sum_{j=1}^d c_{ij}\mathbf{t}_j$. Por lo tanto, si $\mathbf{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_k)$ podemos escribir $\mathbf{V} = \mathbf{CT}$ donde la matriz $\mathbf{C}$ cumple $\mathbf{CC}' = \mathbf{I}_k$ ya que $\mathbf{v}_i$ son una vectores ortonormales. Luego, $\sum_{i=1}^k c_{ij}^2 \leq 1$. Observemos entonces que $\sum_{i=1}^k \mathbf{v}_i'\mathbf{\Sigma}\mathbf{v}_i = \sum_{j=1}^p \lambda_j \sum_{i=1}^k c_{ij}^2$, donde $\sum_{i=1}^k c_{ij}^2 \leq 1$ y $\lambda_1 \geq \lambda_2 \geq \ldots \lambda_p$. Luego, $\sum_{j=1}^p \lambda_j \sum_{i=1}^k c_{ij}^2$ es máximo si $c_{jj} = 1$ para $1 \leq j \leq k$ y $c_{ij} = 0$ para $i \neq j$, o sea si $\mathbf{v}_i = \mathbf{t}_i$. □

3.1.4. Reducción de la dimensión

En esta sección trataremos de formalizar la idea de que las componentes principales proveen una reducción efectiva de la dimensión del vector original. Para ello, dado un $\mathbf{X} \in \mathbb{R}^d$, transformarlo a un vector $\mathbf{Y} \in \mathbb{R}^k$ tal que para una elección adecuada de $\mathbf{A} \in \mathbb{R}^{d\times k}$ valga que $\mathbf{AY}$ sea "cercano" a $\mathbf{X}$, en algún sentido. De alguna manera, $\mathbf{Y}$ representará la información comprimida de $\mathbf{X}$ y $\mathbf{A}$ será entonces la "matriz de descompresión".

Precisamos formalizar bien esta noción de cercanía y para ello fijaremos alguna notación.

Supondremos que $\mathbf{\Sigma}_Y = \mathbf{\Gamma} > 0$, es decir, que $\mathbf{Y}$ tiene matriz de covarianza definida positiva.

Vamos a escribir de una forma distinta $\mathbf{AY}$, el vector "descomprimido" de $\mathbf{Y}$.

$$\mathbf{AY} = \left(\mathbf{A\Gamma}^{\frac{1}{2}}\right)\left(\mathbf{\Gamma}^{-\frac{1}{2}}\mathbf{Y}\right) = \mathbf{A}_0\mathbf{Z}\ ,$$

con $\mathbf{A}_0 = \mathbf{A\Gamma}^{\frac{1}{2}}$ y $\mathbf{Z} = \mathbf{\Gamma}^{-\frac{1}{2}}\mathbf{Y}$. Con esta transformación tenemos que $\mathbf{\Sigma_Z} = \mathbf{I}_k$.

Supondremos como antes que $E(\mathbf{X}) = 0$, por lo tanto, también supondremos $E(\mathbf{Z}) = 0$. Sean

- $\mathcal{P}_0 = \{\mathbf{A} \in \mathbb{R}^{d\times d} / \ \mathbf{A}$ es semidefinida positiva $\}$
- $\mathcal{P} = \{\mathbf{A} \in \mathbb{R}^{d\times d} / \ \mathbf{A}$ es positiva $\}$

tenemos entonces las siguientes definiciones.

Definición 3.1.

- Si $\mathbf{C}, \mathbf{D} \in \mathcal{P}_0$, diremos que $\mathbf{C} \geq \mathbf{D}$ si $\mathbf{C} - \mathbf{D} \in \mathcal{P}_0$.
- Si $\mathbf{C}, \mathbf{D} \in \mathcal{P}_0$, diremos que $\mathbf{C} > \mathbf{D}$ si $\mathbf{C} - \mathbf{D} \in \mathcal{P}$.

Definimos ahora nuestra medida de cercanía.

Definición 3.2. $f : \mathcal{P} \to I\!R$ se dice estrictamente creciente si $\mathbf{C} > \mathbf{D}$ implica que $f(\mathbf{C}) > f(\mathbf{D})$.

Definición 3.3. $f : \mathcal{P} \to I\!R$ se dice invariante por transformaciones ortogonales si para toda matrix $\mathbf{T}$ ortogonal $f(\mathbf{T}'\mathbf{C}\mathbf{T}) = f(\mathbf{C})$.

Definición 3.4. $f : \mathcal{P} \to I\!R$ se dice una medida de cercanía si es creciente e invariante por transformaciones ortogonales.

La traza, la norma de Frobenius y el determinante son ejemplos de medidas cercanía. El siguiente lema provee una caracterización de las medidas de cercanía a partir de su comportamiento con los autovalores de la matriz argumento.

Lema 3.1. *Sea $f : \mathcal{P} \to I\!R$, dada $\mathbf{C} \in \mathcal{P}_0$, sean $\xi_1(\mathbf{C}) \geq \ldots \geq \xi_d(\mathbf{C}) \geq 0$ los autovalores de $\mathbf{C}$. Una condición necesaria y suficiente para que f sea una medida de cercanía es que exista $g : I\!R^d \to I\!R$ estrictamente creciente en cada argumento tal que $f(\mathbf{C}) = g(\xi_1(\mathbf{C}), \ldots, \xi_d(\mathbf{C}))$.*

Demostración. a) Supongamos que f es una medida de cercanía, es decir, una función estrictamente creciente e invariante por transformaciones ortogonales. Aplicando una descomposición espectral sobre $\mathbf{C}$, obtenemos una matriz ortogonal $\mathbf{T}$ tal que $\mathbf{T}'\mathbf{C}\mathbf{T} = \text{diag}(\xi_1, \ldots, \xi_d) = \boldsymbol{\Upsilon}$. Entonces, $f(\mathbf{C}) = f(\mathbf{T}'\mathbf{C}\mathbf{T}) = f(\boldsymbol{\Upsilon})$. Definamos entonces $g(a_1, \ldots, a_d) = f(\text{diag}(a_1, \ldots, a_d))$.

Veamos que la función g así definida resulta ser estrictamente creciente en cada argumento.

Consideremos $\boldsymbol{\Upsilon}_1 = \text{diag}(\xi_1, \ldots, \xi_j + \delta, \ldots, \xi_d)$ con $\delta > 0$. Es inmediato que $\boldsymbol{\Upsilon}_1 > \boldsymbol{\Upsilon}$, luego como f es estrictamente creciente tendremos que

$$g(\xi_1, \ldots, \xi_j + \delta, \ldots, \xi_d) = f(\boldsymbol{\Upsilon}_1) > f(\boldsymbol{\Upsilon}) = g(\xi_1, \ldots, \xi_d)$$

y tenemos así una implicación.

b) Veamos ahora la otra, o sea que si $f : \mathcal{P} \to I\!R$ definida como $f(\mathbf{C}) = g(\xi_1(\mathbf{C}), \ldots, \xi_d(\mathbf{C}))$ con g estrictamente creciente en cada argumento entonces f resulta una medida de cercanía.

Sea $\mathbf{S}$ una matriz ortogonal, $|\mathbf{S}'\mathbf{C}\mathbf{S} - \mu\mathbf{I}_d| = |\mathbf{S}'|\,|\mathbf{C} - \mu\mathbf{I}_d|\,|\mathbf{S}|$ y esto es 0 si y sólo si $|\mathbf{C} - \mu\mathbf{I}_d| = 0$, con lo cual $\mathbf{S}'\mathbf{C}\mathbf{S}$ y $\mathbf{C}$ tienen los mismos autovalores. Por lo tanto, como f depende exclusivamente de los autovalores de $\mathbf{C}$ obtenemos la invariancia frente a transformaciones ortogonales.

Veamos que f es estrictamente creciente. Sea $\mathbf{C} \geq \mathbf{D}$ con $\mathbf{C} \neq \mathbf{D}$. Vale entonces, por el Teorema 6.5, que $\xi_j(\mathbf{C}) \geq \xi_j(\mathbf{D})$ para todo j y existe j_0 tal que $\xi_{j_0}(\mathbf{C}) > \xi_{j_0}(\mathbf{D})$, con lo cual por ser g creciente se obtiene $f(\mathbf{C}) = g(\xi_1(\mathbf{C}), \ldots, \xi_d(\mathbf{C})) > g(\xi_1(\mathbf{D}), \ldots, \xi_d(\mathbf{D})) = f(\mathbf{D})$. $\square$

Fijada entonces una medida de cercanía f, buscamos entonces un vector aleatorio $\mathbf{Y}$ y una matriz $\mathbf{A}$ tal que el vector de la forma $\mathbf{AY}$ o equivalentemente $\mathbf{A}_0\mathbf{Z}$ cumpla que $f(\mathbf{\Sigma_{X-AY}}) = f(\mathbf{\Delta})$ es mínimo, donde $\mathbf{\Delta} = \mathbf{\Sigma_{X-AY}}$. Vamos a ver en primer lugar que una cota para $f(\mathbf{\Delta})$ estará dada por algo que depende de $\lambda_{k+1}, \ldots, \lambda_d$, los menores autovalores de la matriz de covarianza de $\mathbf{X}$, $\mathbf{\Sigma}$ cuya descomposición está dada por (4). Es decir, por algo que en algún sentido se está omitiendo en el proceso de compresión. Todo esto representaría entonces la información perdida que ningún descompresor $\mathbf{A}$ podría recuperar. Expresamos todo eso en el siguiente resultado:

Teorema 3.3. *Sea $f : \mathcal{P} \to I\!R$ una medida de cercanía, $f(\mathbf{C}) = g(\xi_1(\mathbf{C}), \ldots, \xi_d(\mathbf{C}))$ con g estrictamente creciente en cada argumento. Sean $\mathbf{X} \in I\!R^d$, $\mathbf{Y} \in I\!R^k$ vectores aleatorios de media 0, $\mathbf{A} \in I\!R^{d\times k}$ y sea $\mathbf{\Delta} = \mathbf{\Sigma_{X-AY}}$. Sean además $\lambda_1, \ldots \lambda_d$ los autovalores de la matriz $\mathbf{\Sigma} = \mathbf{\Sigma}_X$. Vale entonces que para todo $\mathbf{A}$, $f(\mathbf{\Delta}) \geq g(\lambda_{k+1}, \ldots, \lambda_d)$.*

Demostración. Observemos que $\mathbf{\Delta} = \mathbf{\Sigma_{X-AY}} = Cov(\mathbf{X} - \mathbf{A}_0\mathbf{Z}, \mathbf{X} - \mathbf{A}_0\mathbf{Z})$ desarrollando obtenemos

$$\begin{aligned}\mathbf{\Delta} &= \mathbf{\Sigma_X} - Cov(\mathbf{A}_0\mathbf{Z}, \mathbf{X}) - Cov(\mathbf{X}, \mathbf{A}_0\mathbf{Z}) + Cov(\mathbf{A}_0\mathbf{Z}, \mathbf{A}_0\mathbf{Z}) \\ &= \mathbf{\Sigma_X} - \mathbf{A}_0 Cov(\mathbf{Z}, \mathbf{X}) - Cov(\mathbf{X}, \mathbf{Z})\mathbf{A}_0' + \mathbf{A}_0\mathbf{\Sigma_Z}\mathbf{A}_0' \,.\end{aligned}$$

El último sumando resultará ser $\mathbf{A}_0\mathbf{A}_0'$ pues $\mathbf{\Sigma_Z} = \mathbf{I}_k$. Indiquemos por $\mathbf{B} = Cov(\mathbf{X}, \mathbf{Z})$, luego tenemos

$$\mathbf{\Delta} = \mathbf{\Sigma} - \mathbf{A}_0\mathbf{B}' - \mathbf{B}\mathbf{A}_0' + \mathbf{A}_0\mathbf{A}_0' = (\mathbf{\Sigma} - \mathbf{BB}') + (\mathbf{A}_0 - \mathbf{B})(\mathbf{A}_0 - \mathbf{B})' \,.$$

Por la Proposición 6.1 el segundo sumando será mayor o igual que cero. Nos queda entonces $\mathbf{\Delta} = \mathbf{\Sigma_{X-AY}} \geq \mathbf{\Sigma} - \mathbf{BB}' = \mathbf{\Sigma_{X-BZ}} \geq 0$.

El rango de $\mathbf{B}$ es a lo sumo k, luego por el Teorema 6.6 tenemos que

$$\xi_i(\mathbf{\Sigma} - \mathbf{BB}') \geq \begin{cases} \xi_{k+i}(\mathbf{\Sigma}) = \lambda_{k+i} & i = 1, 2, \ldots, d-k \\ 0 & i = d-k+1, d-k+2, \ldots, d \end{cases} .$$

Por el lema 3.1, se verifica que $f(\mathbf{\Delta}) \geq f(\mathbf{\Sigma} - \mathbf{BB}') = g(\xi_1(\mathbf{\Sigma} - \mathbf{BB}'), \ldots, \xi_d(\mathbf{\Sigma} - \mathbf{BB}')) = g(\lambda_{k+1}, \ldots, \lambda_d, 0, \ldots, 0)$. $\square$

Tendríamos así una cota para $f(\mathbf{\Delta})$ que dice que no se puede achicar por debajo de algo que depende exclusivamente de f y de los autovalores $\lambda_{k+1}, \ldots, \lambda_d$. O sea, de la información que se está en algún sentido omitiendo. El siguiente resultado nos permite afirmar que la cota antes encontrada es efectivamente alcanzable.

Teorema 3.4. *Sea $\mathbf{T} = (\mathbf{T}_1, \mathbf{T}_2)$, con $\mathbf{T}_1 = (\mathbf{t}_1, \ldots, \mathbf{t}_k)$ definido como en (4). Si f es una medida de cercanía entonces $f(\mathbf{\Delta}) = f(\mathbf{\Sigma_{X-AY}})$ se minimiza cuando $\mathbf{AY} = \mathbf{A}_0\mathbf{Z} = \mathbf{T}_1\mathbf{T}_1'\mathbf{X}$.*

Demostración. Dado $\mathbf{AY} = \mathbf{A}_0\mathbf{Z} = \mathbf{T}_1\mathbf{T}_1'\mathbf{X} = \mathbf{DX}$, se verifica que $\mathbf{X} - \mathbf{AY} = (\mathbf{I}_d - \mathbf{D})\mathbf{X} = \mathbf{QX}$, con $\mathbf{Q}$ una matriz de proyección (simétrica e idempotente). Tenemos entonces que $\mathbf{\Delta} = \mathbf{\Sigma_{X-AY}} = \mathbf{\Sigma_{QX}} = \mathbf{Q\Sigma Q}$. Ahora bien, por (4) vale que $\mathbf{\Sigma} = \mathbf{T\Lambda T}'$, entonces

$$\mathbf{D\Sigma} = \mathbf{T}_1\mathbf{T}_1'\mathbf{T\Lambda T}' = \mathbf{T}_1\mathbf{T}_1'(\mathbf{T}_1, \mathbf{T}_2)\begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{pmatrix}\begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_2' \end{pmatrix} = (\mathbf{T}_1, \mathbf{0})\begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{pmatrix}\begin{pmatrix} \mathbf{T}_1' \\ \mathbf{T}_2' \end{pmatrix} = \mathbf{T}_1\mathbf{\Lambda}_1\mathbf{T}_1'$$

donde $\mathbf{\Lambda}_1 \in I\!R^{k\times k}$ son los primeros k autovalores de $\mathbf{\Sigma}$. Como $\mathbf{Q} = \mathbf{I}_d - \mathbf{D} = \mathbf{T}\mathbf{T}' - \mathbf{T}_1\mathbf{T}_1' = \mathbf{T}_2\mathbf{T}_2'$, se obtiene fácilmente que $\mathbf{Q\Sigma} = \mathbf{T}_2\mathbf{\Lambda}_2\mathbf{T}_2'$. Además, a partir de la ortogonalidad de la matriz $\mathbf{T}$, se tiene $\mathbf{T}_2'\mathbf{T}_1 = \mathbf{0}$, de donde $\mathbf{Q\Sigma D} = \mathbf{T}_2\mathbf{\Lambda}_2\mathbf{T}_2'\mathbf{T}_1\mathbf{T}_1' = \mathbf{0}$. Por lo tanto,

$$\mathbf{\Delta} = \mathbf{Q\Sigma Q} = \mathbf{Q\Sigma}(\mathbf{I}_d - \mathbf{D}) = \mathbf{Q\Sigma} - \mathbf{Q\Sigma D} = \mathbf{Q\Sigma} = \mathbf{T}_2\mathbf{\Lambda}_2\mathbf{T}_2' = \mathbf{T}\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{pmatrix}\mathbf{T}'$$

de donde,

$$\begin{aligned} f(\mathbf{\Delta}) &= f\left(\mathbf{T}'\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{pmatrix}\mathbf{T}'\right) = f\begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_2 \end{pmatrix} \\ &= g(\xi_1(\mathbf{\Lambda_2}), \ldots, \xi_{d-k}(\mathbf{\Lambda_2}), 0, \ldots, 0) = g(\lambda_{k+1}, \ldots, \lambda_d, 0, \ldots, 0) \end{aligned}$$

donde la anteúltima igualdad se deduce del lema 3.1. El último miembro de la igualdad es precisamente la cota definida en el teorema 3.3, de donde $f(\mathbf{\Delta})$ alcanza su valor mínimo cuando $\mathbf{AY} = \mathbf{DX}$. □

3.1.5. Caso muestral

En la práctica $\boldsymbol{\mu}$ y $\mathbf{\Sigma}$ son parámetros desconocidos y por lo tanto, deberán ser estimados a partir de una muestra $\mathbf{x}_1, \ldots, \mathbf{x}_n$ de observaciones independientes e indénticamente distribuídas. Consideremos $\widehat{\boldsymbol{\mu}} = \overline{\mathbf{x}}$ y $\widehat{\mathbf{\Sigma}} = \sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}})(\mathbf{x}_i - \overline{\mathbf{x}})'/n$. Sean $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$ los autovalores de $\widehat{\mathbf{\Sigma}}$ y $\widehat{\mathbf{T}} = (\widehat{\mathbf{t}}_1, \ldots, \widehat{\mathbf{t}}_d)$ la matriz ortogonal formada por los autovectores correspondientes. Para cada observación $\mathbf{x}_i$ podemos definir un vector de componentes principales muestrales, $\mathbf{y}_i = \widehat{\mathbf{T}}'(\mathbf{x}_i - \overline{\mathbf{x}})$ obteniendo así una nueva matriz

$$\mathbf{Y} = (\mathbf{y}_1, \ldots, \mathbf{y}_n) = \widehat{\mathbf{T}}'(\mathbf{x}_1 - \overline{\mathbf{x}}, \ldots, \mathbf{x}_n - \overline{\mathbf{x}}).$$

Se esperaría, como en el caso poblacional, que las primeras k componentes principales muestrales de un vector de observaciones nos dé efectivamente el mejor resúmen de esas observaciones que se pueda conseguir con k variables. El siguiente teorema apunta en este sentido:

Teorema 3.5. *Sea* $\mathbf{A}$ *una matriz en* $\mathbb{R}^{d\times k}$ *(matriz de "descompresión"). Tomemos* $\mathbf{y}_i = \mathbf{g}(\mathbf{x}_i - \overline{\mathbf{x}})$, *la "i-ésima componente comprimida", siendo* $\mathbf{g} : \mathbb{R}^d \to \mathbb{R}^k$ *una función medible tal que* $\overline{\mathbf{y}} = 0$. *Si* f *es una medida de cercanía (estrictamente creciente e invariante por transformaciones ortogonales) entonces*

$$f\left\{\frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \overline{\mathbf{x}} - \mathbf{AY})(\mathbf{x}_i - \overline{\mathbf{x}} - \mathbf{AY})'\right\} \tag{5}$$

se minimiza con respecto de $\mathbf{A}$ *y de* $\mathbf{g}$ *eligiendo*

$$\mathbf{A} = (\widehat{\mathbf{t}}_1, \ldots, \widehat{\mathbf{t}}_k) = \widehat{\mathbf{T}}_k \quad y \quad \mathbf{g}(\mathbf{x} - \overline{\mathbf{x}}) = \widehat{\mathbf{T}}_k'(\mathbf{x} - \overline{\mathbf{x}}).$$

Demostración. La estrategia es utilizar los resultados del caso poblacional tomando como medida de probabilidad la distribución empírica. En ese sentido, se define un vector aleatorio que "emulará" el comportamiento de nuestra muestra. Sea $\mathbf{v}$ un vector aleatorio que toma cada valor $\mathbf{x}_i$ con probabilidad $1/n$, $\mathbf{w} = \mathbf{v} - \overline{\mathbf{x}}$ e $\mathbf{y} = \mathbf{g}(\mathbf{w})$. Entonces, $\boldsymbol{\eta} = E(\mathbf{v}) = \sum_{i=1}^{n} \mathbf{x}_i/n = \overline{\mathbf{x}}$, $E(\mathbf{w}) = 0$ y $E(\mathbf{y}) = \overline{\mathbf{y}}$. Por otra parte,

$$\boldsymbol{\Sigma}_{\mathbf{w}} = \boldsymbol{\Sigma}_{\mathbf{v}} = E\left[(\mathbf{v}-\boldsymbol{\eta})(\mathbf{v}-\boldsymbol{\eta})'\right] = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\eta})(\mathbf{x}_i-\boldsymbol{\eta})' = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\overline{\mathbf{x}})(\mathbf{x}_i-\overline{\mathbf{x}})' = \boldsymbol{\Sigma}.$$

Luego, (5) es igual a

$$f\left(E\left[(\mathbf{w}-\mathbf{A}\mathbf{y})(\mathbf{w}-\mathbf{A}\mathbf{y})'\right]\right) = f\left(\boldsymbol{\Sigma}_{\mathbf{w}-\mathbf{A}\mathbf{y}}\right)$$

que, por el Teorema 3.4, se minimiza en todos los vectores aleatorios $\mathbf{y}$ tales que $E(\mathbf{y}) = 0$ (y luego para cualquier $\mathbf{y}$ que es imagen a través de una función medible de $\mathbf{w}$) cuando $\mathbf{A}\mathbf{y} = \widehat{\mathbf{T}}_k\widehat{\mathbf{T}}_k'\mathbf{w}$. Esto último vale cuando $\mathbf{A} = \widehat{\mathbf{T}}_k$ y $\mathbf{g}(\mathbf{w}) = \widehat{\mathbf{T}}_k'\mathbf{w}$. $\square$

Como consecuencia del método utilizado en el teorema anterior, no habría diferencia esencial entre un acercamiento poblacional o uno muestral a las componentes principales. En efecto, podemos reemplazar $\mathbf{x}$, $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ por $\mathbf{v}, \overline{\mathbf{x}}$ y $\widehat{\boldsymbol{\Sigma}}$ para tener así hecha la "conversión" entre ambos casos, es decir, las propiedades vistas para el caso poblacional se heredan automáticamente a nivel muestral. En efecto, podemos por ejemplo, transcribir el Teorema 3.1 como sigue:

Teorema 3.6. *Sea* $\mathbf{Y} = \widehat{\mathbf{T}}'(\mathbf{v} - \overline{\mathbf{x}})$ *entonces se verifica*

a) $\forall \mathbf{a}_1$ *tal que* $\|\mathbf{a}_1\| = 1$, $Var(\mathbf{a}_1'\mathbf{v}) \le Var(\widehat{\mathbf{t}}_1'\mathbf{v}) = \widehat{\lambda}_1$.

b) $\forall \mathbf{a}_j$ *tal que* $\|\mathbf{a}_j\| = 1$ *y* $\mathbf{a}_j \perp <\widehat{\mathbf{t}}_1, \ldots, \widehat{\mathbf{t}}_{j-1}>$, $Var(a_j'\mathbf{v}) \le Var(\widehat{\mathbf{t}}_j'\mathbf{v}) = \lambda_j$.

c) $\sum_{j=1}^{d} Var(Y_j) = \sum_{j=1}^{d} Var(v_j) = tr(\widehat{\boldsymbol{\Sigma}})$.

3.2. Caso funcional

Lo que se hará es básicamente seguir los lineamientos planteados cuando se estudió la perspectiva geométrica de las componentes principales. Consideremos entonces V un elemento aleatorio en un espacio de Hilbert separable $\mathcal{H}$ de dimensión infinita con segundo momento finito. Supongamos, sin pérdida de generalidad, que la media del elemento es 0 e indiquemos por Γ_V el operador de covarianza.

Buscamos un subespacio $\mathcal{L}$ de $\mathcal{H}$ de dimensión k de manera tal que se minimice $E(\|V - P_{\mathcal{L}}V\|^2)$, siendo $P_{\mathcal{L}}$ la proyección ortogonal sobre $\mathcal{L}$.

Sea $\mathcal{L} =< v_1, \ldots, v_k >$ donde $v_1, \ldots, v_k$ es un conjunto ortonormal de vectores que generan $\mathcal{L}$. Completemos $v_1, \ldots, v_k$ a una base ortonormal de $\mathcal{H}$, $\{v_n\}_{n\in\mathbb{N}}$.

Tenemos entonces, $\|V\|^2 = \|V - P_{\mathcal{L}}V\|^2 + \|P_{\mathcal{L}}V\|^2$. Luego, $E(\|V\|^2) = E(\|V - P_{\mathcal{L}}V\|^2) + E(\|P_{\mathcal{L}}V\|^2)$. Ahora bien, estamos interesados en minimizar el primer sumando en $\mathcal{L}$. Como la suma de ambos es constante con respecto a $\mathcal{L}$, entonces bastará con maximizar el segundo sumando, o sea,

$$\begin{aligned} E(\|P_{\mathcal{L}}V\|^2) &= \sum_{i=1}^{k} E((< v_i, V >)^2) = \sum_{i=1}^{k} \left[Var(< v_i, V >) + (E(< v_i, V >))^2 \right] \\ &= \sum_{i=1}^{k} \left[Var(< v_i, V >) + (< v_i, E(V) >)^2 \right] = \sum_{i=1}^{k} Var(< v_i, V >) = \sum_{i=1}^{k} < \Gamma_V v_i, v_i > \end{aligned}$$

puesto que $E(V) = 0$ y $Var(< v_i, V >) =< \Gamma_V v_i, v_i >$. Debemos entonces encontrar $v_1, \ldots, v_k$ vectores ortonormales en $\mathcal{H}$ de modo que maximicen $\sum_{i=1}^{k} < \Gamma_V v_i, v_i >$. Ahora bien, si suponemos Γ_V es un operador compacto, como es autoadjunto, es diagonalizable, o sea,

$$\Gamma_V = \sum_{i=1}^{\infty} \lambda_n \pi_n$$

donde $\lambda_1 \geq \lambda_2 \geq \ldots$ son los autovalores de Γ_V, ordenados de mayor a menor y π_n son las proyecciones ortogonales sobre $Ker(\Gamma_V - \lambda_n)$, es decir, el autoespacio asociado al $n-$ésimo autovalor. Notemos que como Γ_V es semidefinido positivo, $\lambda_n \geq 0$.

Se tiene entonces que

$$\sum_{i=1}^{k} < \Gamma_V v_i, v_i >= \sum_{i=1}^{k} < \sum_{n=1}^{\infty} \lambda_n \pi_n v_i, v_i >= \sum_{i=1}^{k} \sum_{n=1}^{\infty} < \lambda_n \pi_n v_i, v_i >= \sum_{n=1}^{\infty} \lambda_n \sum_{i=1}^{k} < \pi_n v_i, v_i > .$$

Los autovalores λ_n están ordenados de mayor a menor, entonces esto se maximiza considerando $v_i \in Ker(\Gamma_V - \lambda_n)$, $1 \leq n \leq k$, es decir, en cada autoespacio asociado a los k mayores autovalores de Γ_V. Tomando así v_i un autovector de Γ_V asociado al autovalor λ_i, para $1 \leq i \leq k$, podremos entonces maximizar $E(\|P_{\mathcal{L}}V\|^2)$. Luego, el subespacio $\mathcal{L}$ estará formado por los k autovectores del operador de covarianza asociados a los mayores autovalores de Γ_V, y serán esos autovectores los que se considerarán las componentes principales funcionales de V, como en el caso finito–dimensional.

4. Familias elípticas

4.1. Distribuciones esféricas y elípticas en el caso multivariado

4.1.1. Motivación

Buena parte del estudio clásico en el análisis multivariado se concentró en el caso de la normal multivariada. Sin embargo, no siempre es posible validar hipótesis de normalidad, tema de especial interés en lo que concierne a la robustez. La idea será entonces ofrecer una generalización que permita un cierto alejamiento del modelo clásico normal. Posteriormente iremos un paso más allá al extender esto al caso funcional.

4.1.2. Distribuciones Esféricas

Llamaremos $\mathcal{O}(d)$ el grupo de matrices ortogonales de orden d.

Definición 4.1. Un vector aleatorio $\mathbf{X} \in I\!R^d$ se dice que tiene una distribución esférica, o directamente que es esférico, si $\mathbf{X}$ y $\mathbf{HX}$ tienen la misma distribución para toda matriz ortogonal $\mathbf{H} \in \mathcal{O}(d)$.

Es fácil ver que si $\mathbf{X}$ es esférico y cuenta además con una función de densidad entonces, debido a la invariancia de $\mathbf{X}$ respecto a rotaciones, ésta deberá depender exclusivamente de la norma de $\mathbf{x}$, es decir, de $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x}$.

Algunos ejemplos de distribuciones esféricas son:

1. la distribución normal multivariada $N_d(\mathbf{0}, \sigma^2\mathbf{I}_d)$ con función de densidad

$$\frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2\right) ,$$

2. la distribución normal multivariada $N_d(\mathbf{0}, \mathbf{I}_d)$ contaminada con un $\epsilon\,\%$ proveniente de una normal $N_d(\mathbf{0}, \sigma^2\mathbf{I}_d)$, con densidad

$$(1-\epsilon)\frac{1}{(2\pi)^{d/2}} \exp\left(-\frac{1}{2}\|\mathbf{x}\|^2\right) + \epsilon\frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2\right) ,$$

3. la distribución multivariada t con n grados de libertad, $\mathcal{T}_{d,n}$, y densidad

$$\frac{\Gamma\left[\frac{1}{2}(n+d)\right]}{\Gamma\left(\frac{1}{2}\right)(n\pi)^{\frac{d}{2}}} \frac{1}{\left(1+\frac{1}{n}\|\mathbf{x}\|^2\right)^{\frac{n+d}{2}}} .$$

Procedemos a describir una manera de generar distribuciones esféricas. Sean $X_1, \ldots, X_d, Z$ variables aleatorias tales que $Z > 0$ y para un dado valor de $Z = z$, $X_1, \ldots, X_d$ son independientes con distribución $N(0, z)$, o sea, el vector $\mathbf{X} = (X_1, \ldots, X_d)$ es tal que $\mathbf{X}|Z = z \sim N_d(\mathbf{0}, z\ \mathbf{I}_d)$. Supongamos que Z tenga función de distribución G. Luego, la distribución conjunta de $X_1, \ldots, X_d$ es

$$f_{\mathbf{X}}(\mathbf{x}) = \int_0^{+\infty} f_{\mathbf{X}|Z=z}(\mathbf{x}) dG(z) = \int_0^{+\infty} (2\pi z)^{d/2} \exp\left(-\frac{1}{2d}\|\mathbf{x}\|^2\right) dG(z)\ ,$$

que es una distribución esférica ya que depende de $\|\mathbf{x}\|$. A este tipo de distribuciones se las denomina una *mezcla escalada* de distribuciones normales. La clase de distribuciones esféricas obtenidas al variar el G es llamada la clase de distribuciones *normales compuestas*. En particular si $P(Z = 1) = 1 - \epsilon$ y $P(Z = \sigma^2) = \epsilon$ obtenemos la normal contaminada descripta en 2).

Lema 4.1. *Sea* $\mathbf{X} = (X_1, \ldots, X_d)$ *tal que* $\mathbf{X}|Z = z \sim N_d(\mathbf{0}, z\ \mathbf{I}_d)$, *entonces* $\mathbf{X} = Z^{1/2}\mathbf{Y}$ *con* $\mathbf{Y} \sim N_d(\mathbf{0}, \mathbf{I}_d)$. *Además,* Z *e* $\mathbf{Y}$ *son independientes.*

Demostración. Sea $\mathbf{Y} = \mathbf{X}\ Z^{-\frac{1}{2}}$. Entonces $\mathbf{Y}|Z = z \sim z^{-\frac{1}{2}}\mathbf{U}$ siendo $\mathbf{U} \sim N_d(\mathbf{0}, z\ \mathbf{I}_d)$. Por lo tanto, $\mathbf{Y}|Z = z \sim N_d(\mathbf{0}, \mathbf{I}_d)$, con lo cual $\mathbf{Y}$ es independiente de Z y su distribucion es $N_d(\mathbf{0}, \mathbf{I}_d)$. $\square$

Este resultado da una manera de generar observaciones con distribución esférica, bastará generar un vector aleatorio $\mathbf{Y} \sim N_d(\mathbf{0}, \mathbf{I}_d)$ y una variable aleatoria $Z \sim G$ independiente de $\mathbf{Y}$, entonces el vector $\mathbf{X} = Z\mathbf{Y}$ será esférico con distribución *normal compuesta*.

Daremos a continuación algunos otros resultados sobre distribuciones esféricas.

Teorema 4.1. *Si* $\mathbf{X} \in I\!R^d$ *tiene una distribución esférica tal que* $P(\mathbf{X} = 0) = 0$ *y* $R = \|\mathbf{X}\|$, $\mathbf{T}(\mathbf{X}) = \|\mathbf{X}\|^{-1}\mathbf{X}$, *entonces* $\mathbf{T}$ *está uniformemente distribuido en* $\mathcal{S}^{d-1} = \{\mathbf{y} \in I\!R^d\ :\ \|\mathbf{y}\| = 1\}$ *y además* $\mathbf{T}(\mathbf{X})$ *y* R *son independientes.*

Demostración. Para toda matriz ortogonal $\mathbf{H} \in \mathcal{O}(d)$ tenemos que

$$\mathbf{T}(\mathbf{H}\mathbf{X}) = \|\mathbf{H}\mathbf{X}\|^{-1}\mathbf{H}\mathbf{X} = \|\mathbf{X}\|^{-1}\mathbf{H}\mathbf{X} = \mathbf{H}\mathbf{T}(\mathbf{X})\ ,$$

con lo cual $\mathbf{T}(\mathbf{H}\mathbf{X})$ y $\mathbf{H}\mathbf{T}(\mathbf{X})$ tienen la misma distribución. Como $\mathbf{X}$ tiene distribución esférica, $\mathbf{X}$ y $\mathbf{H}\mathbf{X}$ tienen la misma distribución y por lo tanto, $\mathbf{T}(\mathbf{X})$ y $\mathbf{T}(\mathbf{H}\mathbf{X})$ tiene idéntica distribución. Consecuentemente, $\mathbf{T}(\mathbf{X})$ y $\mathbf{H}\mathbf{T}(\mathbf{X})$ tendrán la misma distribución. Luego, $\mathbf{T}(\mathbf{X})$ es un vector aleatorio de $\mathcal{S}^{d-1}$ con distribución invariante frente a transformaciones ortogonales, pero la única distribución que satisface eso es la distribución uniforme sobre dicho conjunto, tendremos así el primer resultado.

Veamos ahora la independencia. Definamos una medida μ en $\mathcal{S}^{d-1}$ como

$$\mu(B) = P(\mathbf{T}(\mathbf{X}) \in B | R \in C)$$

con C un boreliano fijo en $I\!R$ tal que $P(R \in C) \neq 0$, y B un boreliano de $\mathcal{S}^{d-1}$. Es inmediato que μ es una medida invariante frente a transformaciones ortogonales, de done μ es una distribución uniforme sobre $\mathcal{S}^{d-1}$, o sea que tiene la misma distribución que $\mathbf{T}(\mathbf{X})$. Es decir, $\mu(B) = P(\mathbf{T}(\mathbf{X}) \in B)$, o sea, $P(\mathbf{T}(\mathbf{X}) \in B | R \in C) = P(\mathbf{T}(\mathbf{X}) \in B)$, de donde $\mathbf{T}(\mathbf{X})$ y R son independientes. $\square$

Podemos entonces probar este teorema que generaliza resultados clásicos de variables aleatorias normales.

Teorema 4.2. *Sea* $\mathbf{X} \in I\!R^d$ *con distribución esférica tal que* $P(\mathbf{X} = \mathbf{0}) = 0$.

a) Si $W = \boldsymbol{\alpha}'\mathbf{X}/\|\mathbf{X}\|$, *con* $\boldsymbol{\alpha} \in I\!R^d$, $\|\boldsymbol{\alpha}\| = 1$, *entonces*

$$Y = \frac{(d-1)^{\frac{1}{2}} W}{(1 - W^2)^{\frac{1}{2}}} \sim \mathcal{T}_{1,d-1}$$

b) Si $\mathbf{B} \in I\!R^{d\times d}$ *es una matriz simétrica e idempotente de rango* k, *entonces*

$$Z = \frac{\mathbf{X}'\mathbf{B}\mathbf{X}}{\|\mathbf{X}\|^2}$$

tiene distribución beta con parámetros $\frac{1}{2}k$ *y* $\frac{1}{2}(d-k)$.

Demostración. Usaremos el Teorema 4.1 para ambos incisos, notando que tanto Z como Y son funciones del vector aleatorio $\mathbf{T}(\mathbf{X}) = \mathbf{X}/\|\mathbf{X}\|$, que es un vector aleatorio con distribución uniforme en $\mathcal{S}_{d-1}$.

a) Observemos que $W = \boldsymbol{\alpha}'\mathbf{T}(\mathbf{X})$, luego, la distribución de W depende de la distribución uniforme en $\mathcal{S}_{d-1}$ y no cambiará entonces con la distribución particular de $\mathbf{X}$ ni con la elección de $\boldsymbol{\alpha}$ pues $\|\boldsymbol{\alpha}\| = 1$. Podemos suponer entonces sin pérdida de generalidad que $\mathbf{X} \sim N_d(\mathbf{0}, \mathbf{I}_d)$ y tomar $\boldsymbol{\alpha} = (1, 0, \ldots, 0)'$. Entonces,

$$Y = \frac{(d-1)^{\frac{1}{2}} X_1}{\left(\sum_{i=2}^d X_i^2\right)^{\frac{1}{2}}}$$

que tiene distribución $\mathcal{T}_{1,d-1}$, como queríamos probar.

b) Observemos que $Z = \mathbf{T}(\mathbf{X})'\mathbf{B}\mathbf{T}(\mathbf{X})$, con lo cual como antes, podemos suponer que $\mathbf{X} \sim N_d(\mathbf{0}, \mathbf{I}_d)$. Como $\mathbf{B}$ es una matriz simétrica e idempotente de rango k, existe una matriz $\mathbf{H} \in \mathcal{O}(d)$ tal que

$$\mathbf{H}\mathbf{B}\mathbf{H}' = \begin{bmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

Sea $\mathbf{U} = \mathbf{H}\mathbf{X}$, tenemos entonces que

$$Z = \frac{\mathbf{X}'\mathbf{B}\mathbf{X}}{\|\mathbf{X}\|^2} = \frac{\mathbf{U}'\mathbf{H}\mathbf{B}\mathbf{H}'\mathbf{U}}{\|\mathbf{U}\|^2} = \frac{\sum_{i=1}^k U_i^2}{\sum_{i=1}^d U_i^2} = \frac{V_1}{V_1 + V_2}$$

con $V_1 = \sum_{i=1}^k U_i^2 \sim \chi_k^2$, $V_2 = \sum_{i=k+1}^d U_i^2 \sim \chi_{d-k}^2$, siendo ambos independientes. Luego Z tendrá una distribución beta de parámetros $k/2$ y $(d-k)/2$. $\square$

4.1.3. Función característica de una distribución esférica

Se deducirá en esta sección una expresión para la función característica de una variable esférica. Esto será necesario para la posterior definición de la distribución elíptica, nuestro material de trabajo.

Proposición 4.1. *Sea* $\mathbf{X}$ *un vector aleatorio esféricamente distribuido, entonces la función característica de* $\mathbf{X}$*,* $\varphi_{\mathbf{X}}(\mathbf{t})$*, sólo depende de la norma de* $\mathbf{t}$*. Más precisamente,*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \int_0^\infty \psi_{\mathbf{U}}(r^2\mathbf{t}'\mathbf{t})dF_R(r),$$

donde $R = \|\mathbf{X}\|$ *y* $\psi_{\mathbf{U}}(s^2) = \varphi_{\mathbf{U}}(s\,\mathbf{e}_1)$*, siendo* $\mathbf{U}$ *un vector uniformemente distribuído en* $\mathcal{S}^{d-1}$ *con característica* $\varphi_{\mathbf{U}}$ *y* $\mathbf{e}_1$ *el primer vector de la base canónica. Si* $\Omega_d = \psi_{\mathbf{U}}$*, tendremos que* $\varphi_{\mathbf{X}}(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t}'\mathbf{t})$ *donde* $\phi_{\mathbf{X}} : \mathbb{R} \to \mathbb{R}$ *está definida por*

$$\phi_{\mathbf{X}}(s) = \int_0^\infty \Omega_d(r^2\,s)dF_R(r).$$

Demostración. Sea $\mathbf{t} \in \mathbb{R}^d$. Tenemos que $\mathbf{t}'\mathbf{X} = \|\mathbf{X}\|\mathbf{t}'(\mathbf{X}/\|\mathbf{X}\|)$. Sea $\mathbf{U}(\mathbf{X}) = \|\mathbf{X}\|^{-1}\mathbf{X}$, entonces por el Teorema 4.1, $\mathbf{U}$ está uniformemente distribuido en $\mathcal{S}^{d-1}$ y además $\mathbf{T}(\mathbf{X})$ y R son independientes. Entonces, la función característica se obtendrá como

$$\varphi_{\mathbf{X}}(\mathbf{t}) = E(\exp(i\mathbf{t}'\mathbf{X})) = E(\exp(i\|\mathbf{X}\|\,\mathbf{t}'(\mathbf{X}/\|\mathbf{X}\|))) = E(\exp(iR\,\mathbf{t}'\mathbf{U}(\mathbf{X}))).$$

Usando las propiedades de la esperanza condicional, obtenemos

$$\begin{aligned} E(\exp(iR\,\mathbf{t}'\mathbf{U}(\mathbf{X}))) &= E(E(\exp(iR\,\mathbf{t}'\mathbf{U}(\mathbf{X})|R)) \\ &= \int_0^\infty E(\exp(i\,r\,\mathbf{t}'\mathbf{U}))dF_R(r) \\ &= \int_0^\infty \varphi_{\mathbf{U}}(r\mathbf{t})dF_R(r) = \int_0^\infty \varphi_{r\mathbf{U}}(\mathbf{t})dF_R(r)\,, \end{aligned}$$

pues $\mathbf{U}$ y R son independientes, siendo $\varphi_{\mathbf{U}}$ la función característica de $\mathbf{U}$. Ahora bien, por ser $\mathbf{X}$ esférica, $\mathbf{U}$ está uniformemente distribuida en $\mathcal{S}^{d-1}$ y por lo tanto, tenemos que $\varphi_{\mathbf{U}}(s\mathbf{v}) = \varphi_{\mathbf{v}'\mathbf{U}}(s)$ no dependerá de $\mathbf{v}$ siempre que $\|\mathbf{v}\| = 1$. Además, como $\varphi_{\mathbf{U}}((-s)\mathbf{v}) = \varphi_{\mathbf{U}}(s(-\mathbf{v}))$ y $-\mathbf{v} \in \mathcal{S}^{d-1}$, si $\|\mathbf{v}\| = 1$, tampoco dependerá del signo de s sino de su valor absoluto, $|s|$, o bien, de su cuadrado s^2. Entonces, podremos hallar una función $\psi_{\mathbf{U}}(s^2)$ tal que $\varphi_{\mathbf{U}}(s\mathbf{v}) = \psi_{\mathbf{U}}(s^2)$, de modo tal que

$$\varphi_{\mathbf{U}}(\mathbf{t}) = \varphi_{\mathbf{U}}\left(\|\mathbf{t}\|\frac{\mathbf{t}}{\|\mathbf{t}\|}\right) = \psi_{\mathbf{U}}(\|\mathbf{t}\|^2) = \psi_{\mathbf{U}}(\mathbf{t}'\mathbf{t})$$

con lo cual $\varphi_{r\mathbf{U}}(\mathbf{t}) = \psi_{\mathbf{U}}(r^2\mathbf{t}'\mathbf{t})$ y por lo tanto,

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \int_0^\infty \psi_{\mathbf{U}}(r^2\mathbf{t}'\mathbf{t})dF_R(r)\ ,$$

como queríamos ver. Cabe destacar que la función característica depende únicamente de la dimensión del espacio. Luego, si definimos $\Omega_d = \psi_{\mathbf{U}}$ tendremos que $\varphi_{\mathbf{X}}(\mathbf{t}) = \phi_{\mathbf{X}}(\mathbf{t}'\mathbf{t})$ donde $\phi_{\mathbf{X}} : \mathbb{R} \to \mathbb{R}$ está definida por

$$\phi_{\mathbf{X}}(s) = \int_0^\infty \Omega_d(r^2\, s) dF_R(r).$$

□

Llamaremos "generador característico de $\mathbf{X}$" a $\phi_{\mathbf{X}}$.

La siguiente proposición ofrece una primera motivación para lo que será la definición de las distribuciones elípticas.

Proposición 4.2. *Sea* $\mathbf{X}$ *un vector aleatorio esféricamente distribuido con generador característico* $\phi_{\mathbf{X}}$. *Sea además* $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times k}$ *y* $\boldsymbol{\mu} \in \mathbb{R}^d$. *Entonces, la función característica* $\varphi_{\mathbf{Y}}$ *de* $\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{X}$ *está dada por*

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = \exp(i\, \mathbf{t}'\boldsymbol{\mu})\phi_{\mathbf{X}}(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$$

con $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}'$.

Demostración. La función característica de $\mathbf{Y}$ es

$$\varphi_{\mathbf{Y}}(\mathbf{t}) = E(\exp(i\mathbf{t}'(\boldsymbol{\mu}+\boldsymbol{\Lambda}\mathbf{X}))) = \exp(i\mathbf{t}'\boldsymbol{\mu})\varphi_{\mathbf{X}}(\boldsymbol{\Lambda}'\mathbf{t}) = \exp(i\mathbf{t}'\boldsymbol{\mu})\, \phi_{\mathbf{X}}\left((\boldsymbol{\Lambda}'\mathbf{t})'(\boldsymbol{\Lambda}'\mathbf{t})\right) = \exp(i\mathbf{t}'\boldsymbol{\mu})\, \phi_{\mathbf{X}}(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}).$$

□

4.2. Distribuciones elípticas

Definición 4.2. Sea $\mathbf{X} \in \mathbb{R}^d$ un vector aleatorio. Se dirá que $\mathbf{X}$ tiene distribución elíptica si existe un vector $\boldsymbol{\mu} \in \mathbb{R}^d$, una matriz $\boldsymbol{\Sigma} \in {}^{d\times d}$ semidefinida positiva y una función $\phi : \mathbb{R}_+ \to \mathbb{R}$ tal que la función característica de $\mathbf{X} - \boldsymbol{\mu}$ es $\varphi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t}) = \phi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$, para todo $\mathbf{t} \in \mathbb{R}^d$. Se lo indicará $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

En algunas situaciones, por simplicidad escribiremos simplemente $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ y diremos que $\mathbf{X}$ es elíptica de parámetros $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$ entendiendo que existe una función $\phi : \mathbb{R}_+ \to \mathbb{R}$ que es la generadora de la función característica de $\mathbf{X} - \boldsymbol{\mu}$.

Observemos que un vector $\mathbf{Y} \sim \varepsilon_d(0, \mathbf{I}_d, \phi)$ tiene distribución esférica. Además, por la Proposición 4.2, toda transformación afín de un vector esférico será elíptico. El siguiente resultado establece una recíproca en el caso de que la matriz de transformación sea de rango completo. En lo que sigue $X \stackrel{\mathcal{D}}{=} Y$ indicará que X e Y tienen idéntica distribución.

Proposición 4.3. $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ *con rango*$(\boldsymbol{\Sigma}) = k$ *si y sólo si*

$$\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}\ ,$$

con $\mathbf{U} \in \mathbb{R}^k$ *un vector aleatorio uniformemente distribuido en* $\mathcal{S}^{k-1}$, $\mathcal{R}$ *un variable aleatoria no negativa independiente de* $\mathbf{U}$, $\boldsymbol{\mu} \in \mathbb{R}^d$ *y* $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times k}$ *con rango*$(\boldsymbol{\Lambda}) = k$. *Es decir,* $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{Y}$ *con* $\mathbf{Y} \sim \varepsilon_k(0, \mathbf{I}_k, \phi)$, *esférico.*

Demostración. Probaremos solamente que si $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ con rango$(\boldsymbol{\Sigma}) = k$ entonces $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + R\boldsymbol{\Lambda}\mathbf{U}$, puesto que la recíproca se deduce inmediatamente de la Proposición 4.2.

Como $\boldsymbol{\Sigma}$ es semidefinida positiva de rango k, existe $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times k}$ tal que $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}$, es decir, si $k = d$, $\boldsymbol{\Lambda}$ una raíz de $\boldsymbol{\Sigma}$. Consideramos $\boldsymbol{\Lambda}^-$ una pseudo-inversa de $\boldsymbol{\Lambda}$ y definamos $\mathbf{Y} = \boldsymbol{\Lambda}^-(\mathbf{X} - \boldsymbol{\mu})$. Por ser una pseudo-inversa, se verifica que $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^-\boldsymbol{\Lambda} = \boldsymbol{\Lambda}$ y más aún, $\boldsymbol{\Lambda}^-\boldsymbol{\Lambda} \in \mathbb{R}^{k\times k}$ es inversible por ser de rango completo. Es fácil ver que $\boldsymbol{\Lambda}^-\boldsymbol{\Lambda} = \mathbf{I}_k$. Entonces, como $\varphi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t}) = \phi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$ se tiene

$$\begin{aligned}
\varphi_{\mathbf{Y}}(\mathbf{t}) &= \varphi_{\mathbf{X}-\boldsymbol{\mu}}((\boldsymbol{\Lambda}^-)'\mathbf{t}) = \phi(((\boldsymbol{\Lambda}^-)'\mathbf{t})'\boldsymbol{\Sigma}(\boldsymbol{\Lambda}^-)'\mathbf{t}) \\
&= \phi(\mathbf{t}'\boldsymbol{\Lambda}^-\boldsymbol{\Sigma}(\boldsymbol{\Lambda}^-)'\mathbf{t}) = \phi(\mathbf{t}'\boldsymbol{\Lambda}^-(\boldsymbol{\Lambda}\boldsymbol{\Lambda}')(\boldsymbol{\Lambda}')^-\mathbf{t}) \\
&= \phi(\mathbf{t}'(\boldsymbol{\Lambda}^-\boldsymbol{\Lambda})(\boldsymbol{\Lambda}^-\,\boldsymbol{\Lambda})'\mathbf{t}) = \phi(\mathbf{t}'\mathbf{t}),
\end{aligned}$$

luego $\mathbf{Y} \in \mathbb{R}^k$ tiene una distribución esférica con generador característico ϕ y que puede ser representado por un $R\mathbf{U}$ estocásticamente, donde $\mathcal{R} = \|\mathbf{Y}\|$ y $\mathbf{U}$ son independientes y $\mathbf{U}$ es uniforme en $\mathcal{S}^{k-1}$. Por lo tanto, $\mathbf{X} = \boldsymbol{\mu} + \boldsymbol{\Lambda}\mathbf{Y} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}$. □

Debido a la matriz de transformación $\boldsymbol{\Lambda}$, el vector aleatorio esférico $\mathbf{U}$ producirá superificies de nivel elipsoidales, mientras que la variable $\mathcal{R}$ determina la forma de la distribución y en particular, el peso de sus colas. Llamaremos a $\mathcal{R}$ la variable generadora de $\mathbf{X}$. Notemos que $\boldsymbol{\mu}$ indica la posición de $\mathbf{X}$ mientras que la matriz $\boldsymbol{\Sigma}$ se llama la matriz de dispersión o de escala de $\mathbf{X}$. Por ejemplo, si $\mathbf{X} \sim N_d(\mathbf{0}, \mathbf{I}_d)$ su variable generadora $\mathcal{R} \sim \sqrt{\chi^2_d}$. En particular, para generar observaciones con distribución elíptica, bastará generar un vector aleatorio $\mathbf{X}_1 \sim N_d(\mathbf{0}, \boldsymbol{\Sigma})$ y una variable aleatoria $Z \sim G$ independiente de $\mathbf{X}_1$, entonces el vector $\mathbf{X} = \boldsymbol{\mu} + Z\,\mathbf{X}_1$ tendrá distribución elíptica.

Una buena propiedad de las distribuciones elípticas es que su densidad se puede expresar a través de la densidad de la variable generadora ($\mathcal{R}$), suponiendo que esta última es absolutamente continua. Dada una matriz $\boldsymbol{\Lambda} \in \mathbb{R}^{d\times k}$, de rango k, sean $\lambda_1(\boldsymbol{\Lambda}) \geq \lambda_2(\boldsymbol{\Lambda}) \geq \ldots \geq \lambda_k(\boldsymbol{\Lambda}) > 0$ los autovalores de $\boldsymbol{\Lambda}'\boldsymbol{\Lambda}$. En lo que sigue, $\det(\boldsymbol{\Lambda}) = \prod_{i=1}^k \sqrt{\lambda_i} = (\det(\boldsymbol{\Lambda}'\boldsymbol{\Lambda}))^{\frac{1}{2}}$. Si la matriz $\mathbf{A} \in \mathbb{R}^{d\times d}$ indicaremos $|\mathbf{A}| = \det(\mathbf{A})$. En lo que sigue, como en Frahm (2004), permitiremos que las funciones de densidad estén definidas no sólo en $\mathbb{R}^d$ cuando $\mathbf{X} \in \mathbb{R}^d$ sino en subespacios de dimensión menor o en superficies. Tenemos el siguiente resultado.

Teorema 4.3. *Sea $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$, con $\boldsymbol{\mu} \in \mathbb{R}^d$ y $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ una matriz semidefinida positiva tal que rango$(\boldsymbol{\Sigma}) = k$. Supongamos que $\mathcal{R}$ es una variable aleatoria absolutamente continua con densidad $f_{\mathcal{R}}$ y sea $\mathcal{S}_{\boldsymbol{\Lambda}}$ el subespacio de $\mathbb{R}^d$ generado por las columnas $\boldsymbol{\Lambda}$. Luego, la función de densidad de $\mathbf{X}$ está dada por*

$$f_{\mathbf{X}}(\mathbf{x}) = |det(\boldsymbol{\Lambda})|^{-1} \cdot g_{\mathcal{R}}((\mathbf{x} - \boldsymbol{\mu})'(\boldsymbol{\Sigma})^-(\mathbf{x} - \boldsymbol{\mu}))\ ,\ \forall \mathbf{x} \text{ tal que } \mathbf{x} - \boldsymbol{\mu} \in \mathcal{S}_{\boldsymbol{\Lambda}} \setminus \mathbf{0},$$

donde

$$g_{\mathcal{R}}(t) = \frac{\Gamma(\frac{k}{2})}{2\pi^{\frac{k}{2}}}\, t^{-\frac{k-1}{2}}\, f_{\mathcal{R}}(\sqrt{t}),\ \ t > 0\ .$$

Debido a que rango$(\boldsymbol{\Sigma}) = k$ entonces valdrá la Proposición 4.3 por lo que $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}$ con $\mathbf{U} \in \mathbb{R}^k$ uniforme en $\mathcal{S}^{k-1}$, $\mathcal{R}$ una variable nonegativa independiente de $\mathbf{U}$ y $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}$.

Demostración. $\Gamma(\frac{k}{2})/(2\pi^{k/2})$ corresponde a la densidad de una variable uniforme distribuida en $\mathcal{S}^{k-1}$, es decir, de $\mathbf{U}$. Entonces, la densidad conjunta de $(\mathcal{R}, \mathbf{U})$, utilizando la independencia, será

$$f_{(\mathcal{R},\mathbf{U})}(r, \mathbf{u}) = \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}}\, f_{\mathcal{R}}(r), \quad r > 0, \mathbf{u} \in \mathcal{S}^{k-1}.$$

Definamos $\mathbf{Y} = \mathcal{R}\mathbf{U}$, procedemos a calular su densidad utilizando la transformación $h : (0, \infty) \times \mathcal{S}^{k-1} \to \mathbb{R}^k \setminus \mathbf{0}$, $h(r, \mathbf{u}) = r\mathbf{u} = \mathbf{y}$. Observemos que h es injectiva, luego biyectiva con su imagen, con lo cual por la fórmula de cambio de variables obtendremos que

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{(\mathcal{R},\mathbf{U})}(h^{-1}(\mathbf{y})) \cdot |J_h|^{-1}, \quad \mathbf{y} \neq \mathbf{0},$$

siendo J_h el determinante jacobiano de $\partial h(r, \mathbf{u})/\partial(r, \mathbf{u})$. Definimos la hiperesfera de radio r como $\mathcal{S}_r^{k-1} = \{\mathbf{x} \in \mathbb{R}^k, \|\mathbf{x}\| = r > 0\}$. Puesto que la derivada parcial $\partial r\mathbf{u}/\partial r$ tienen longitud unitaria y es ortogonal a cada plano tangente $\partial r\mathbf{u}/\partial\mathbf{u}'$ en $\mathcal{S}_r^{k-1}$, tenemos que

$$|J_h| = \det\left(\begin{bmatrix} 1 & \mathbf{0}' \\ \mathbf{0} & r\,\mathbf{I}_{k-1} \end{bmatrix}\right) = r^{k-1} = \|\mathbf{y}\|^{k-1}.$$

Más aún, $h^{-1}(\mathbf{y}) = (\|\mathbf{y}\|, \mathbf{y}/\|\mathbf{y}\|)$ luego la densidad de $\mathbf{Y}$ será

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{(\mathcal{R},\mathbf{U})}(\|\mathbf{y}\|, \mathbf{y}/\|\mathbf{y}\|)\ \|\mathbf{y}\|^{k-1} = \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}}\ \|\mathbf{y}\|^{-(k-1)}\ f_{\mathcal{R}}(\|\mathbf{y}\|), \quad \mathbf{y} \neq \mathbf{0}.$$

Definamos ahora $q : \mathbb{R}^k \backslash \mathbf{0} \to \mathcal{S}_{\mathbf{\Lambda}} \backslash \{\mu\}$, como $q(\mathbf{y}) = \boldsymbol{\mu} + \mathbf{\Lambda}\mathbf{y}$, llamaremos $\mathbf{x}$ a $q(\mathbf{y})$. La transformación q será inyectiva puesto que $\mathbf{\Lambda}^-\mathbf{\Lambda} = \mathbf{I}_k$. El valor absoluto del determinante jacobiano de $\partial(\boldsymbol{\mu} + \mathbf{\Sigma}\mathbf{y})\partial\mathbf{y}$ será $|J_q| = |\det(\mathbf{\Lambda})|$, y luego la densidad de $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathbf{\Lambda}\mathbf{Y} = \boldsymbol{\mu} + \mathcal{R}\mathbf{\Lambda}\mathbf{U}$ será

$$f_{\mathbf{X}}(\mathbf{x}) = f_{\mathbf{Y}}(q^{-1}(\mathbf{x}))\ |J_q|^{-1} = f_{\mathbf{Y}}(\mathbf{\Lambda}^-(\mathbf{x} - \boldsymbol{\mu}))|\det(\mathbf{\Lambda})|^{-1}, \quad \mathbf{x} \in \mathcal{S}_\Lambda \setminus \{\boldsymbol{\mu}\}.$$

Entonces, la función de densidad de $\mathbf{X}$ resulta ser

$$f_{\mathbf{X}}(\mathbf{x}) = |\det(\mathbf{\Lambda})|^{-1}\ \frac{\Gamma(\frac{k}{2})}{2\pi^{k/2}}\ \|\mathbf{\Lambda}^-(\mathbf{x} - \boldsymbol{\mu})\|^{-(k-1)}\ f_{\mathcal{R}}(\|\mathbf{\Lambda}^-(\mathbf{x} - \boldsymbol{\mu})\|),$$

con $\mathbf{x} \in \mathcal{S}_{\mathbf{\Lambda}} \setminus \{\boldsymbol{\mu}\}$. Como $\|\mathbf{\Lambda}^-(\mathbf{x} - \boldsymbol{\mu})\|^2 = (\mathbf{x} - \boldsymbol{\mu})'\mathbf{\Lambda}'^-\mathbf{\Lambda}^-(\mathbf{x} - \boldsymbol{\mu})$ y

$$\mathbf{\Lambda}'^-\mathbf{\Lambda}^- = \big((\mathbf{\Lambda}\mathbf{\Lambda}')^-\mathbf{\Lambda}\big)\ \big((\mathbf{\Lambda}\mathbf{\Lambda}')^-\mathbf{\Lambda}\big)' = (\mathbf{\Sigma}^-\mathbf{\Lambda})(\mathbf{\Sigma}^-\mathbf{\Lambda})' = \mathbf{\Sigma}^-\mathbf{\Sigma}\mathbf{\Sigma}^- = \mathbf{\Sigma}^-\ ,$$

se obtiene la fórmula deseada. □

La función $g_{\mathcal{R}}$ se la llamará "generadora de densidades" de $\mathbf{X}$. Notemos que las superficies de nivel de esta función corresponden a superficies elipsoidales.

Para el caso en que la matriz $\mathbf{\Sigma}$ sea inversible, tenemos entonces el siguiente corolario.

Corolario 4.1. *Sea* $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$*, con* $\boldsymbol{\mu} \in \mathbb{R}^d$ *y* $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ *una matriz definida positiva, o sea, tal que rango*$(\Sigma) = d$*. Supongamos además que* $\mathcal{R}$ *es una variable aleatoria absolutamente continua. Luego, la función de densidad de* $\mathbf{X}$ *está dada por*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}}\, g_{\mathcal{R}}\left((\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right), \quad \mathbf{x} \neq \boldsymbol{\mu},$$

donde

$$g_{\mathcal{R}}(t) = \frac{\Gamma(\frac{d}{2})}{2\pi^{\frac{d}{2}}}\, t^{-\frac{d-1}{2}} \cdot f_{\mathcal{R}}(\sqrt{t}),\ t > 0$$

con $f_{\mathcal{R}}$ *la función de densidad de* $\mathcal{R}$*.*

Como rango$(\Sigma) = d$ luego valdrá la Proposición 4.3 por lo que $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}$ con $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}$.

Demostración. Consiste en aplicar el teorema anterior luego de substituir k por d y considerando que

$$|\det(\boldsymbol{\Lambda})| = \sqrt{\det(\boldsymbol{\Lambda}'\boldsymbol{\Lambda})} = \sqrt{\det(\boldsymbol{\Lambda}\boldsymbol{\Lambda}')} = \sqrt{\det(\Sigma)} = \sqrt{|\Sigma|},$$

puesto que $\boldsymbol{\Lambda}$ es inversible. $\square$

4.2.1. Momentos de una distribución elíptica

Teorema 4.4. *Sea* $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$*, con* $\boldsymbol{\mu} \in \mathbb{R}^d$ *y* $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ *una matriz semidefinida positiva tal que rango*$(\boldsymbol{\Sigma}) = k$*.*

a) Si existe $E(\mathbf{X})$ *y* $E(\mathcal{R})$*, se cumple que* $E(\mathbf{X}) = \boldsymbol{\mu}$*. O sea, el parámetro* $\boldsymbol{\mu}$ *es la esperanza de* $\mathbf{X}$ *en caso de existir.*

b) Si existen los momentos de orden 2*,*

$$Var(\mathbf{X}) = \alpha\boldsymbol{\Sigma} = \frac{E(\mathcal{R}^2)}{k}\,\boldsymbol{\Sigma}.$$

Como rango$(\boldsymbol{\Sigma}) = k$ valdrá entonces la Proposición 4.3 por lo que $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}$ con $\mathbf{U} \in \mathbb{R}^k$ uniforme en $\mathcal{S}^{k-1}$, $\boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \boldsymbol{\Sigma}$ y $\mathcal{R}$ una variable nonegativa independiente de $\mathbf{U}$.

Demostración. Tenemos que $\mathbf{X} = \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}$, con $\mathbf{U}$ uniforme sobre la esfera $\mathcal{S}^{k-1}$, $\mathcal{R}$ variable no negativa independiente de $\mathbf{U}$ y $\boldsymbol{\Lambda}$ una raíz de $\boldsymbol{\Sigma}$.
a) Entonces, si existe $E(\mathbf{X})$ y $E(\mathcal{R})$, se cumple que $E(\mathbf{X}) = E(\boldsymbol{\mu}+\mathcal{R}\boldsymbol{\Lambda}\mathbf{U}) = \boldsymbol{\mu}+\boldsymbol{\Lambda}E(\mathcal{R})E(\mathbf{U})$, puesto que $\mathbf{U}$ y $\mathcal{R}$ son independientes. Como $E(\mathbf{U}) = 0$ por ser $\mathbf{U}$ esférico, obtenemos que $E(\mathbf{X}) = \boldsymbol{\mu}$.
b) Supongamos ahora que exiten los momentos de orden dos y calculemos la covarianza de $\mathbf{X}$. Utilizando la independencia entre $\mathbf{U}$ y $\mathcal{R}$ obtenemos

$$Var(\mathbf{X}) = Var(\boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}) = Var(\mathcal{R}\boldsymbol{\Lambda}\mathbf{U}) = E\left((\mathcal{R}\boldsymbol{\Lambda}\mathbf{U})(\mathcal{R}\boldsymbol{\Lambda}\mathbf{U})'\right) = E(\mathcal{R}^2)\,\boldsymbol{\Lambda}E(\mathbf{U}\mathbf{U}')\boldsymbol{\Lambda}',$$

donde hemos supuesto que $\mathcal{R}$ tiene segundo momento finito. Ahora bien, $\mathbf{Z} = \sqrt{\chi^2_k}\mathbf{U} \sim N_k(0, \mathbf{I}_k)$, pues la variable generadora de una $\mathbf{Z} \sim N_k(0, \mathbf{I}_k)$ tiene la misma distribución que $\sqrt{\chi^2_k}$. Luego,

$$\mathbf{I}_k = Var(\mathbf{Z}) = E\left((\sqrt{\chi^2_k}\mathbf{U})(\sqrt{\chi^2_k}\mathbf{U})' \right) = E(\chi^2_k)\ E(\mathbf{U}\mathbf{U}') = k \cdot E(\mathbf{U}\mathbf{U}') \ ,$$

de donde, $E(\mathbf{U}\mathbf{U}') = \mathbf{I}_k/k$, por lo tanto

$$Var(\mathbf{X}) = \frac{E(\mathcal{R}^2)}{k}\ \mathbf{\Sigma}.$$

□

Cabe aclarar que k no es necesariamente la dimensión de $\mathbf{X}$ o de $\mathbf{\Sigma}$ sino la cantidad de componentes de $\mathbf{U}$. Tenemos por lo tanto, que el parámetro de dispersión es una constante positiva multiplicada por la matriz de covarianza de $\mathbf{X}$. Más aún, es fácil ver que la constante $\alpha = E(\mathcal{R}^2)/k$ es igual a $-2\phi'(0)$.

Para lograr que ambos parámetros coincidan se debería re-definir $\mathcal{R}$, es decir cambiar la función ϕ. Nosotros consideraremos la parametrización dada por ϕ, o sea, salvo una constante de proporcionalidad.

4.2.2. Otras propiedades de las distribuciones elípticas: distribución condicional

Proposición 4.4. *Todas las distribuciones marginales de un vector aleatorio* $\mathbf{X} \sim \varepsilon_d(\boldsymbol{\mu}, \mathbf{\Sigma}, \phi_{\mathbf{X}})$ *elíptico son a su vez elípticos con generador característico* $\phi_{\mathbf{X}}$.

Demostración. Consideramos $\mathbf{X} \in \mathbb{R}^d$ particionado como sigue

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}, \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} \\ \mathbf{\Sigma}_{21} & \mathbf{\Sigma}_{22} \end{pmatrix}$$

con $\mathbf{X}_1, \boldsymbol{\mu}_1 \in \mathbb{R}^k$ y $\mathbf{\Sigma}_{11} \in \mathbb{R}^{k \times k}$. Calculemos la función característica de $\mathbf{X}_1$. Sea $\tilde{\mathbf{t}} = (\mathbf{t}', \mathbf{0}')'$

$$\begin{aligned} \varphi_{\mathbf{X}_1}(\mathbf{t}) &= E(\exp(i\mathbf{t}'\mathbf{X}_1)) = E(\exp(i(\mathbf{t}', \mathbf{0}')(\mathbf{X}_1, \mathbf{X}_2))) = E(\exp(i\tilde{\mathbf{t}}'\mathbf{X})) \\ &= \varphi_{\mathbf{X}}(\tilde{\mathbf{t}}) = \exp(i\tilde{\mathbf{t}}'(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2))\phi(\tilde{\mathbf{t}}'\mathbf{\Sigma}\tilde{\mathbf{t}}) = \exp(i\mathbf{t}'\mu_1)\phi(\mathbf{t}'\mathbf{\Sigma}_{11}\mathbf{t}). \end{aligned}$$

de donde X_1 es una variable $\varepsilon(\boldsymbol{\mu}_1, \mathbf{\Sigma}_{11}, \phi_{\mathbf{X}})$. □

Es interesante destacar que el generador característico ϕ de $\mathbf{X}_1$ es el mismo que el de $\mathbf{X}$. Tendremos así el siguiente corolario:

Corolario 4.2. *Si* $\mathbf{X}$ *es un vector aleatorio elítipico tal que alguna de sus marginales es normal, entonces* $\mathbf{X}$ *es normal multivariada.*

Demostración. Si una de las marginales supongamos $\mathbf{X}_1$ es normal multivariada, luego eso implica que en su función característica $\phi_{\mathbf{X}_1}$ será la correspondiente a la de una normal. Pero como $\phi_{\mathbf{X}} = \phi_{\mathbf{X}_1}$ tendremos así que $\mathbf{X}$ también es normal. □

Teorema 4.5. *Sea $\mathbf{X} \in \mathbb{R}^d$ un vector aleatorio con distribución $\varepsilon_d(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ y $\mathbf{B} \in \mathbb{R}^{k\times d}$. Luego, $\mathbf{BX}$ será un vector aleatorio con distribución $\varepsilon_d(\mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}', \phi)$.*

Demostración. Sea $\mathbf{Y} = \mathbf{BX}$, calculemos la función característica de $\mathbf{Y}$.

$$\begin{aligned} \varphi_{\mathbf{Y}}(\mathbf{t}) &= E\left(\exp(i\mathbf{t}'\mathbf{Y})\right) = E\left(\exp(i\mathbf{t}'\mathbf{BX})\right) = E\left(\exp(i(\mathbf{B}'\mathbf{t})'\mathbf{X})\right) = \varphi_{\mathbf{X}}(\mathbf{B}'\mathbf{t}) \\ &= \exp(i(\mathbf{B}'\mathbf{t})'\boldsymbol{\mu})\phi_{\mathbf{X}}((\mathbf{B}'\mathbf{t})'\boldsymbol{\Sigma}(\mathbf{B}'\mathbf{t})) = \exp(i\mathbf{t}'(\mathbf{B}'\boldsymbol{\mu}))\phi_{\mathbf{X}}(\mathbf{t}'(\mathbf{B}\boldsymbol{\Sigma}\mathbf{B}')\mathbf{t}) \end{aligned}$$

y obtenemos así lo deseado. □

Procedemos ahora a estudiar propiedades sobre la distribución condicional de un vector elíptico. Estos resultados serán de vital importancia para el capítulo que viene, cuando consideremos puntos autoconsistentes. Como primer paso daremos una caracterización de la distribución condicional de un vector esférico, posteriormente esto nos permitirá inferir resultados sobre el caso elíptico.

Teorema 4.6. *Sea $\mathbf{X} \stackrel{\mathcal{D}}{=} \mathcal{R}\mathbf{U}^{(d)} \sim \varepsilon(0, \mathbf{I}_d, \phi)$ (o sea, con distribución esférica) donde $\mathbf{U}^{(d)}$ tiene distribución uniforme sobre la esfera $\mathcal{S}^{d-1}$ y $\mathcal{R}$ una variable aleatoria no negativa independiente de $\mathbf{U}^{(d)}$. Consideremos $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2)$ con $\mathbf{X}_1$ el vector formado por las primeras k componentes de $\mathbf{X}$. Suponiendo que el vector aleatorio condicional $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$ exista, se cumple que $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$ tiene también una distribución esférica y podrá ser estocásticamente representado como*

$$\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1 \stackrel{\mathcal{D}}{=} \mathcal{R}^*\mathbf{U}^{(d-k)} \ ,$$

donde $\mathbf{U}^{(d-k)}$ es un vector aleatorio uniformemente distribuido sobre $\mathcal{S}^{d-k-1}$ y

$$\mathcal{R}^* \stackrel{\mathcal{D}}{=} \mathcal{R}\sqrt{1-\beta}\,\Big|(\mathcal{R}\sqrt{\beta}\mathbf{U}^{(k)} = \mathbf{x}_1) \ ,$$

con $\beta \sim Beta(\frac{k}{2}, \frac{d-k}{2})$, $\mathcal{R}$, β,$\mathbf{U}^{(k)}$ y $\mathbf{U}^{(d-k)}$ mutuamente independientes y $\mathbf{U}^{(k)}$ uniformemente distribuído en $\mathcal{S}^{k-1}$.

Demostración. Sea $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2) \sim N_d(0, \mathbf{I}_d)$, luego

$$\mathbf{U}^{(d)} = \begin{pmatrix} \mathbf{U}_1^{(d)} \\ \mathbf{U}_2^{(d)} \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} \frac{\|\mathbf{Z}_1\|}{\|\mathbf{Z}\|} \cdot \frac{\mathbf{Z}_1}{\|\mathbf{Z}_1\|} \\ \frac{\|\mathbf{Z}_2\|}{\|\mathbf{Z}\|} \cdot \frac{\mathbf{Z}_2}{\|\mathbf{Z}_2\|} \end{pmatrix} = \begin{pmatrix} \sqrt{\beta}\,\mathbf{U}^{(k)} \\ \sqrt{1-\beta}\,\mathbf{U}^{(d-k)} \end{pmatrix}$$

donde

$$\mathbf{U}^{(k)} = \frac{\mathbf{Z}_1}{\|\mathbf{Z}_1\|}, \quad \mathbf{U}^{(d-k)} = \frac{\mathbf{Z}_2}{\|\mathbf{Z}_2\|}, \quad \sqrt{\beta} = \frac{\|\mathbf{Z}_1\|}{\|\mathbf{Z}\|}.$$

Como $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2) \sim N_d(0, \mathbf{I}_d)$, $\mathbf{U}^{(k)}$ y $\mathbf{U}^{(d-k)}$ son independientes. Por otra parte, por el Teorema 4.2, $\beta \sim \text{Beta}(k/2, (d-k)/2)$. Más aún, $\beta^2 = \|\mathbf{Z}_1\|^2/(\|\mathbf{Z}_1\|^2 + \|\mathbf{Z}_2\|^2)$, como $\mathbf{U}^{(k)}$ es independiente de $\mathbf{Z}_2$ y de $\|\mathbf{Z}_1\|$, lo es de β. De igual forma, $\mathbf{U}^{(d-k)}$ es independiente de $\mathbf{Z}_a$ y de $\|\mathbf{Z}_2\|$, por lo tanto, de β.

Como $\mathbf{X} \stackrel{\mathcal{D}}{=} \mathcal{R}\mathbf{U}^{(d)}$ tendremos entonces

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} \mathcal{R}\sqrt{\beta}\mathbf{U}^{(k)} \\ \mathcal{R}\sqrt{1-\beta}\mathbf{U}^{(d-k)} \end{pmatrix}$$

en donde $\mathcal{R}$, β, $U^{(k)}$ y $U^{(d-k)}$ son mutuamente independientes y $\beta \sim \text{Beta}(\frac{k}{2}, \frac{d-k}{2})$. Basta ahora condicionar para tener el resultado deseado. □

Extenderemos ahora el resultado para el caso elíptico.

Teorema 4.7. *Sea $\mathbf{X} \sim \varepsilon(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ con $\boldsymbol{\mu} \in \mathbb{R}^d$ y $\boldsymbol{\Sigma} \in \times$ una matriz semidefinida positiva con rango$(\boldsymbol{\Sigma}) = r$. Además, sea $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$ y $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$ con $\mathbf{X}_1$ el vector de las primeras k coordenadas de $\mathbf{X}$ ($k < r$) de modo tal que $\boldsymbol{\Sigma}_{11}$ sea no singular. Entonces, por la Proposición 4.3, tendremos que $\mathbf{X}$ se puede representar estocásticamente por $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\mathbf{C}\mathbf{U}^{(r)}$ con*

$$\mathbf{C} = \begin{bmatrix} \mathbf{C}_{11} & 0 \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{bmatrix} \in \mathbb{R}^{d \times r}$$

la raíz de Cholesky generalizada de $\boldsymbol{\Sigma}$ con submatrices $\mathbf{C}_{11} \in \mathbb{R}^{k \times k}$, $\mathbf{C}_{21} \in \mathbb{R}^{(d-k) \times k}$ y $\mathbf{C}_{22} \in \mathbb{R}^{(d-k) \times (r-k)}$. Suponiendo que el vector aleatorio condicional $\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1$ existe, entonces $\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1$ tendrá también una distribución elíptica que puede ser representada estocásticamente por

$$\mathbf{X}_2 | \mathbf{X}_1 = \mathbf{x}_1 \stackrel{\mathcal{D}}{=} \boldsymbol{\mu}^* + \mathcal{R}^* \mathbf{C}_{22} \mathbf{U}^{(r-k)},$$

con $\mathbf{U}^{(r-k)}$ uniformemente distribuida en $\mathcal{S}^{r-k-1}$ y

$$\mathcal{R}^* \stackrel{\mathcal{D}}{=} \mathcal{R}\sqrt{1-\beta} \Big| (\mathcal{R}\sqrt{\beta}\mathbf{U}^{(k)} = \mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)),$$

donde $\mathbf{U}^{(k)}$ se encuentra uniformemente distribuida en $\mathcal{S}^{k-1}$ y $\beta \sim$ Beta$(\frac{k}{2}, \frac{r-k}{2})$ con $\mathcal{R}$, β, $\mathbf{U}^{(k)}$ y $\mathbf{U}^{(r-k)}$ mutuamente independientes. Por otra parte, el vector de posición corresponde a $\boldsymbol{\mu}^* = \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$.

Demostración. Definamos $\mathbf{U}^{(r)}$ como en la demostración anterior, substituyendo d por r. Como $\mathbf{X} \stackrel{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\mathbf{C}\mathbf{U}^{(r)}$ obtenemos fácilmente que

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} \stackrel{\mathcal{D}}{=} \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{C}_{11}\mathcal{R}\mathbf{U}_1^{(r)} \\ \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathcal{R}\mathbf{U}_1^{(r)} + \mathbf{C}_{22}\mathcal{R}\mathbf{U}_2^{(r)} \end{pmatrix}.$$

Utilizando la caracterización de $\mathbf{U}^{(r)} = (\mathbf{U}_1^{(r)}, \mathbf{U}_2^{(r)})$ dada en el Teorema anterior, tendremos que

$$\mathbf{X} \stackrel{\mathcal{D}}{=} \begin{pmatrix} \boldsymbol{\mu}_1 + \mathbf{C}_{11}\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)} \\ \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)} + \mathbf{C}_{22}\mathcal{R}\sqrt{1-\beta}\,\mathbf{U}^{(r-k)} \end{pmatrix}.$$

Bajo la condición $\mathbf{X}_1 = \mathbf{x}_1$, despejando $\mathbf{x}_1 = \boldsymbol{\mu}_1 + \mathbf{C}_{11}\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)}$ se obtendrá que el vector aleatorio $\mathcal{R}\sqrt{\beta}\mathbf{U}^{(k)}$ se degenera en la constante $\mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$, pues $\mathbf{C}_{11}$ es inversible por hipótesis ya que $\mathbf{C}_{11}\mathbf{C}_{11}' = \boldsymbol{\Sigma}_{11}$. Definimos entonces

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)} = \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1).$$

Tendremos entonces que $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$ tendrá la misma distribución que

$$\begin{aligned}
& \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)} + \mathbf{C}_{22}\mathcal{R}\sqrt{1-\beta}\,\mathbf{U}^{(r-k)}\Big|(\mathbf{X}_1 = \mathbf{x}_1) \\
\overset{\mathcal{D}}{=}\; & \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) + \mathbf{C}_{22}\mathcal{R}\sqrt{1-\beta}\,\mathbf{U}^{(r-k)}\Big|(\mathbf{X}_1 = \mathbf{x}_1) \\
\overset{\mathcal{D}}{=}\; & \boldsymbol{\mu}^* + \mathbf{C}_{22}\mathcal{R}\sqrt{1-\beta}\,\mathbf{U}^{(r-k)}|(\mathbf{X}_1 = \mathbf{x}_1) \\
\overset{\mathcal{D}}{=}\; & \boldsymbol{\mu}^* + \mathbf{C}_{22}\mathcal{R}^*\sqrt{1-\beta}\,\mathbf{U}^{(r-k)}
\end{aligned}$$

donde $\mathcal{R}^* \overset{\mathcal{D}}{=} \mathcal{R}\sqrt{1-\beta}\,\Big|(\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)} = \mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1))$, como queríamos probar. □

El siguiente corolario nos permite expresar el resultado anterior sin necesidad de recurrir a la raíz generalizada, expresando la distribución condicional en términos de los componentes de $\boldsymbol{\Sigma}$.

Corolario 4.3. *Sea $\mathbf{X} \sim \varepsilon(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$ con $\boldsymbol{\mu} \in \mathbb{R}^d$ y $\boldsymbol{\Sigma} \in \times$ una matriz semidefinida positiva con rango$(\boldsymbol{\Sigma}) = r$. Además, sea $\boldsymbol{\mu} = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')'$ y $\mathbf{X} = (\mathbf{X}_1', \mathbf{X}_2')'$ con $\mathbf{X}_1$ el vector de las primeras k coordenadas de $\mathbf{X}$ ($k < r$) de modo tal que $\boldsymbol{\Sigma}_{11}$ sea no singular donde*

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}' = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \in \mathbb{R}^{d\times d}$$

con submatrices $\boldsymbol{\Sigma}_{11} \in \mathbb{R}^{k\times k}$, $\boldsymbol{\Sigma}_{21} \in \mathbb{R}^{(d-k)\times k}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}' \in \mathbb{R}^{k\times(d-k)}$ y $\boldsymbol{\Sigma}_{22} \in \mathbb{R}^{(d-k)\times(r-k)}$ (luego, por la Proposición 4.3, tendremos que $\mathbf{X} \overset{\mathcal{D}}{=} \boldsymbol{\mu} + \mathcal{R}\boldsymbol{\Lambda}\mathbf{U}^{(r)}$). Suponiendo que el vector aleatorio condicional $\mathbf{X}_2|\mathbf{X}_1 = \mathbf{x}_1$ existe, tendrá también una distribución elíptica $\varepsilon_{d-k}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, \phi^*)$ con

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$$

y

$$\boldsymbol{\Sigma}^* = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12},$$

y ϕ^* corresponde al generador característico de $\mathcal{R}^*\mathbf{U}^{(r-k)}$ con

$$\mathcal{R}^* \overset{\mathcal{D}}{=} \mathcal{R}\sqrt{1-\beta}\,\Big|(\mathcal{R}\sqrt{\beta}\,\mathbf{U}^{(k)} = \mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)).$$

Aquí $\mathbf{C}_{11}$ corresponde a la raíz de Cholesky de $\boldsymbol{\Sigma}_{11}$, $\mathbf{U}^{(k)}$ está uniformemente distribuido en $\mathcal{S}^{k-1}$ y $\beta \sim Beta(\frac{k}{2}, \frac{r-k}{2})$ con $\mathcal{R}$, β, $\mathbf{U}^{(k)}$ y $\mathbf{U}^{(r-k)}$ mutuamente independientes.

Demostración. Consideramos el teorema anterior y hacemos notar que $\mathbf{C}_{21}\mathbf{C}_{11}^{-1} = (\mathbf{C}_{21}\mathbf{C}_{11}')((\mathbf{C}_{11}^{-1})'\mathbf{C}_{11}^{-1}) = \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}$, luego

$$\boldsymbol{\mu}^* = \boldsymbol{\mu}_2 + \mathbf{C}_{21}\mathbf{C}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1) = \boldsymbol{\mu}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1).$$

Además, $\mathbf{C}_{11}\mathbf{C}_{11}' = \mathbf{\Sigma}_{11}$ y

$$\begin{aligned}\mathbf{\Sigma}^* &= \mathbf{C}_{22}\mathbf{C}_{22}' = \mathbf{C}_{21}\mathbf{C}_{21}' + \mathbf{C}_{22}\mathbf{C}_{22}' - \mathbf{C}_{21}\mathbf{C}_{21}' \\ &= (\mathbf{C}_{21}\mathbf{C}_{21}' + \mathbf{C}_{22}\mathbf{C}_{22}') - (\mathbf{C}_{21}\mathbf{C}_{11}')((\mathbf{C}_{11}^{-1})'\mathbf{C}_{11}^{-1})(\mathbf{C}_{11}\mathbf{C}_{21}') = \mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}.\end{aligned}$$

lo que concluye la demostración. □

Del Corolario 4.3 se deduce inmediatamente que bajo las condiciones del mismo, si existen $E(\mathbf{X}_2|\mathbf{X}_1) = \boldsymbol{\mu}_2 + \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1)$ y $Var(\mathbf{X}_2|\mathbf{X}_1) = g(\mathbf{X}_1)\left(\mathbf{\Sigma}_{22} - \mathbf{\Sigma}_{21}\mathbf{\Sigma}_{11}^{-1}\mathbf{\Sigma}_{12}\right)$ para alguna función g. Más aún se puede probar, que si la matriz de covarianza condicional no depende de $\mathbf{X}_2$ entonces, $\mathbf{X}$ debe ser normal multivariada. Esta propiedad caracteriza a la distribución normal dentro de la clase de las elípticas (ver, Kelker, 1970).

4.3. Caso Funcional

Vamos a extender la definición de distribución elíptica para el caso en que se trabaje con un espacio de Hilbert $\mathcal{H}$ separable. La definición se basará en la dada para el caso multivariado.

Definición 4.3. Sea V un elemento aleatorio en un espacio de Hilbert separable $\mathcal{H}$. Diremos que V tiene una distribución elíptica de parámetros $\mu \in \mathcal{H}$ y $\mathbf{\Gamma} : \mathcal{H} \to \mathcal{H}$, con $\mathbf{\Gamma}$ autoadjunto, semidefinido positivo y compacto y lo indicaremos $\mathbf{V} \sim \mathcal{E}(\mu, \mathbf{\Gamma})$ si para todo $A : \mathcal{H} \to I\!R^p$ lineal y acotado (o sea, tal que $\sup_{\|x\|=1} \|Ax\| < \infty$) se cumple que AV tiene distribución elíptica multivariada de parámetros $A\mu$ y $\mathbf{\Sigma} = A\mathbf{\Gamma}A^*$.

Probaremos primero que los dos parámetros, μ y $\mathbf{\Gamma}$, que caracterizan al elemento V son la esperanza y el operador de covarianza, en caso de existir.

Lema 4.2. *Sea V un elemento aleatorio en un espacio de Hilbert separable $\mathcal{H}$ con distribución $\mathcal{E}(\mu, \mathbf{\Gamma})$. Si existe $E(V)$ entonces, $E(V) = \mu$.*

Demostración. Sea $f \in \mathcal{H}^*$ por hipótesis se tiene que $E(|f(V)|) < \infty$ y como $f : \mathcal{H} \to I\!R$ es lineal y continuo, es lineal y acotado. Por lo tanto, $f(V)$ es elíptico de parámetros $f(\mu)$ y $f\mathbf{\Gamma}f^*$. Como existe $E(V)$, existe $E(f(V))$ y además $E(f(V)) = f(E(V))$. Pero $E(f(V)) = f(\mu)$ luego, por la unicidad, se tiene que $E(V) = \mu$. □

Lema 4.3. *Sea V un elemento aleatorio en un espacio de Hilbert separable $\mathcal{H}$ con distribución $\mathcal{E}(\mu, \mathbf{\Gamma})$. Si existe el operador de covarianza se cumple $\mathbf{\Gamma}_V = \alpha\, \mathbf{\Gamma}$, para algún $\alpha \in I\!R$.*

Para demostrar que la covarianza es efectivamente $\mathbf{\Gamma}$ debemos primero probar el siguiente resultado auxiliar.

Lema 4.4. *Sea $V \sim \varepsilon(\mu, \mathbf{\Gamma})$ un elemento aleatorio en $\mathcal{H}_1$ elíptico de parámetros μ y $\mathbf{\Gamma}$ y $A : \mathcal{H}_1 \to \mathcal{H}_2$ lineal y acotada. Luego AV es un elemento aleatorio elíptico en $\mathcal{H}_2$ de parámetros $A\mu$ y $A\mathbf{\Gamma}A^*$.*

Demostración. Sea $B : \mathcal{H}_2 \to I\!R^d$ lineal y acotado, veamos que BAV es elíptico multivariado de media $BA\mu$ y matriz de covarianza $BA\mathbf{\Gamma}A^*B^*$.

Sea $B \circ A : \mathcal{H}_1 \to I\!R^d$ el operador composición. Luego, $B \circ A$ es lineal y acotado y por lo tanto, $BAV = (B \circ A)(V)$ es elíptico de parámetros $B \circ A(\mu) = BA\mu$ y $(B \circ A)\mathbf{\Gamma}(B \circ A)^* = BA\mathbf{\Gamma}A^*B^*$, teniendo así lo deseado. □

La demostración del Lema 4.3 se deducirá de la Proposición 2.9 y del Lema 4.4 por la unicidad de la covarianza.

Para ello, será conveniente tener definidas una serie de operadores especiales. Como $\mathcal{H}$ es separable, admite una base ortonormal numerable o finita, o sea que existen $\{\phi_n\}_{n \in I\!N}$ (eventualmente finitos si el espacio es de dimensión finita) ortonormales y que generan $\mathcal{H}$.

Podemos definir entonces $P_n = P_{<\phi_1, \ldots, \phi_n>}$, la proyección ortogonal sobre el subespacio $\mathcal{H}_1$ generado por $\phi_1, \ldots, \phi_n$. Como en realidad, nos interesará llegar finalmente a $I\!R^n$, vamos a efectuar una composición con el operador natural que identifica este subespacio con $I\!R^n$. Es decir, consideramos el operador $T_n : \mathcal{H} \to I\!R^n$ definido como

$$T_n(\phi_i) = \begin{cases} \mathbf{e}_i & i = 1, 2, \ldots, n \\ 0 & i > n \end{cases} , \tag{6}$$

con $\mathbf{e}_1, \ldots, \mathbf{e}_n$ los vectores de la base canónica de $I\!R^n$. En definitiva, dado $x \in \mathcal{H}$, $T_n(x) = \sum_{j=1}^n < x, \phi_j > \mathbf{e}_j$. Usaremos T_n en lugar de P_n como proyector, pues directamente llega a $I\!R^n$ y a cada uno de ellos los llamaremos "proyectores truncantes".

Ahora sí, probemos el Lema 4.3.

Demostración. del Lema 4.3. Queremos ver que $\mathbf{\Gamma}_V = \alpha\mathbf{\Gamma}$, o sea que dados $u, v \in \mathcal{H}$ se cumple $< \alpha\mathbf{\Gamma}u, v >_{\mathcal{H}} = a_V(u, v) = Cov(< v, V >_{\mathcal{H}}, < v, V >_{\mathcal{H}})$, donde especificamos en que espacio se toma el producto interno por claridad.

Sea $\{\phi_n\}_{n \in I\!N}$ la base de $\mathcal{H}$ formada por las autofunciones de $\mathbf{\Gamma}$ asociada a los autovalores $\lambda_1 \geq \lambda_2 \geq \ldots$ y $\{T_n\}_{n \in I\!N}$ la sucesión de proyecciones truncantes asociadas.

Sin pérdida de generalidad podemos suponer $\lambda_1 > 0$ ya que en caso contrario, $P(V = E(V)) = 1$ y el resultado es trivial.

Sea $d \in I\!N$ fijo. Por ser V elíptico, T_dV es elíptico multivariado con parámetros $T_d\mu$ y $T_d\mathbf{\Gamma}T_d^*$. Como V tiene segundo momento finito, entonces T_dV tiene segundo momento finito y $E(T_dV) = T_d\mu$ y la covarianza, $\mathbf{\Sigma}_d$, de T_dV es proporcional a $T_d\mathbf{\Gamma}T_d^*$. Por lo tanto, existe $\alpha_d \in I\!R$ tal que $\mathbf{\Sigma}_d = \alpha_d T_d \mathbf{\Gamma} T_d^*$.

Veamos que α_d no depende de d. Es fácil ver que $T_d\mathbf{\Gamma}u = \sum_{i=1}^d \lambda_i < \phi_i, u >_{\mathcal{H}} \phi_i$. Por lo tanto, usando que $< \phi_i, T_d^*\mathbf{x} >_{\mathcal{H}} = < T_d\phi_i, \mathbf{x} >_{I\!R^d} = < \mathbf{e}_i, \mathbf{x} >_{I\!R^d} = x_i$ para todo $\mathbf{x} \in I\!R^d$, se deduce que

$$T_d\mathbf{\Gamma}T_d^* = \operatorname{diag}(\lambda_1, \ldots, \lambda_d) \ . \tag{7}$$

Sea $k \leq d$ y $\pi_k : \mathbb{R}^d \to \mathbb{R}^k$ la proyección usual $\pi(\mathbf{x}) = \pi_k((x_1, \ldots, x_p)') = (x_1, \ldots, x_k)' = \mathbf{A}_k \mathbf{x}$. Luego $\pi_k T_d V = T_k V$ y por lo tanto, $Cov(T_k V) = \mathbf{\Sigma}_k = \mathbf{A}_k \mathbf{\Sigma}_d \mathbf{A}_k'$ de donde, $\alpha_k T_k \mathbf{\Gamma} T_k^* = \alpha_d \mathbf{A}_k (T_d \mathbf{\Gamma} T_d^*) \mathbf{A}_k'$, con $\mathbf{A}_k = \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}$. Usando (7) obtenemos

$$\alpha_k \mathrm{diag}\,(\lambda_1, \ldots, \lambda_k) = \alpha_d \mathbf{A}_k \mathrm{diag}\,(\lambda_1, \ldots, \lambda_d)\, \mathbf{A}_k' = \alpha_d \mathrm{diag}\,(\lambda_1, \ldots, \lambda_k)$$

de donde $\alpha_k = \alpha_d$. Luego existe $\alpha \in \mathbb{R}$ tal que para todo $d \in \mathbb{N}$ la covarianza, $\mathbf{\Sigma}_d$, de $T_d V$ es igual a $\alpha T_d \mathbf{\Gamma} T_d^*$, de donde

$$< \alpha T_d \mathbf{\Gamma} T_d^* \mathbf{x}, \mathbf{y} >_{\mathbb{R}^d} = Cov(< \mathbf{x}, T_d V >_{\mathbb{R}^d}, < \mathbf{y}, T_d V >_{\mathbb{R}^d}).$$

Utilizando la definición de la adjunta de T_d, tenemos que $< T_d \mathbf{\Gamma} T_d^* \mathbf{x}, \mathbf{y} >_{\mathbb{R}^d} = < \mathbf{\Gamma} T_d^* \mathbf{x}, T_d^* \mathbf{y} >_{\mathcal{H}}$ mientras que el miembro derecho de la igualdad se puede escribir como

$$Cov(< \mathbf{x}, T_d V >_{\mathbb{R}^d}, < \mathbf{y}, T_d V >_{\mathbb{R}^d}) = Cov(< T_d^* \mathbf{x}, V >_{\mathcal{H}}, < T_d^* \mathbf{y}, V >_{\mathcal{H}})\ .$$

Luego, tenemos que para todo $d \in \mathbb{N}$, $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^d$,

$$< \alpha \mathbf{\Gamma} T_d^* \mathbf{x}, T_d^* \mathbf{y} >_{\mathcal{H}} = Cov(< T_d^* \mathbf{x}, V >_{\mathcal{H}}, < T_d^* \mathbf{y}, V >_{\mathcal{H}}) = a_V(T_d^* \mathbf{x}, T_d^* \mathbf{y}).$$

Sean $u, v \in \mathcal{H}$, sean $u_d = P_d u$, $v_d = P_d v$. Sabemos además que $\lim_{d \to \infty} \|u - u_d\| = 0$ y $\lim_{d \to \infty} \|v - v_d\| = 0$. Sean $\mathbf{x} = T_d u = T_d u_d$ e $\mathbf{y} = T_d v = T_d v_d$. Luego, es inmediato que $u_d = T_d^* \mathbf{x}$, $v_d = T_d^* \mathbf{y}$, de donde Entonces,

$$< \alpha \mathbf{\Gamma} u_d, v_d >_{\mathcal{H}} = < \alpha \mathbf{\Gamma} T_d^* \mathbf{x}, T_d^* \mathbf{y} >_{\mathcal{H}} = a_V(T_d^* \mathbf{x}, T_d^* \mathbf{y}) = a_V(u_d, v_d)$$

Por ser a_V continuo se tiene que $\lim_{d \to \infty} a_V(u_d, v_d) = a_V(u, v)$. Por ser $\mathbf{\Gamma}$ un operador autoadjunto y compacto $\lim_{d \to \infty} < \mathbf{\Gamma} u_d, v_d >_{\mathcal{H}} = < \mathbf{\Gamma} u, v >_{\mathcal{H}}$, lo que concluye la demostración. $\square$

En base a los resultados vistos para el caso finito dimensional, tenemos que una manera de obtener elementos aleatorios elípticos es a través de la siguiente transformación. Sea $\mathbf{V}_1$ un elemento gaussiano en $\mathcal{H}$ de media cero y operador de covarianza $\mathbf{\Gamma}_{V_1}$ y sea Z una variable aleatoria con distribución G independiente de $\mathbf{V}_1$. Sea $\boldsymbol{\mu} \in \mathcal{H}$, luego $\mathbf{V} = \boldsymbol{\mu} + Z\, \mathbf{V}_1$ tiene distribuición elíptica y si existe $E(Z^2)$ entonces $\mathbf{\Gamma}_V = E(Z^2)\, \mathbf{\Gamma}_{V_1}$.

4.4. Distribución condicional

Vamos a demostrar algunas propiedades sobre la distribución condicional en familias elípticas. En nuestro caso, tendremos V un elemento aleatorio perteneciente a una familia elíptica de parámetros μ y $\mathbf{\Gamma}$. Consideremos en $\mathcal{H}$ la base ortonormal, $\{\phi_n\}_{n \in \mathcal{I}}$ ($\mathcal{I}$ finito o numerable) formada por las autofunciones $\mathbf{\Gamma}$ asociadas a los autovalores $\lambda_1 \geq \lambda_2 \geq \ldots$. Dado $d \in \mathcal{I}$ fijo, definamos

$$\mathcal{H}_1 = < \phi_1, \ldots, \phi_d > \qquad \mathcal{H}_2 = < \phi_1, \ldots, \phi_d >^{\perp}$$

Ambos son subespacios cerrados, el primero por ser de dimensión finita y el segundo por ser el ortogonal de un conjunto. Por lo tanto, ambos serán a su vez espacios de Hilbert. Definimos sobre estos espacios proyecciones truncantes, es decir, $P_{\mathcal{H}_1} : \mathcal{H} \to \mathcal{H}_1$ y $P_{\mathcal{H}_2} : \mathcal{H} \to \mathcal{H}_2$ definidos como

$$P_{\mathcal{H}_1}(\phi_i) = \begin{cases} \phi_i & i = 1, 2, \ldots, d \\ 0 & i > d \end{cases} \qquad P_{\mathcal{H}_2}(\phi_i) = \begin{cases} \phi_i & i > d \\ 0 & i = 1, 2, \ldots, d \end{cases} . \tag{8}$$

Por lo tanto, $P_{\mathcal{H}_1} = P_d$. En base a estas proyecciones podemos construir $V_1 = P_{\mathcal{H}_1} V \in \mathcal{H}_1$, $W_1 = T_d V \in I\!R^d$ y $V_2 = P_{\mathcal{H}_2} V \in \mathcal{H}_2$ elementos aleatorios, ambos elípticos por el Lema 4.4.

Lo que hemos hecho es "partir" al elemento aleatorio en dos partes, una de ellas de dimensión finita. Esto nos permitirá definir entonces una distribución condicional $V_2|V_1$ siguiendo los lineamientos descriptos anteriormente.

Teorema 4.8. *Sea $\mathcal{H}$ un espacio de Hilbert separable. Sea V un elemento aleatorio en $\mathcal{H}$ con distribución $\mathcal{E}(\mu, \mathbf{\Gamma})$, segundo momento finito $\mathbf{\Gamma}$ Hilbert Schmidt (ver 6.13) para que $\sum_{i=1}^{\infty} \lambda_i < \infty$. Sea $d \in \mathcal{I}$ fijo y consideremos $V_1 = P_{\mathcal{H}_1} V$, $W_1 = T_d V$ y $V_2 = P_{\mathcal{H}_2} V$ con $P_{\mathcal{H}_1}$ y $P_{\mathcal{H}_2}$ definidos en (8) y T_d definido en (6). Sean $\lambda_1 \geq \lambda_2 \geq \ldots$ los autovalores de $\mathbf{\Gamma}$ y supongamos que $\lambda_d > 0$. Entonces,*

a) la matriz de covarianza de W_1, $\mathbf{\Sigma}_{W_1} = T_d \mathbf{\Gamma} T_d^* = diag(\lambda_1, \ldots, \lambda_d)$ es no singular y

b) $E(V_2|W_1) = \mu_2 + \mathbf{\Gamma}_{V_2,W_1} \mathbf{\Sigma}_{W_1}^{-1} (W_1 - \boldsymbol{\mu}_1)$,

donde $\mathbf{\Gamma}_{V_2,W_1}$ es el operador de covarianza entre V_2 y W_1, $\boldsymbol{\mu}_1 = E(W_1)$ y $\mu_2 = E(V_2)$.

Demostración. a) Es inmediato pues $\lambda_d > 0$.

b) Sin pérdida de generalidad podemos suponer que $\mathbf{\Gamma}$ es el operador de covarianza. Debemos probar que la esperanza de $V_2|W_1$ es efectivamente lo que se quiere. Vamos a verificar las condiciones de la esperanza. Sea $f \in \mathcal{H}^*$, debemos ver que $E(|f(V_2|W_1)|) < \infty$ y que $E(f(V_2)|W_1) = f(\mu_2 + \mathbf{\Gamma}_{V_2,W_1} \mathbf{\Sigma}_{W_1}^{-1} (W_1 - \mu_1))$. Ahora bien, sabemos que entonces $f(V_2)$ será elíptico real, más aún, $\mathbf{W} = (W_1, f(V_2)) = TV$, con T lineal y acotada, y por lo tanto, $\mathbf{W}$ es elíptico de parámetros $(\boldsymbol{\mu}_1, f(\mu_2))$ y $T\mathbf{\Gamma}T^* = \begin{pmatrix} \mathbf{\Sigma}_{W_1} & Cov(W_1, f(V_2)) \\ Cov(W_1, f(V_2))' & f\mathbf{\Gamma}f^* \end{pmatrix}$. Por el caso multivariado, del Corolario 4.3 obtenemos $f(V_2)|W_1$ tiene una distribución elíptica con esperanza $f(\mu_2) + Cov(f(V_2), W_1) \mathbf{\Sigma}_{W_1}^{-1} (W_1 - \boldsymbol{\mu}_1)$. Usando la Proposición 2.9 deducimos que $Cov(f(V_2), W_1) = f\mathbf{\Gamma}_{V_2,W_1}$. Por lo tanto,

$$\begin{aligned} E(f(V_2)|W_1) &= f(\mu_2) + Cov(f(V_2), W_1) \mathbf{\Sigma}_{W_1}^{-1} (W_1 - \boldsymbol{\mu}_1) \\ &= f(\mu_2) + f\mathbf{\Gamma}_{V_2,W_1} \mathbf{\Sigma}_{W_1}^{-1} (W_1 - \boldsymbol{\mu}_1) = f(\mu_2 + \mathbf{\Gamma}_{V_2,W_1} \mathbf{\Sigma}_{W_1}^{-1} (W_1 - \boldsymbol{\mu}_1)) \end{aligned}$$

y se obtiene el resultado deseado. □

5. Puntos Autoconsistentes y Puntos Principales

5.1. Definiciones y propiedades

Definición 5.1. Dado un conjunto $\mathcal{W} = \{y_1, \ldots, y_k\}$ con $y_i \in \mathcal{H}$, $1 \leq i \leq k$ definimos la distancia mínima de un elemento aleatorio V al conjunto $\mathcal{W}$ como $d(V|y_1, \ldots, y_k) = \text{mín}_{1\leq j\leq k} \|V - y_j\|$.

A partir de un conjunto $\mathcal{W}$ y de esta distancia, induciremos una partición del espacio explicitada por los dominios de atracción.

Definición 5.2. Para un conjunto $\mathcal{W} = \{y_1, \ldots, y_k\}$, el dominio de atracción $\mathcal{D}_j$ de y_j consiste en todos los elementos de $\mathcal{H}$ que tienen a y_j como punto más cercano de $\mathcal{W}$, es decir, $\mathcal{D}_j = \{x \in \mathcal{H} : \|x - y_j\| < \|x - y_\ell\| \quad \ell \neq j\}$.
En caso de empates, se asigna el punto x al conjunto con menor índice j.

Definición 5.3. Sea V un elemento aleatorio tal que existe $E(V)$. Un conjunto $\mathcal{W} = \{y_1, \ldots, y_k\}$ se dice autoconsistente para V si $E(V|V \in \mathcal{D}_j) = y_j$.

Podemos extender esta definición para elementos aleatorios.

Definición 5.4. Sea V un elemento aleatorio tal que existe $E(V)$. Un elemento aleatorio W se dice autoconsistente para V si $E(V|W) = W$.

Lema 5.1. *Sea V un elemento aleatorio tal que existe $E(V)$. Si V es un elemento aleatorio con un conjunto autoconsistente de puntos $\{y_1, \ldots, y_k\}$ entonces $E(V)$ es una combinación convexa de $y_1, \ldots, y_k$.*

Demostración. Sea $\mathcal{D}_j$ el dominio de atracción de y_j, es claro que $\bigcup_{j=1}^k \mathcal{D}_j$ es una partición de $\mathcal{H}$, con lo cual, $E(V) = \sum_{j=1}^k E(V \mathbb{1}_{\mathcal{D}_j}(V) = \sum_{j=1}^k \pi_j E(V|V \in \mathcal{D}_j) = \sum_{i=1}^k \pi_j y_j$ con $\mathbb{1}_A(V) = 1$ si $V \in A$ y $\mathbb{1}_A(V) = 0$ si $V \notin A$, $\pi_j = P(V \in \mathcal{D}_j)$. Como los $\mathcal{D}_j$ forman una partición de $\mathcal{H}$, $\sum_{j=1}^k \pi_j y_j$ es una combinación convexa de $y_1, \ldots, y_k$. $\square$

Por el Lema 5.1, si $k = 1$ y V es un elemento aleatorio con un conjunto autoconsistente $\{y_1\}$ entonces $y_1 = E(V)$

Definición 5.5. Los elementos $\xi_1, \ldots, \xi_k$ son llamados puntos principales si

$$P_V(k) = E(d^2(V|\xi_1, \ldots, \xi_k)) = \min_{y_j \in H} E(d^2(V|y_1, \ldots, y_k))$$

Lema 5.2. *Un conjunto $\mathcal{W} = \{\xi_1, \ldots, \xi_k\}$ de puntos principales del elemento aleatorio V es autoconsistente.*

Demostración. Sean $\xi_1, \ldots, \xi_k$ los puntos principales de V, con $\mathcal{D}_1, \ldots, \mathcal{D}_j$ los respectivos dominios de atracción. Sea $\pi_j = P(V \in \mathcal{D}_j)$. Entonces, por la definición de los dominios de atracción tenemos

$$\begin{aligned} E(d^2(V|\xi_1, \ldots, \xi_k)) &= E\left(\min_{1\le j\le k} \|V - \xi_j\|^2\right) = \sum_{i=1}^{k} \pi_i E(\min_{1\le j\le k} \|V - \xi_j\|^2 \; |V \in \mathcal{D}_i) \\ &= \sum_{i=1}^{k} \pi_i E(\|V - \xi_i\|^2 \; |V \in \mathcal{D}_i) \; . \end{aligned} \tag{9}$$

Por otra parte, por ser $\xi_1, \ldots, \xi_k$ puntos principales de V

$$\begin{aligned} E(d^2(V|\xi_1, \ldots, \xi_k)) &= \min_{y_1\ldots y_k \in H} E(d^2(V|y_1, \ldots, y_k)) = \min_{y_1\ldots y_k \in H} E\left(\min_{1\le j\le k} \|V - y_j\|^2\right) \\ &= \min_{y_1\ldots y_k \in H} \sum_{i=1}^{k} \pi_i E(\min_{1\le j\le k} \|V - y_j\|^2 \; |V \in \mathcal{D}_i) \\ &\le \min_{y_1\ldots y_k \in H} \sum_{i=1}^{k} \pi_i E(\|V - y_i\|^2 \; |V \in \mathcal{D}_i) \; . \end{aligned} \tag{10}$$

De (9), (10) y del hecho que $\operatorname{argmin}_{y_i \in H} E(\|V - y_i\|^2 \; |V \in \mathcal{D}_i) = E(V|V \in \mathcal{D}_i)) = \xi_{0,i}$, obtenemos

$$\begin{aligned} E(d^2(V|\xi_1, \ldots, \xi_k)) &= \sum_{i=1}^{k} \pi_i E(\|V - \xi_i\|^2 \; |V \in \mathcal{D}_i) \\ &= \min_{y_1\ldots y_k \in H} \sum_{i=1}^{k} \pi_i E(\|V - y_i\|^2 \; |V \in \mathcal{D}_i) \\ &= \sum_{i=1}^{k} \pi_i \min_{y_i \in H} E(\|V - y_i\|^2 \; |V \in \mathcal{D}_i) = \sum_{i=1}^{k} \pi_i E(\|V - \xi_{0,i}\|^2 \; |V \in \mathcal{D}_i) \\ &= E(d^2(V|\xi_{0,1}, \ldots, \xi_{0,k})) \; . \end{aligned}$$

Luego, $\xi_{0,i} = E(V|V \in \mathcal{D}_i))$ son puntos principales.

Sea $\xi_0 = \sum_{i=1}^{k} \xi_{0,i} I_{V\in\mathcal{D}_i}$ y $\xi = \sum_{i=1}^{k} \xi_i I_{V\in\mathcal{D}_i}$. Luego, $E(d^2(V|\xi_1, \ldots, \xi_k)) = E\|V - \xi\|^2$. De donde, $E\|V - \xi_0\|^2 = E\|V - \xi\|^2$. Con lo cual, usando que $\xi_{0,i} = E(V|V \in \mathcal{D}_i))$ y por lo tanto $E(< V - \xi_0, \xi - \xi_0 >) = 0$, obtenemos

$$E\|V - \xi\|^2 = E\|V - \xi_0\|^2 + E\|\xi - \xi_0\|^2 + 2E(< V - \xi_0, \xi - \xi_0 >) = E\|V - \xi\|^2 + E\|\xi - \xi_0\|^2$$

de donde $E\|\xi - \xi_0\|^2 = 0$. Por lo tanto, $P(\xi_0 = \xi) = 1$ y entonces los puntos principales son autoconsistentes. $\square$

Luego, puntos autoconsistentes existen cada vez que existan puntos principales.

Lema 5.3. *Sea V un elemento aleatorio de $\mathcal{H}$ y sea $V_2 = \nu + \rho\, UV$ con $\nu \in \mathcal{H}$, ρ un escalar y $U : \mathcal{H} \to \mathcal{H}$ un operador unitario (suryectivo e isométrico). Entonces, se verifica*

a) Si $\mathcal{W} = \{y_1, \ldots, y_k\}$ es un conjunto de k puntos autoconsistentes de V entonces $\mathcal{W}_2 = \{\nu + \rho\, Uy_1, \ldots, \nu + \rho\, Uy_k\}$ lo será de V_2.

b) Si $\mathcal{W} = \{y_1, \ldots, y_k\}$ es un conjunto de k puntos principales de V entonces $\mathcal{W}_2 = \{\nu + \rho\, Uy_1, \ldots, \nu + \rho\, Uy_k\}$ lo serán de V_2 y además $E(d^2(V_2|\mathcal{W}_2)) = \rho^2 E(d^2(V|\mathcal{W}))$.

Demostración. a) Sabemos que $\mathcal{W}$ es autoconsistente para V, entonces tendremos que $E(V|V \in \mathcal{D}_j) = y_j$. Notemos que por ser U unitaria, $V \in \mathcal{D}_j$ si y solamente si $V_2 = \nu + \rho\, UV \in \widetilde{\mathcal{D}_j}$ donde $\widetilde{\mathcal{D}_j}$ es el dominio de atracción de $\nu + \rho\, Uy_j$. En particular, $\nu + \rho\, U\mathcal{D}_j \subset \widetilde{\mathcal{D}_j}$. Entonces,

$$\begin{aligned} E(V_2|V_2 \in \widetilde{\mathcal{D}_j}) &= E(\nu + \rho UV|V_2 \in \widetilde{D_j}) = \nu + \rho U E(V|\nu + \rho UV \in \nu + \rho U\mathcal{D}_j) \\ &= \nu + \rho U E(V|V \in \mathcal{D}_j) = \nu + \rho U y_j. \end{aligned}$$

b) Sean $\xi_1, \ldots, \xi_k$ un conjunto de puntos cualesquiera de $\mathcal{H}$ y sean $\mathcal{A}_1, \ldots, \mathcal{A}_k$ sus respectivos dominios de atracción. Debemos ver que $E(d^2(V_2|\mathcal{W}_2)) \le E(d^2(V_2|\xi_1, \ldots, \xi_k))$. Sea z_j tal que $\xi_j = \nu + \rho\, Uz_j$. Tenemos

$$\begin{aligned} E(d^2(V_2|\xi_1, \ldots, \xi_k)) &= E(\min_{1\le j\le k} \|V_2 - \xi_j\|^2) = E(\min_{1\le j\le k} \|\nu + \rho\, UV - \xi_j\|^2) \\ &= E(\min_{1\le j\le k} \|\nu + \rho\, UV - \nu - \rho\, Uz_j\|^2) = E(\min_{1\le j\le k} \|\rho\, UV - \rho\, Uz_j\|^2) \\ &= \rho^2 E(\min_{1\le j\le k} \|UV - Uz_j\|^2) = \rho^2 E(\min_{1\le j\le k} \|V - z_j\|^2) \end{aligned}$$

donde la última igualdad se deduce del hecho de ser U una isometría. Pero sabemos que $\{y_1, \ldots, y_k\} = \operatorname{argmin}_{z_1,\ldots,z_k} E(\min_{1\le j\le k} \|V - z_j\|^2)$ por ser los puntos principales de V. Luego,

$$E(d^2(V_2|\xi_1, \ldots, \xi_k)) = \rho^2 E(\min_{1\le j\le k} \|V - z_j\|^2) \ge \rho^2 E(\min_{1\le j\le k} \|V - y_j\|^2) = E(\min_{1\le j\le k} \|V_2 - \xi_{0,j}\|^2)$$

donde $\xi_{0,j} = \nu + \rho U y_j$, o sea, $\mathcal{W}_2$ son puntos principales de V_2. Por otra parte, el cálculo hecho permite ver que $E(d^2(V_2|W_2)) = \rho^2 E(d^2(V|W))$. □

Este resultado nos permitirá suponer en todo momento que el elemento aleatorio tiene esperanza 0.

Lema 5.4. *Sea V un elemento aleatorio con esperanza 0. Sean $y_1, \ldots, y_k$ un conjunto de k puntos autoconsistentes de V, de manera tal que estos puntos generan un subespacio $\mathcal{M}$ de dimensión q, con base ortonormal $\{e_1, \ldots, e_q\}$. Luego, el vector aleatorio de $\mathbb{R}^q$ definido por $\mathbf{X} = (X_1, \ldots, X_q)'$ con $X_i =< e_i, V >$ tendrá a $\mathcal{W} = \{\mathbf{w}_j\}_{1\le j\le k}$ con $\mathbf{w}_j = (w_{1j}, \ldots, w_{1q})$ y $w_{ij} =< e_i, y_j >$ como conjunto autoconsistente.*

Demostración. Para facilitar la notación, definimos $A : \mathcal{H} \to I\!R^q$ como $A(v) = \mathbf{v} = (v_1, \ldots, v_q)$ con $v_i =< e_i, v >$. Queremos probar que $\mathbf{X} = AV \in I\!R^q$ tiene como conjunto autoconsistente a $\mathcal{W} = \{\mathbf{w}_1, \ldots, \mathbf{w}_k\} = \{Ay_1, \ldots, Ay_k\}$. En efecto, si $\widetilde{\mathcal{D}_j}$ es el dominio de atracción de Ay_j y $\mathcal{D}_j$ el de y_j, tendremos que $V \in \mathcal{D}_j$ si y sólo si $AV \in \widetilde{\mathcal{D}_j}$, con lo cual

$$E(AV|AV \in \widetilde{\mathcal{D}_j}) = AE(V|AV \in \widetilde{\mathcal{D}_j}) = AE(V|V \in \mathcal{D}_j) = Ay_j = \mathbf{w}_j.$$

□

Definición 5.6. Sea $W : \Omega \to \mathcal{H}$ un elemento aleatorio discreto, distribuído conjuntamente con el elemento aleatorio $V : \Omega \to \mathcal{H}$. Indiquemos por $\mathcal{S}(W)$ el soporte de W.

El elemento aleatorio W es una mejor aproximación de k puntos de V si $\mathcal{S}(W)$ contiene exactamente a k puntos distintos $y_1, \ldots, y_k$ y vale que $E(\|V - W\|^2) \leq E(\|V - Z\|^2)$ para cualquier $Z : \Omega \to \mathcal{H}$ discreto con $\#\mathcal{S}(Z) \leq k$, o sea, con a lo sumo k puntos en su soporte.

Lema 5.5. *Sea W una mejor aproximación de k puntos de V. Luego, vale que*

a) Si $V \in \mathcal{D}_j$ luego W debe valer y_j con probabilidad 1. O sea que

$$W = \sum_{i=1}^{k} y_i \mathbb{1}_{\mathcal{D}_j}(V).$$

b) $\|V - W\| \leq \|V - y_j\|$ c.t.p. para todo $y_j \in \mathcal{S}(W)$.

c) $E(V|W) = W$ c.t.p. (o sea que W es autoconsistente para V)

Demostración. a) Supondremos siempre que la probabilidad de que V se encuentre en la frontera de dos dominios de atracción ($\mathcal{D}_i \cap \mathcal{D}_j$) es 0.

Supongamos que el resultado no fuese cierto, eso es, que el conjunto $A = \{V \in \mathcal{D}_j\} \cap \{W = y_r\}$ tiene probabilidad estrictamente positiva, con $r \neq j$. Vamos a definir un nuevo elemento $Z(\omega)$ que vale $W(\omega)$ si ω no está en A y que vale y_i si lo está. Veamos entonces que Z aproxima mejor a V que W.

$$\begin{aligned} E(||V - Z||^2) &= E(||V - Z||^2 \mathbb{1}_{A^c}) + E(||V - Z||^2 \mathbb{1}_A) \\ &= E(||V - W||^2 \mathbb{1}_{A^c}) + E(||V - Z||^2 \mathbb{1}_A) \\ &= E(||V - W||^2 \mathbb{1}_{A^c}) + E(||V - y_j||^2 \mathbb{1}_A) \end{aligned}$$

Ahora bien, estando en el conjunto A tenemos que

$$||V - y_j||^2 \mathbb{1}_A < ||V - y_r||^2 \mathbb{1}_A = ||V - W||^2 \mathbb{1}_A$$

y como A tiene probabilidad positiva luego obtenemos que

$$E(||V - y_j||^2 \mathbb{1}_A) < E(||V - W||^2 \mathbb{1}_A).$$

Siguiendo con la cuenta de antes esto nos dice que

$$\begin{aligned} E(||V - W||^2 \mathbb{1}_{A^c}) + E(||V - y_j||^2 \mathbb{1}_A) < \\ E(||V - W||^2 \mathbb{1}_{A^c}) + E(||V - W||^2 \mathbb{1}_A) = \\ E(||V - W||^2). \end{aligned}$$

O sea que $E(||V - Z||^2) < E(||V - W||^2)$ con lo que Z sería una mejor aproximación que W con soporte de cardinal k, absurdo pues W era la mejor aproximación. Se tiene así lo deseado.

b) Si $V(\omega) \in \mathcal{D}_j$ luego por lo de antes eso implicará, salvo medida nula, que $W(\omega) = y_j$. Entonces, $||V - W||(\omega) = ||V(\omega) - y_j||$ y esto es menor o igual que $||V(\omega) - y_i||$ para cualquier y_i pues precisamente en dicho ω tenemos que el $V(\omega) \in \mathcal{D}_j$ y en esa región el punto y_j es el punto más cercano a $V(\omega)$.

c) Pensemos en que $E(V|W)$ es por definición una función $g(W)$ medible que minimiza la distancia esperada al cuadrado entre V y cualquier función medible $h(W)$. Ahora bien, una función $h(W)$ tendrá necesariamente un soporte de k o menos puntos pues W tiene un soporte de k puntos. Entonces, por definición de mejor aproximación, tendremos que W aproxima mejor a $h(W)$ para toda función medible h, con el criterio de la distancia esperada al cuadrado. Entonces, la función $g(W) = E(V|W)$ coincide con W.

□

Por último, notemos que dado un conjunto autoconsistente $\{y_1, \dots, y_k\}$ de V podemos definir de manera natural una variable aleatoria Y con soporte $\{y_1, \dots, y_k\}$ tal que $P(Y = y_j) = P(V \in \mathcal{D}_j)$, es decir, $Y = \sum_{i=1}^{k} y_i I_{V \in \mathcal{D}_i}$. Por ser un conjunto autoconsistente, se cumplirá que $E(V|Y) = Y$, con probabilidad 1. En principio no tendría por qué ser una mejor aproximación, eso ocurrirá en caso de que el conjunto sea de puntos principales.

5.2. Relación de los puntos autoconsistentes con componentes principales

Como hemos visto si V es un elemento aleatorio con un conjunto autoconsistente de $k = 1$ puntos $\{y_1\}$ entonces $E(V) = y_1$, por lo que si suponemos $E(V) = 0$ tendremos $y_1 = 0$. El siguiente resultado trata de caracterizar el subespacio generado por los puntos autoconsistentes si $k > 1$ generaliza un resultado análogo dado por Tarpey y Flury (1995) para el caso finito–dimensional y extendido al caso de procesos gausianos por Tarpey y Kinateder (2003), si bien la demostración de estos últimos autores no garantiza que la matriz $\mathbf{\Gamma}_{W_1,W_1}$ que se define en la demostración del Teorema sea inversible. Establecemos en primer lugar una serie de lemas auxiliares.

Lema 5.6. *Supongamos $E(V) = 0$. Luego, si $h \in Ker(\mathbf{\Gamma}_V)$ entonces $< h, V > \equiv 0 \in I\!R$ c.t.p.*

Demostración.

$$E(< h, V >) = < h, E(V) >= 0.$$

$$Var(< h, V >) = Cov(< h, V >, < h, V >) = a_V(h, h) = < \mathbf{\Gamma}_V h, h >= 0$$

pues $h \in Ker(\mathbf{\Gamma}_V)$. Entonces, como su varianza es cero luego es constante para casi todo punto. Como su esperanza es cero luego dicha constante es el 0. □

Lema 5.7. *$E(V) = 0$, $\mathcal{H}$ es separable. Sea*

$$\mathcal{H}_1 = Ker(\mathbf{\Gamma}_V),$$

$$\mathcal{H}_2 = Ker(\mathbf{\Gamma}_V)^{\perp}.$$

$V = P_{\mathcal{H}_1}V + P_{\mathcal{H}_2}V$. Luego $P_{\mathcal{H}_1}V \equiv 0$ c.t.p. ($P_{\mathcal{H}_2}V \equiv V$ c.t.p.).

Demostración. Supondremos ambos espacios de dimensión infinita. El caso en que alguno de ellos sea de dimensión finita se procede análogamente pero usando finitos índices.

$$\mathcal{H}_1 = Ker\mathbf{\Gamma}_V = < e_1, e_2, \ldots > .$$

Lo completamos a una base ortonormal de todo el espacio

$$\mathcal{H}_2 = Ker\mathbf{\Gamma}_V{}^{\perp} = < f_1, f_2, \ldots > .$$

Entonces $Var(P_{\mathcal{H}_1}V) = \sum_{i=1}^{\infty} < \mathbf{\Gamma}_V e_i, e_i >$ y acá cada sumando vale cero pues $e_i \in Ker(\mathbf{\Gamma}_V)$. Además, $E(P_{\mathcal{H}_1}V) = P_{\mathcal{H}_1}E(V) = 0$ con lo cual se deduce que $P_{\mathcal{H}_1}V$ es un elemento aleatorio de varianza constante. Como $Var(X) = E(||X - E(X)||^2)$ luego deducimos que $||V - E(V)||^2 = 0$ c.t.p. con lo cual $V = E(V) = 0$ c.t.p. y listo.

□

Corolario 5.1. $P(V = P_{\mathcal{H}_2}V) = 1.$

Demostración. Definición de c.t.p. □

Corolario 5.2. $P(V \in Ker(\mathbf{\Gamma}_V)^{\perp} = 1.$

Demostración.

$$\{V = P_{\mathcal{H}_2}V\} \subset \{V \in Ker(\mathbf{\Gamma}_V)\}$$

con lo cual

$$1 = P(V = P_{\mathcal{H}_2}V) \leq P(V \in Ker(\mathbf{\Gamma}_V))$$

y listo. □

Corolario 5.3. *Si $A \cap Ker(\mathbf{\Gamma}_V)^{\perp} = \varnothing$ entonces $P_V(A) = 0$.*

Demostración. Es claro pues $P(V \in Ker(\mathbf{\Gamma}_V)^{\perp} = 1$. $\square$

Teorema 5.1. *Sea $W = \{y_1, \ldots, y_k\}$ un conjunto de k puntos autoconsistentes. Luego todos ellos deberán estar necesariamente en $Ker(\mathbf{\Gamma}_V)^{\perp}$.*

Demostración. $Ker(\mathbf{\Gamma}_V)^{\perp}$ es un subespacio y por lo tanto un conjunto convexo. Los dominios de atracción de cada punto $\mathcal{D}_j$ también son conjuntos convexos. La intersección de ambos será un conjunto convexo. El soporte del elemento aleatorio V estará incluido en $Ker(\mathbf{\Gamma}_V)^{\perp}$ pues este tiene probabilidad 1.

$$y_j = E(V|V \in \mathcal{D}_j) = E(V|V \in \mathcal{D}_j, V \in Ker(\mathbf{\Gamma}_V)^{\perp}) = E(V|V \in \mathcal{D}_j \cap Ker(\mathbf{\Gamma}_V)^{\perp}).$$

Es una esperanza de un elemento aleatorio que se mueve en un conjunto convexo, luego esto caerá en el conjunto convexo (habría que probarlo ...) $\mathcal{D}_j \cap Ker(\mathbf{\Gamma}_V)^{\perp} \subset Ker(\mathbf{\Gamma}_V)^{\perp}$. Luego, $y_j \in Ker(\mathbf{\Gamma}_V)^{\perp}$ y listo.

$\square$

Corolario 5.4. *Sea $W = \{y_1, \ldots, y_k\}$ un conjunto de k puntos autoconsistentes. Sea $\mathcal{V}$ el subespacio generado por ellos. Entonces, $Ker(\mathbf{\Gamma}_V) \cap \mathcal{V} = \{0\}$.*

Demostración. Por el teorema anterior, $W = \{y_1, \ldots, y_k\} \subset Ker(\mathbf{\Gamma}_V)^{\perp}$. Por lo tanto, $\mathcal{V} \subset Ker(\mathbf{\Gamma}_V)^{\perp}$. Entonces, $Ker(\mathbf{\Gamma}_V) \cap \mathcal{V} = \{0\}$. $\square$

Vale también que $\mathbf{\Gamma}(\mathcal{V}) \cap \mathcal{V}^{\perp} = \{0\}$. Interesante destacar que esto ya sería una propiedad general de operadores semidefinidos y diagonalizables, no es necesario recurrir a ningún concepto probabilístico como en la otra hipótesis.

Demostración. Sea $z \in \mathbf{\Gamma}(\mathcal{V}) \cap \mathcal{V}^{\perp}$, entonces $z = \mathbf{\Gamma}v$ con $v \in \mathcal{V}$. Quiero ver que $z = 0$. Sea $\{\phi_j\}_{j \in \mathbb{N}}$ una base ortonormal de autofunciones de $\mathbf{\Gamma}$, asociados a los autovalores $\mu_1, \ldots, \mu_j, \ldots$ Entonces

$$v = \sum_{j=1}^{\infty} < v, \phi_j > \phi_j$$

luego

$$z = \mathbf{\Gamma}v = \sum_{j=1}^{\infty} < v, \phi_j > \mu_j \phi_j.$$

Ahora bien, como $v \in \mathcal{V}^{\perp}$ entonces $< z, w >= 0 \ \forall \ w \in \mathcal{V}^{\perp}$. En particular, podemos tomar $w = v$ y entonces tendremos que

$$0 =< z, v >=< \sum_{j=1}^{\infty} < v, \phi_j > \mu_j \phi_j, \sum_{j=1}^{\infty} < v, \phi_j > \phi_j >$$

$$= \sum_{j=1}^{\infty} \mu_j < v, \phi_j >^2 .$$

Los μ_j son no negativos, cada sumando deberá entonces dar 0. Esto está diciéndonos que $< v, \phi_j >=$ 0 siempre que μ_j sea distinto de 0. Obtendríamos así que $\mathbf{\Gamma} v = 0$, con lo cual $v \in Ker(\mathbf{\Gamma})$, luego z debe valer 0 y es lo que se quería probar. $\square$

Tenemos así el siguiente corolario final:

Corolario 5.5. *Sea V un elemento elíptico con $E(V) = 0$ y operador de covarianza compacto $\mathbf{\Gamma}$. Sea $\mathcal{V}$ el subespacio generado por un conjunto $\{y_1, \ldots, y_k\}$ de $k > 1$ puntos autoconsistentes. Luego, $ker(\mathbf{\Gamma}) \cap \mathcal{V} = \{0\}$ y $\mathbf{\Gamma}\mathcal{V} \cap \mathcal{V}^{\perp} = \{0\}$.*

Podemos ahora sí encarar el teorema principal.

Teorema 5.2. *Sea V un elemento elíptico con $E(V) = 0$ y operador de covarianza compacto $\mathbf{\Gamma}$. Sea $\mathcal{V}$ el subespacio generado por un conjunto $\{y_1, \ldots, y_k\}$ de $k > 1$ puntos autoconsistentes. Luego $\mathcal{V}$ estará generado por un conjunto de autovectores de $\mathbf{\Gamma}$.*

Demostración. Sea $q = dim(\mathcal{V})$, consideramos $\{v_1, \ldots, v_q\}$ una base ortonormal de $\mathcal{V}$. Definimos $A_1 : \mathcal{H} \to I\!R^q$ como

$$A_1(h) = \begin{pmatrix} < v_1, h > \\ \vdots \\ < v_q, h > \end{pmatrix} = \sum_{i=1}^{q} < v_i, h > \mathbf{e}_i \ ,$$

siendo $\mathbf{e}_i$ la base canónica de $I\!R^q$. Luego, $A_1^* : I\!R^q \to \mathcal{H}$, será igual a $A_1^*\mathbf{x} = \sum_{i=1}^{q} x_i v_i$ con $\mathbf{x} = (x_1, \ldots, x_q)$, es decir la imagen de A_1^* es $\mathcal{V}$. Definimos también $A_2 : \mathcal{H} \to I\!R^{\mathbb{N}}$ como

$$A_2(h) = \begin{pmatrix} < v_{q+1}, h > \\ < v_{q+2}, h > \\ \vdots \end{pmatrix} .$$

Para poder asegurar la continuidad del segundo operador debemos dotar de una norma adecuada a $I\!R^{\mathbb{N}}$. Para que además sea un espacio de Hilbert, consideramos la norma ℓ^2, es decir, la norma dada por la raíz de la suma de los cuadrados de la sucesión, lo indicaremos entonces por $I\!R^{\mathbb{N}} = \ell^{\infty}$.

Notemos, por la identidad de Parseval, que $\|h\|^2 = \sum_{k=1}^{\infty} < h, v_k >^2$. Entonces,

$$\|A_2(h)\|^2 = \sum_{k=q+1}^{\infty} < h, v_k >^2 \leq \sum_{k=1}^{\infty} < h, v_k >^2 = \|h\|^2 .$$

Tenemos así que A_2 es continua, pues es acotado de norma menor o igual a 1. De hecho, la norma será exactamente 1, evaluando en elementos unitarios de $\mathcal{V}^{\perp}$ se alcanzará dicha norma.

Definimos Y un elemento aleatorio de manera tal que $P(Y = y_j) = P(V \in \mathcal{D}_j)$, con $\mathcal{D}_j$ el dominio de atracción del punto y_j con respecto a V, es decir, $Y = \sum_{i=1}^{k} y_i I_{V \in \mathcal{D}_i}$. Usando que A_2 es lineal y continuo, y que y_j es un elemento del conjunto autoconsistente, obtenemos que

$$E(A_2 V|Y = y_j) = E(A_2 V|V \in \mathcal{D}_j) = A_2 E(V|V \in \mathcal{D}_j) = A_2 y_j = 0$$

donde la última igualdad vale pues $y_j \in \mathcal{V}$ que es el núcleo de A_2. O sea que, $P(E(A_2 V|Y) = 0) = 1$ puesto que Y únicamente tomará los valores $y_1, \ldots, y_k$.

Por otra parte, tenemos

$$0 = E(A_2 V|Y) = E(E(A_2 V|A_1 V)|Y). \tag{11}$$

Ahora bien, si definimos $A : \mathcal{H} \to \ell^\infty$ y $W = (W_1, W_2) = (A_1(V)', A_2(V)') = A(V)$, tenemos que W es un elemento elíptico en ℓ^∞ con esperanza $A(E(V)) = 0$ y operador de covarianza $\mathbf{\Gamma}_{AV} = A\mathbf{\Gamma}_V A^*$ y además $W = (W_1, W_2)$ donde W_1 es de dimensión finita. Si llamamos $\eta_2 = E(W_2) = 0$ y $\eta_1 = E(W_1) = 0$, por el Teorema 4.8, tenemos que

$$E(A_2(V)|A_1(V)) \;\; = \;\; E(W_2|W_1) = \eta_2 + \mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1}(W_1 - \eta_1) = \mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1} A_1(V)$$

donde $\mathbf{\Gamma}_{W_2,W_1} = A_2 \mathbf{\Gamma} A_1^*$ y $\mathbf{\Gamma}_{W_1,W_1} = A_1 \mathbf{\Gamma} A_1^*$.

Esto último vale siempre y cuando $\mathbf{\Gamma}_{W_1,W_1}$ sea inversible, hecho que probaremos ahora utilizando que $ker(\mathbf{\Gamma}) \cap \mathcal{V} = \{0\}$ y que $\mathbf{\Gamma}\mathcal{V} \cap \mathcal{V}^\perp = \{0\}$. Notemos en primer lugar que $\mathbf{\Gamma}_{W_1,W_1}$ es un endomorfismo entre espacios vectoriales de dimensión finita. Luego, basta probar que es un monomorfismo para tener automáticamente que es un isomorfismo. Veamos entonces la inyectividad de este operador.

Supongamos que $A_1 \mathbf{\Gamma} A_1^* \mathbf{h} = \mathbf{0}$, con $\mathbf{h} \in I\!R^q$. Queremos ver que $\mathbf{h} = \mathbf{0}$. Como $A_1 \mathbf{\Gamma} A_1^* \mathbf{h} = \mathbf{0}$ entonces $\mathbf{\Gamma} A_1^* \mathbf{h} \in Ker(A_1) = \mathcal{V}^\perp$. Ahora bien, $A_1^* \mathbf{h} \in \mathcal{V}$ entonces lo que se tiene es que $\mathbf{\Gamma} A_1^* \mathbf{h} \in \mathbf{\Gamma}(\mathcal{V}) \cap \mathcal{V}^\perp$, y esto por hipótesis vale $\{0\}$ con lo cual $\mathbf{\Gamma} A_1^* \mathbf{h} = 0$. Entonces, $A_1^* \mathbf{h} \in Ker(\mathbf{\Gamma})$ pero como $A_1^* \mathbf{h} \in \mathcal{V}$ entonces $A_1^* \mathbf{h} \in \; Ker(\mathbf{\Gamma}) \cap \mathcal{V}$ que vale $\{0\}$ por hipótesis. Luego $A_1^* \mathbf{h} = 0$ y como A_1^* es inyectiva de aquí sale que $\mathbf{h} = \mathbf{0}$ que era lo que se quería probar.

Tenemos así que $\mathbf{\Gamma}_{W_1,W_1}$ es inversible, luego tiene sentido plantear

$$E(A_2(V)|A_1(V)) \;\; = \;\; \mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1} A_1(V) \; .$$

Luego, de (11) obtenemos

$$\begin{aligned} 0 = E(A_2 V|Y) &= E(\mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1} A_1(V)|Y) = \mathbf{\Gamma}_{W_2,W_1}(\mathbf{\Gamma}_{W_1,W_1})^{-1} A_1 \; E(V|Y) \\ &= A_2 \mathbf{\Gamma} A_1^* (A_1 \mathbf{\Gamma} A_1^*)^{-1} A_1 \; Y \quad \text{c.t.p.} \end{aligned}$$

donde la última igualdad es por el Lema 5.5.

El soporte de Y genera $\mathcal{V}$, entonces el soporte de $A_1 Y$ genera $I\!R^q$. Entonces, como $(A_1 \mathbf{\Gamma} A_1^*)^{-1}$ es un endomorfismo de $I\!R^q$, se cumple que $\{A_1^*(A_1 \mathbf{\Gamma} A_1^*)^{-1} A_1 y_1, \ldots, A_1^*(A_1 \mathbf{\Gamma} A_1^*)^{-1} A_1 y_k\}$ genera $I\!R^q$. Luego, como $P\left(A_2 \mathbf{\Gamma} A_1^* (A_1 \mathbf{\Gamma} A_1^*)^{-1} A_1 Y = 0\right) = 1$, se tiene

$$A_2 \mathbf{\Gamma} A_1^* (A_1 \mathbf{\Gamma} A_1^*)^{-1} A_1 \; y_1 = 0 \quad \ldots \quad A_2 \mathbf{\Gamma} A_1^* (A_1 \mathbf{\Gamma} A_1^*)^{-1} A_1 \; y_k = 0.$$

es decir valdria que $\mathbf{\Gamma}\mathcal{V} \cap \mathcal{V}^{\perp} = \{0\}$, y por lo tanto, $\forall \mathbf{x} \in I\!R^q$, $A_2\mathbf{\Gamma}A_1^*\mathbf{x} = 0$, o equivalentemente, $A_2\mathbf{\Gamma}A_1^* : I\!R^q \to \ell^\infty$ es el operador nulo. Valdrá lo mismo para $A_1\mathbf{\Gamma}A_2^* : \ell^\infty \to I\!R^q$, por ser el adjunto de $A_2\mathbf{\Gamma}A_1^*$.

Definamos $P_\mathcal{V} = A_1^*A_1 : \mathcal{H} \to \mathcal{V}$, o sea, la proyección sobre $\mathcal{V}$ y $P_{\mathcal{V}^\perp} = A_2^*A_2 : \mathcal{H} \to \mathcal{V}^\perp$, la proyección sobre $\mathcal{V}^\perp$. Luego, por lo anterior, $P_{\mathcal{V}^\perp}\mathbf{\Gamma}P_\mathcal{V} = 0$ y $P_\mathcal{V}\mathbf{\Gamma}P_{\mathcal{V}^\perp} = 0$. Entonces esto nos dice que $\mathbf{\Gamma}(\mathcal{V}) \subset \mathcal{V}$ y que $\mathbf{\Gamma}(\mathcal{V}^\perp) \subset \mathcal{V}^\perp$. O sea que $\mathcal{V}$ y $\mathcal{V}^\perp$ son $\mathbf{\Gamma}-$invariantes. Como comentario esto diría que $\mathcal{V}$ descompone a $\mathbf{\Gamma}$.

Tiene sentido pensar en $\mathbf{\Gamma} : \mathcal{V} \to \mathcal{V}$. Ahora bien, $\mathbf{\Gamma}$ es compacto y autoadjunto, luego será diagonalizable por el Teorema 6.12. Pero además, $\mathbf{\Gamma}$ restringido a $\mathcal{V}$ también será compacto y autoadjunto, será entonces a su vez diagonalizable con las mismas autofunciones. Entonces tenemos que $\mathcal{V}$, el dominio de $\mathbf{\Gamma}|\mathcal{V}$ estará generado por un conjunto de autofunciones de $\mathbf{\Gamma}$. Como $\mathcal{V}$ tiene dimensión q, luego serán q autofunciones y tenemos así lo deseado. $\square$

Teorema 5.3. *Sea V un elemento elíptico con $E(V) = 0$ y operador de covarianza compacto $\mathbf{\Gamma}$. Si k de puntos principales de V generan un subespacio $\mathcal{V}$ de dimensión q entonces este subespacio estará además generado por las q autofunciones de $\mathbf{\Gamma}$ asociados a los q autovalores más grandes.*

Demostración. Sean $\lambda_1 \geq \lambda_2 \dots$ los autovalores ordenados de $\mathbf{\Gamma}$, con los correspondientes autovectores $\beta_1, \beta_2, \dots$. Sea $\{y_1, \dots, y_k\}$ un conjunto de k puntos principales que generan $\mathcal{V}$. Por el teorema anterior, $\mathcal{V}$ estará generado por q autofunciones de $\mathbf{\Gamma}$. Sea r un entero tal que $\{\beta_1, \dots, \beta_r\}$ contiene a las q autofunciones que generan $\mathcal{V}$.

Entonces, para cada punto principal y_j existen coeficientes a_{ji} de manera tal que $y_j = \sum_{i=1}^r a_{ji}\beta_i$. Definimos $\mathbf{a}_j = (a_{j1}, \dots, a_{jr})'$ y $\mathbf{X} = (< \beta_1, V >, \dots, < \beta_r, V >)'$.

Las autofunciones conforman un conjunto ortonormal, luego tendremos que

$$||V - y_j||^2 = ||\sum_{i=1}^{\infty} < \beta_i, V > \beta_i - \sum_{i=1}^{r} a_{ji}\beta_i||^2 = ||\sum_{i=1}^{r}(< \beta_i, V > -a_{ji})\beta_i||^2 + ||\sum_{i=r+1}^{\infty} < \beta_i, V > \beta_i||^2$$

$$= \sum_{i=1}^{r}(< \beta_i, V > -a_{ji})^2 + \sum_{i=r+1}^{\infty} \lambda_i = ||\mathbf{X} - \mathbf{a}_j||_r^2 + \sum_{i=r+1}^{\infty} \lambda_i.$$

donde $||\cdot||_r$ es la norma euclídea en $\mathbb{R}^r$. Por definición, los k puntos principales son los k puntos tales que minimizan el siguiente error cuadrático medio:

$$MSE(V, \{y_1, \dots, y_k\}) = E(min_j||V - y_j||^2) = E(min_j||\mathbf{X} - \mathbf{a}_j||_r^2) + \sum_{i=r+1}^{\infty} \lambda_i.$$

De acá obtenemos que

$$MSE(\mathbf{X}, \{\mathbf{a}_1, \dots, \mathbf{a}_k\}) = MSE(V, \{y_1, \dots, y_k\}) - \sum_{i=r+1}^{\infty} \lambda_i.$$

Pero $\{y_1, \ldots, y_k\}$ son los puntos que minimizan $MSE(V, \cdot)$. Veamos entonces que $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\}$ son los que minimizan $MSE(\mathbf{X}, \cdot)$. En efecto, sean $\{\mathbf{b}_1, \ldots, \mathbf{b}_k\}$ otro conjunto.

Veamos que $MSE(V, \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}) \leq MSE(V, \{\mathbf{b}_1, \ldots, \mathbf{b}_k\})$. Definimos $x_j = \sum_{i=1}^{r} b_{ji}\beta_i$. Repitiendo la cuenta anterior obtendremos que

$$||V - x_j||^2 = ||\mathbf{X} - \mathbf{b}_j||^2 + \sum_{i=r+1}^{\infty} \lambda_i.$$

Entonces,

$$MSE(\mathbf{X}, \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}) = MSE(V, \{y_1, \ldots, y_k\}) - \sum_{i=r+1}^{\infty} \lambda_i \leq$$

$$MSE(V, \{x_1, \ldots, x_k\}) - \sum_{i=r+1}^{\infty} \lambda_i = MSE(\mathbf{X}, \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}).$$

O sea,

$$MSE(\mathbf{X}, \{\mathbf{a}_1, \ldots, \mathbf{a}_k\}) \leq MSE(\mathbf{X}, \{\mathbf{b}_1, \ldots, \mathbf{b}_k\}).$$

Tenemos así que $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\})$ es un conjunto de k puntos principales de $\mathbf{X}$. Ahora bien, como V es elíptico entonces $\mathbf{X}$ también será elíptico, pues consiste en una aplicación lineal y continua sobre el elemento V. Estamos en un caso multivariado, luego por [13] tendremos que los k puntos principales de $\mathbf{X}$ estarán en un subespacio asociado a los q mayores autovalores de la covarianza de $\mathbf{X}$.

Bien, estudiemos con detalle $\mathbf{X}$. Definimos $A : \mathcal{H} \to \mathbb{R}^r$, dado por

$$A(v) = (< \beta_1, v >, \ldots, < \beta_r, v >)'.$$

Es sencillo demostrar que A es lineal. Utilizando la norma infinito en $\mathbb{R}^r$ sale bien que es acotado, y por ende es acotado en cualquier norma que se aplique en $\mathbb{R}^r$ por ser un espacio de dimensión finita. Tendremos su operador adjunto $A' : \mathbb{R}^r \to \mathcal{H}$ dado por

$$A'(\mathbf{x}) = \sum_{i=1}^{r} x_i\beta_i.$$

Nótese que $\mathbf{X} = AV$ y que $\mathbf{a}_j = Ay_j$. Luego, $\mathbf{\Gamma}_X = A\mathbf{\Gamma}_V A'$. Entonces,

$$\mathbf{\Gamma}_X(\mathbf{z}) = \mathbf{\Gamma}_X(z_1, \ldots, z_r)' = A\mathbf{\Gamma}_V A(z_1, \ldots, z_r)' = A\mathbf{\Gamma}_V(\sum_{i=1}^{r} z_i\beta_i) = A(\sum_{i=1}^{r} \lambda_i z_i \beta_i) = (\lambda_1 z_1, \ldots, \lambda_r z_r)'$$

de lo cual tenemos completamente caracterizado el operador $\mathbf{\Gamma}_X$ y tendremos que sus q mayores autovalores coincidirán precisamente con los q mayores autovalores de $\mathbf{\Gamma}_V$, y además los autovectores de $\mathbf{\Gamma}_X$ serán los elementos de la forma $A\beta_i$.

Juntando todo tendremos así que los $\{\mathbf{a}_1, \ldots, \mathbf{a}_k\} = \{Ay_1, \ldots, Ay_k\}$ son los k puntos principales de $\mathbf{X} = AV$ y que el subespacio que ellos generan está también generado por elementos de la pinta $A\beta_i$ con $i = 1, \ldots, q$. Es fácil probar que restringiendo A al espacio $\mathcal{M} =< \beta_1, \ldots, \beta_r >$ tendremos que esto es una isometría suryectiva. Luego, si consideramos $\tilde{A} : \mathcal{M} \to \mathbb{R}^r$ como la restricción de A al subespacio $\mathcal{M}$, su inversa estará dada por $\tilde{A}' : \mathbb{R}^r \to \mathcal{M}$, que es en esencia igual que A' salvo por el espacio de llegada.

Como los y_j se encuentran en $\mathcal{M}$ entonces $\{Ay_1, \ldots, Ay_k\} = \{\tilde{A}y_1, \ldots, \tilde{A}y_k\}$. A su vez, esto estará incluido en $< \tilde{A}\beta_1, \ldots, \tilde{A}\beta_q >$. Aplicando $\tilde{A}^{-1} = \tilde{A}'$ tendremos así que $\{y_1, \ldots, y_k\}$ estará incluido en $< \beta_1, \ldots, \beta_q >$ que son los autovectores asociados a los q mayores autovalores de $\mathbf{\Gamma}_V$, terminando así de probar el resultado.

$\square$

5.3. Propiedades de los puntos principales. Resolución para el caso $k = 2$

Para el caso $k = 1$ vimos que la opción óptima consiste en tomar como punto principal la media. El objetivo será obtener fórmulas explícitas para los puntos principales en el caso $k = 2$. No se conoce un resultado general para cualquier valor de k, ni siquiera en los casos univariado o multivariado. El siguiente resultado será de utilidad para la tarea en cuestión y generaliza un resultado dado, para el caso de vectores de dimensión finita, por Flury (1990). Notemos que, para este resultado, no es necesario pedirle al elemento aleatorio que tenga distribución elíptica.

Teorema 5.4. *Sea $V : \Omega \to \mathcal{H}$ un elemento aleatorio de media μ y con k puntos principales $\xi_1, \ldots, \xi_k \in \mathcal{H}$. Entonces, rango$(\xi_1 - \mu, \ldots, \xi_k - \mu) < k$.*

Demostración. Sean $c_1, \ldots, c_k \in \mathcal{H}$ elementos arbitrarios, y sea $b_i = c_i - c_k$, $1 \leq i \leq k - 1$. Sea $m = \text{rango}(b_1, \ldots, b_{k-1}) \leq k - 1$. Sea $a_1, \ldots, a_m$ una base ortonormal del subespacio $\mathcal{M}_1$ generado por $b_1, \ldots, b_{k-1}$ y sea $\mathcal{M}_2 = \mathcal{M}_1^{\perp}$. $\mathcal{M}_2$ estará generado por $a_{m+1}, a_{m+2}, \ldots,$ una completación ortogonal de $\mathcal{M}_1$ a todo el espacio $\mathcal{H}$.

Consideramos $A_1 : \mathcal{H} \to \ell^\infty$ definido como

$$A_1(a_i) = \begin{cases} (e_i)_{\ell^\infty} & i = 1, 2, \ldots, m \\ 0 & i > m \end{cases}$$

Análogamente, tendremos $A_2 : \mathcal{H} \to \ell^\infty$ definido como

$$A_2(a_i) = \begin{cases} 0 & i = 1, 2, \ldots, m \\ (e_i)_{\ell^\infty} & i > m \end{cases}$$

En base a eso definimos $A : \mathcal{H} \to \ell^\infty$, $A(h) = A_1(h) + A_2(h)$. Notemos entonces que $A(a_i) = e_i$, con lo cual A resultará ser una isometría suryectiva, por lo tanto una aplicación unitaria, su inversa será su conjugada, que será un operador $A^* : \ell^\infty \to \mathcal{H}$ tal que $A^*(e_i) = a_i$, la inversa de A. Definimos

$$d_i = Ac_i = A_1c_i + A_2c_i = d_i^{(1)} + d_i^{(2)}.$$

Por otro lado,
$$d_i^{(2)} - d_j^{(2)} = A_2 c_i - A_2 c_j = A_2(c_i - c_j)$$
pero $c_i - c_j = b_i - b_j$ está en $\mathcal{M}_1$, luego su imagen por A_2 será 0. Entonces, vale que $d_i^{(2)} - d_j^{(2)} = 0$, o sea que todos los $d_i^{(2)}$ son iguales a un valor dado que llamaremos $d^{(2)}$. Definimos
$$W_1 = A_1 V, \quad W_2 = A_2 V, \quad W = W_1 + W_2 = AV.$$
Como A es unitaria, se cumple
$$E(d^2(V|c_1, \ldots, c_k)) = E(d^2(AV|Ac_1, \ldots, Ac_k)) = E(d^2(W|d_1, \ldots, d_k)). \tag{12}$$
Usaremos la descomposición ortogonal de los d_i como $d_i = d_i^{(1)} + d^{(2)}$, con los sumandos ortogonales entre sí. De la misma forma, $W = W_1 + W_2$, también una descomposición ortogonal. Luego, $\|W - d_i\|^2 = \|W_1 - d_i^{(1)}\|^2 + \|W_2 - d^{(2)}\|^2$, con lo cual tendremos que (12) es expresable como
$$E(d^2(V|c_1, \ldots, c_k)) = E(d^2(W_1|d_1^{(1)}, \ldots, d_k^{(1)})) + E(d^2(W_2|d^{(2)}, \ldots, d^{(2)})).$$
Ahora bien, V tiene media μ, luego $W = AV$ tendrá media $A\mu$, y en particular W_2 tendrá media $A_2\mu$. Podemos suponer $\mu = 0$, sin perder generalidad puesto que en esencia lo que se hace es trasladar el origen de coordenadas. En el caso general bastará aplicar el resultado al elemento restado con su media y apelar a la invariancia traslacional. Así, nos quedará que W_2 tendrá esperanza 0. El segundo sumando de la expresión de arriba involucra k puntos que son todos iguales, luego es como tener $E(d^2(W_2|d^{(2)}))$, y esto se minimiza cuando $d^{(2)} = E(W_2) = A_2\mu = 0$. Tenemos entonces que
$$E(d^2(V|c_1, \ldots, c_k)) \geq E(d^2(W_1|d_1^{(1)}, \ldots, d_k^{(1)})) + E(d^2(W_2|0, \ldots, 0)),$$
alcanzándose la igualdad cuando $d^{(2)} = 0$. Definimos así
$$c_i^* = A_1^* d_i^{(1)} + A_2^* A_2 \mu = A_1^* A_1 c_i + A_2^* A_2 \mu = P_{\mathcal{M}_1} c_i + P_{\mathcal{M}_2} \mu = P_{\mathcal{M}_1} c_i.$$
Tenemos que
$$\begin{aligned}
E(d^2(V|c_1, \ldots, c_k)) &\geq E(d^2(W_1|d_1^{(1)}, \ldots, d_k^{(1)})) + E(d^2(W_2|0, \ldots, 0)) \\
&= E(d^2(A_1^* W_1|A_1^* d_1^{(1)}, \ldots, A_1^* d_k^{(1)})) + E(d^2(A_2^* W_1|0, \ldots, 0)) \\
&= E(d^2(V_1|A_1^* d_1^{(1)}, \ldots, A_1^* d_k^{(1)})) + E(d^2(V_2|0, \ldots, 0)) \\
&= E(d^2(V_1|A_1^* A_1 c_1, \ldots, A_1^* A_1 c_k)) + E(d^2(V_2|0, \ldots, 0))
\end{aligned}$$
con $V_1 = A_1^* W_1$ y $V_2 = A_2^* W_2$ ($V = V_1 + V_2$). Usando nuevamente la ortogonalidad de la descomposición, tendremos que esto es
$$E(d^2(V|A_1^* A_1 c_1 + 0, \ldots, A_1^* A_1 c_k + 0)) = E(d^2(V|c_1^*, \ldots, c_k^*)).$$

Resumiendo,
$$E(d^2(V|c_1^*,\ldots,c_k^*)) \leq E(d^2(V|c_1,\ldots,c_k))$$
valiendo la igualdad si $c_i = A_1^* A_1 c_i = P_{\mathcal{M}_1} c_i$. En los puntos principales vale la igualdad puesto que son los que minimizan esa expresión. Deducimos entonces que si los c_i corresponden a puntos principales (o sea ξ_i) entonces cada c_i deberá residir en $\mathcal{M}_1$, pero como $\mathcal{M}_1$ tiene dimensión menor o igual a $k-1$ deducimos entonces que rango$(\xi_1-\mu,\ldots,\xi_k-\mu) < k$. □

Este resultado dice que todo hiperplano que contenga a los k puntos principales necesariamente deberá contener a la media μ. Notemos que en el caso $k=1$, el subespacio resultante es de dimensión cero, por lo tanto el punto principal deberá coincidir con la media. Tendríamos así el siguiente corolario

Corolario 5.6. *Para $k=1$, la media será un punto principal.*

Estudiamos ahora el caso $k=2$. Precisaremos la elipticidad de la distribución para los resultados que siguen. Antes, enunciamos y demostramos un lema auxiliar propio de variables aleatorias reales.

Lema 5.8. *Sean Y_1 e Y_2 dos variables aleatorias reales tal que Y_2 tiene la misma distribución que ρY_1 para algún valor de ρ. Luego,*
$$P_{Y_1}(k)/Var(Y_1) = P_{Y_2}(k)/Var(Y_2)$$

Demostración. Y_2 tiene la misma distribución que ρY_1, entonces vale que $P_{\rho Y_1}(k) = P_{Y_2}(k)$. Ahora bien,
$$P_{\rho Y_1}(k) = E(d^2(\rho Y_1|\xi_1,\ldots,\xi_k)) = \rho^2 E(d^2(Y_1|\xi_1,\ldots,\xi_k)) = \rho^2 P_{Y_1}(k).$$
Pero $Y_2 \stackrel{\mathcal{D}}{=} \rho Y_1$, luego $Var(Y_2) = Var(\rho Y_1) = \rho^2 Var(Y_1)$, con lo cual obtenemos que $\rho^2 = Var(Y_2)/Var(Y_1)$. Entonces,
$$P_{Y_2}(k) = P_{\rho Y_1}(k) = \rho^2 P_{Y_1}(k) = P_{Y_1}(k)\ Var(Y_2)/Var(Y_1)\ ,$$
y se deduce así lo deseado. □

Ahora sí, encaramos el resultado principal que generaliza el Teorema 2 de Flury (1990) dado para el caso de vectores de dimensión finita. Precisaremos calcular de manera auxiliar los puntos principales de variables aleatorias reales. Las condiciones para que los puntos principales de una variable aleatoria real existan están dados en el Teorema 1 de Flury (1990). Entre las condiciones pedidas está la de existir una densidad continua y simétrica alrededor de su media.

Teorema 5.5. *Sea V un elemento aleatorio elíptico de media μ y operador de covarianza $\mathbf{\Gamma}$ con traza finita. Supongamos que variable aleatoria real $Y = < v, V-\mu >$ tiene dos puntos principales para todo $v \in \mathcal{H}$. Luego, V tiene dos puntos principales de la forma $y_1 = \mu + \gamma_1\beta$ e $y_2 = \mu + \gamma_2\beta$, con $\beta \in \mathcal{H}$ un autovector de $\mathbf{\Gamma}$ de norma 1, asociado al mayor autovalor. Además, γ_1, γ_2 son los dos puntos principales de la variable aleatoria real $< \beta, V-\mu >$.*

Demostración. Supondremos sin pérdida de generalidad $\mu = 0$. Para simplificar la notación definimos $P_V(c_1, c_2) = E(d^2(V|c_1, c_2))$.

Primero, demostremos que P_V se minimiza si los dos elementos $c_1, c_2 \in \mathcal{H}$ se eligen de manera tal que queden sobre una recta que pasa por la media, en nuestro caso $\mu = 0$ y con dirección $c_2 - c_1$. Para esto basta recurrir al Teorema 5.4, en cuya demostración vimos que cada uno de los puntos principales y_i si existen pertenece al subespacio $\mathcal{M}_1$ generado por $y_2 - y_1$. O sea, ambos puntos caen en una recta que pasa por el origen (que corresponde a la media) y con dirección $a_1 = (y_2 - y_1)/\|y_2 - y_1\|$.

Dados $c_1, c_2 \in \mathcal{H}$, $c_1 \neq c_2$, sea $\mathcal{M}_1$ el subespacio de dimensión 1 generado por $a_1 = (c_2 - c_1)/\|c_2 - c_1\|$ y $\mathcal{M}_2 = \mathcal{M}_1^{\perp}$ con base ortonormal $\{a_j : j \geq 2\}$. Ahora bien, retomando nuevamente la notación de la demostración del Teorema 5.4, consideramos $A_1 : \mathcal{H} \to \ell^\infty$ definido como $A_1(a_j) = 0$ si $j \geq 2$ y $A_1(a_1) = e_1$ siendo e_j el elemento de ℓ^∞ con su $j-$ésima coordenada igual a 1 y las demás iguales a 0 y $A_2 : \mathcal{H} \to \ell^\infty$ definido como $A_2(a_1) = 0$, $A_2(a_i) = e_i$ si $i > 1$. Sea $W_1 = A_1 V =< a_1, V > e_1$, $W_2 = A_2 V$, $W = W_1 + W_2$. Como $\mathcal{M}_1$ es un subespacio de dimensión 1 entonces W_1 tiene la misma distribución que la variable aleatoria $Y_1 =< a_1, V >$ que es elíptica pues V es elíptico. Siendo elíptica univariada, entonces es simétrica alrededor del cero pues $\mu = 0$. Recordemos que si $d_i = Ac_i = A_1 c_i + A_2 c_i = d_i^{(1)} + d_i^{(2)}$, entonces $d_1^{(2)} = d_2^{(2)} = d^{(2)}$ y

$$\begin{aligned} E(d^2(V|c_1, c_2)) &= E(d^2(W|d_1, d_2)) = E(d^2(W_1|d_1^{(1)}, d_2^{(1)})) + E(d^2(W_2|d^{(2)}, d^{(2)})) \\ &\geq E(d^2(W_1|d_1^{(1)}, d_2^{(1)})) + E(d^2(W_2|0, 0)). \end{aligned}$$

Luego, para c_1 y c_2 fijos, $E(d^2(W_1|d_1^{(1)}, d_2^{(1)}))$ se puede minimizar tomando como $d_1^{(1)}$ y $d_2^{(1)}$ los puntos principales de $Y_1 =< a_1, V >$. Sean ξ_1 y ξ_2 los puntos principales de Y_1. Definamos $d_1^* = \xi_1 e_1$, $d_2^* = \xi_2 e_1$. Se sigue así que

$$E(d^2(W|d_1^*, d_2^*)) \leq E(d^2(W|d_1, d_2))$$

valiendo la igualdad en el caso $d_1 = d_1^*$ y $d_2 = d_2^*$.

Como $W = AV$ entonces $V = A^* W$. Así,

$$c_1^* = A^* d_1^* = A^* \xi_1 e_1 = \xi_1 A_1^* e_1 = \xi_1 a_1 = \xi_1 (c_2 - c_1)/\|c_2 - c_1\|.$$

Análogamente, $c_2^* = \xi_2 (c_2 - c_1)/\|c_2 - c_1\|$. Entonces,

$$E(d^2(V|\xi_1 a_1, \xi_2 a_1)) = E(d^2(V|c_1^*, c_2^*)) \leq E(d^2(V|c_1, c_2)) \ .$$

Ahora bien, dado $a \in \mathcal{H}$ tal que $\|a\| = 1$, para cualquier par c_1, c_2 tal que $c_2 - c_1$ sea proporcional al elemento a, se tendrá

$$E(d^2(V|\xi_1 a, \xi_2 a)) = E(d^2(V|c_1^*, c_2^*)) \leq E(d^2(V|c_1, c_2)) \ .$$

Tenemos así que para cualquier elemento $a \in \mathcal{H}$ tal que $\|a\| = 1$, entonces es posible determinar los puntos principales de V calculando primero los de $W_1 = A_1 V$ y luego definiendo $c_1^* = \xi_1 a$ y $c_2^* = \xi_2 a$ y luego minimizando en a.

Queda entonces como cuestión final determinar el elemento $a \in \mathcal{H}$. Tengamos presente que el operador A_1 depende de ese elemento a, pues se define en base de $c_2 - c_1$ normalizado que no es otra cosa que a. Podemos así considerar $A_1^{(a)}$ para explicitar la dependencia con el elemento a, y análogamente $W_1^{(a)} = A_1^{(a)} V =< a, V > e_1 = Y_1^{(a)} e_1$. Ahora bien,

$$\Sigma_{Y_1^{(a)}} = Var(< a, V >) =< a, \mathbf{\Gamma} a > .$$

Por lo hecho en el primer paso, tendremos que los puntos principales estarán en una recta que pasa por el origen, con una cierta dirección normalizada a que es nuestro objetivo encontrar. Los elementos de dicha recta serán de la forma λa, con $\lambda \in \mathbb{R}$.

Por el Lema 5.8, entonces valdrá que

$$P_{Y_1^{(a)}}(2) =< a, \mathbf{\Gamma} a > \frac{P_{\lambda Y_1^{(a)}}(2)}{Var(\lambda Y_1^{(a)})}.$$

Ahora bien, por otro lado tendremos que

$$P_{\lambda Y_1^{(a)}}(2) = \min_{\eta_1, \eta_2 \in \mathbb{R}} E(\min_{i=1,2} \{\|\lambda Y_1^{(a)} - \eta_1\|^2, \|\lambda Y_1^{(a)} - \eta_2\|^2\}.$$

Esto se está minimizando sobre todos los pares $\eta_1,\ \eta_2$ en $\mathbb{R}$, será por lo tanto menor o igual al caso particular en que se considere $\eta_1 = 0,\ \eta_2 = 0$. Pero ahí lo que obtendríamos sería precisamente la varianza de $\lambda Y_1^{(a)}$. Así, tendremos que

$$\frac{P_{\lambda Y_1^{(a)}}(2)}{Var(\lambda Y_1^{(a)})} < 1.$$

Podemos mejorar un poco la acotación. En primer lugar, es claro que lo de arriba no dependerá de λ. Este término se factoriza tanto en el numerador como en el numerador elevado al cuadrado y se puede simplificar. Es en el fondo lo que afirma el lema 5.8, tendríamos que

$$\frac{P_{\lambda Y_1^{(a)}}(2)}{Var(\lambda Y_1^{(a)})} = \frac{P_{Y_1^{(a)}}(2)}{Var(Y_1^{(a)})}.$$

Es posible mejorar más la acotación para tener independencia no sólo con respecto a λ sino también con respecto a a.

Por ser V elíptico, dado $B : \mathcal{H} \to I\!R^p$ lineal y acotado $\mathbf{Y} = BV$ tiene distribución elíptica de parámetros $B\mu = 0$ y $\mathbf{\Sigma} = B\mathbf{\Gamma}B^*$, luego su función característica es de la forma $\varphi_{\mathbf{Y}}(\mathbf{y}) = \phi(\mathbf{y}'\mathbf{\Sigma}\mathbf{y})$ con ϕ independiente de B. Entonces, $\varphi_{Y_1^{(a)}}(b) = \phi(b^2 < a, \mathbf{\Gamma} a >) = \phi(b^2 Var(Y_1^{(a)}))$. Esto implica que $Z_a = Y_1^{(a)} / \sqrt{Var(Y_1^{(a)})}$ tendrá siempre la misma distribución para cualquier elemento $a \in \mathcal{H}$.

Entonces,

$$\begin{aligned}
\frac{P_{Y_1^{(a)}}(2)}{Var(Y_1^{(a)})} &= \frac{\text{mín}_{\eta_1,\eta_2\in\mathbb{R}} E(\text{mín}\{|Y_1^{(a)}-\eta_1|^2, |Y_1^{(a)}-\eta_2|^2\})}{Var(Y_1^{(a)})} \\
&= \min_{\eta_1,\eta_2\in\mathbb{R}} E(\text{mín}\{\left(\frac{|Y_1^{(a)}-\eta_1|^2}{\sqrt{Var(Y_1^{(a)})}}\right)^2, \left(\frac{|Y_1^{(a)}-\eta_2|^2}{\sqrt{Var(Y_1^{(a)})}}\right)^2\}) \\
&= \min_{\eta_1',\eta_2'\in\mathbb{R}} E(\text{mín}\{\left(|Z_a-\eta_1'|^2\right)^2, \left(|Z_a-\eta_2'|^2\right)^2\})
\end{aligned}$$

y por tanto, no depende del elemento a. Entonces, podemos afirmar que

$$\frac{P_{\lambda Y_1^{(a)}}(2)}{Var(\lambda Y_1^{(a)})} = g < 1$$

con g independiente de a. Concluimos así que $P_{Y_1^{(a)}}(2) = g < a, \mathbf{\Gamma} a >$, con $g < 1$ e independiente del elemento a.

Podemos escribir

$$V = Y_1^{(a)}\, a + (V - Y_1^{(a)}\, a) = P_{<a>}V + P_{<a>^\perp}V.$$

Entonces, considerando ξ_i^a los puntos principales de $Y_1^{(a)}$, obtenemos

$$\|V - \xi_i^{(a)}\, a\|^2 = \|Y_1^{(a)}\, a - \xi_i^{(a)}\, a + (V - Y_1^{(a)}\, a)\|^2 = \|a\cdot(Y_1^{(a)} - \xi_i^a) + P_{<a>^\perp}V\|^2.$$

Por el teorema de Pitágoras (el primer sumando está en $< a >$ y el segundo en su ortogonal) y usando además que $\|a\| = 1$ tendremos que

$$\|V - \xi_i^{(a)}\, a\|^2 = (Y_1^{(a)} - \xi_i^a)^2 + \|P_{<a>^\perp}V\|^2.$$

Tomando mínimo entre i=1,2 y luego aplicando esperanza llegamos a lo siguiente

$$E(d^2(V|\xi_1^{(a)}\, a, \xi_2^{(a)}\, a)) = E(\min_{i=1,2}\|V - \xi_i^{(a)}\, a\|^2) = P_{Y_1^{(a)}}(2) + E(\|P_{<a>^\perp}V\|^2) = g < a, \mathbf{\Gamma} a > + E(\|P_{<a>^\perp}V\|^2).$$

pues $P_{Y_1^{(a)}}(2) = g < a, \mathbf{\Gamma} a >$. Vamos a estudiar el otro sumando. Llamemos $Z = P_{<a>^\perp}V$ y sea β_n la base ortonormal de $\mathcal{H}$ formada por las autofunciones de $\mathbf{\Gamma}$ asociadas a los autovalores $\lambda_1 \geq \lambda_2 \geq \ldots$, entonces,

$$\begin{aligned}
E(\|P_{<a>^\perp}V\|^2) &= E(\|Z\|^2) = E(\|V\|^2) - E(\|P_{<a>}V\|^2) = Var(V) - Var(Y_1^{(a)}) \\
&= \sum_{n=1}^{\infty} < \mathbf{\Gamma}\beta_n, \beta_n > - Var(Y_1^{(a)}) = \sum_{j\geq 1}\lambda_j - Var(Y_1^{(a)}) \\
&= tr(\mathbf{\Gamma}) - Var(Y_1^{(a)}) = tr(\mathbf{\Gamma}) - < a, \mathbf{\Gamma} a > .
\end{aligned}$$

Por lo tanto,

$$E(d^2(V|a \cdot \xi_1^a, a \cdot \xi_2^a)) = g < a, \mathbf{\Gamma} a > + tr(\mathbf{\Gamma}) - < a, \mathbf{\Gamma} a >= tr(\mathbf{\Gamma}) - (1-g) < a, \mathbf{\Gamma} a > .$$

Para minimizar esto se debe maximizar $< a, \mathbf{\Gamma} a >$ entre todos los elementos a de norma 1. Apelando a la compacidad del operador de covarianza $\mathbf{\Gamma}$ obtendríamos que esto se consigue escogiendo a como la autofunción asociada al mayor autovalor de $\mathbf{\Gamma}$.

Por lo tanto, se empieza el procedimiento considerando a como la autofunción asociada al mayor autovalor de $\mathbf{\Gamma}$, obtenemos los puntos principales de $W_a = A_1 V$ y luego multiplicamos a cada punto principal (real) por el elemento a para tener así los dos puntos principales de V. $\square$

6. Apéndice

6.1. Álgebra Matricial

Los siguientes resultados pueden encontrarse en Seber (1984).

Teorema 6.1. *Descomposición Espectral*
Sea $\mathbf{A} \in R^{d\times d}$ *una matriz simétrica. Luego existe una matriz ortogonal* $\mathbf{T} = (\mathbf{t}_1, \ldots, \mathbf{t}_d)$ *tal que* $\mathbf{T}'\mathbf{A}\mathbf{T} = diag(\lambda_1, \ldots, \lambda_d) = \mathbf{\Lambda}$*, con* $\lambda_1 \geq \ldots \geq \lambda_d$ *los autovalores de* $\mathbf{A}$.

Teorema 6.2. *Para matrices compatibles con el producto, los autovalores no nulos de* $\mathbf{AB}$ *son los mismos que los de* $\mathbf{BA}$*. Si las matrices son cuadradas entonces serán todos idénticos.*

Definición 6.1. Una matriz simétrica $\mathbf{A}$ se dice semidefinida (definida) positiva si $\mathbf{x}'\mathbf{A}\mathbf{x} \geq 0$ (> 0) para todo $\mathbf{x} \neq 0$. Diremos entonces que $\mathbf{A} \geq \mathbf{0}$ $(\mathbf{A} > \mathbf{0})$ con $\mathbf{0}$ la matriz de ceros.

Proposición 6.1. *Dado* $\mathbf{A} \geq 0$*, luego* $\mathbf{C}'\mathbf{A}\mathbf{C} \geq 0$*. En particular* $\mathbf{C}'\mathbf{C} \geq 0$.

Teorema 6.3. *Si* $\mathbf{M}$ *y* $\mathbf{N}$ *son dos matrices definidas positivas entonces*

$$\sup_{\mathbf{x},\mathbf{y}} \frac{(\mathbf{x}'\mathbf{L}\mathbf{y})^2}{\mathbf{x}'\mathbf{M}\mathbf{x} \cdot \mathbf{y}'\mathbf{N}\mathbf{y}} = \theta_{max} \tag{13}$$

con θ_{max} *el mayor autovalor de* $\mathbf{M}^{-1}\mathbf{L}\mathbf{N}^{-1}\mathbf{L}'$ *y de* $\mathbf{N}^{-1}\mathbf{L}'\mathbf{M}^{-1}\mathbf{L}$*. El supremo se alcanza cuando* $\mathbf{x}$ *es un autovector de* $\mathbf{M}^{-1}\mathbf{L}\mathbf{N}^{-1}\mathbf{L}'$ *correspondiente a* θ_{max} *e* $\mathbf{y}$ *lo es de* $\mathbf{N}^{-1}\mathbf{L}'\mathbf{M}^{-1}\mathbf{L}$*, también correspondiendo a* θ_{max}.

Teorema 6.4. *Sea* $\mathbf{C} \in R^{p\times q}$ *una matriz simétrica de rango* m *y sean* $\rho_1^2 \geq \ldots \geq \rho_m^2 > 0$ *los autovalores no nulos de* $\mathbf{C}\mathbf{C}'$ *(y de* $\mathbf{C}'\mathbf{C}$*). Sean* $\mathbf{t}_1, \ldots, \mathbf{t}_m$ *los correspondientes autovectores de* $\mathbf{C}\mathbf{C}'$ *y sean* $\mathbf{w}_1, \ldots, \mathbf{w}_m$ *los de* $\mathbf{C}'\mathbf{C}$*. Si* $\mathbf{T}_k = (\mathbf{t}_1, \ldots, \mathbf{t}_k)$ *y* $\mathbf{W}_k = (\mathbf{w}_1, \ldots, \mathbf{w}_k)$ $(k < m)$*, entonces*

$$\sup_{\mathbf{T}_k'\mathbf{x}=0, \mathbf{W}_k'\mathbf{y}=0} \frac{(\mathbf{x}'\mathbf{C}\mathbf{y})^2}{\mathbf{x}'\mathbf{x} \cdot \mathbf{y}'\mathbf{y}} = \rho_{k+1}^2$$

el supremo se alcanza cuando $\mathbf{x} = \mathbf{t}_{k+1}$ *e* $\mathbf{y} = \mathbf{w}_{k+1}$.

Teorema 6.5. *Sean* $\mathbf{A}$ *y* $\mathbf{B}$ *matrices de* $I\!R^d$ *simétricas con autovalores* $\mu_1(\mathbf{A}) \geq \ldots \geq \mu_d(\mathbf{A})$ *y* $\mu_1(\mathbf{B}) \geq \ldots \geq \mu_d(\mathbf{B})$ *respectivamente. Si* $\mathbf{A} - \mathbf{B} \geq 0$ *entonces vale:*

a) $\mu_i(\mathbf{A}) \geq \mu_i(\mathbf{B})$

b) $tr(\mathbf{A}) \geq tr(\mathbf{B})$

c) $|\mathbf{A}| \geq |\mathbf{B}|$

d) $\|\mathbf{A}\| \geq \|\mathbf{B}\|$ *con* $\|\mathbf{A}\| = (tr(\mathbf{A}\mathbf{A}'))^{1/2}$*, la norma de Frobenius.*

Teorema 6.6. *Sean* $\mathbf{A}$*,* $\mathbf{B}$ *y* $\mathbf{A}-\mathbf{B}$ *matrices de* $I\!R^{d\times d}$ *semidefinidas positivas, con rango*$(\mathbf{B}) \leq r$*, y sea* $u_i(\cdot)$ *el i-ésimo autovalor más grande. Entonces,*

$$\mu_i(\mathbf{A}-\mathbf{B}) \geq \begin{cases} \mu_{r+i}(\mathbf{A}) & i = 1, 2, \ldots, n-r \\ \\ 0 & i = n-r+1, n-r+2, \ldots, n \end{cases} \tag{14}$$

La igualdad ocurre si $\mathbf{B} = \sum_{i=1}^{r} \mu_i(\mathbf{A})\mathbf{t}_i\mathbf{t}_i'$*, con* $\mathbf{t}_1, \ldots, \mathbf{t}_n$ *autovectores mutuamente ortogonales correspondientes a* $\mu_1(\mathbf{A}), \ldots, \mu_n(\mathbf{A})$*.*

6.2. Análisis Funcional

Nos basaremos en el clásico texto de Conway (1990). En todo lo que sigue $\mathbb{F}$ será el cuerpo de los números reales o complejos.

Definición 6.2. Diremos que un espacio normado $\mathcal{X}$ es un espacio de Banach si resulta ser completo con la métrica inducida por la norma.

Definición 6.3. Si $\mathcal{X}$ y $\mathcal{Y}$ son espacios de Banach y $A : \mathcal{X} \to \mathcal{Y}$ 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 6.4. Dado $\mathcal{X}$, un funcional lineal acotado $f : \mathcal{X} \to \mathbb{F}$ es una aplicación lineal continua. El conjunto de funcionales lineales acotados de $\mathcal{X}$ será notado como $\mathcal{X}$.

Teorema 6.7. *(Teorema de Hahn-Banach) Sea* $\mathcal{X}$ *un espacio normado y* $M \subset \mathcal{X}$ *un subespacio. Consideremos* $f : M \to \mathbb{F}$ *un funcional lineal y acotado, luego existe* $F \in \mathcal{X}^*$ *tal que* $F|M = f$ *y* $\|F\| = \|f\|$*.*

Corolario 6.1. *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 6.5. Un espacio de Hilbert es un espacio vectorial $\mathcal{H}$ sobre un cuerpo $\mathbb{F}$ junto con un producto interno $< \cdot, \cdot >$ de manera tal que la métrica inducida por el mismo es completa.

Definición 6.6. Si $\mathcal{H}$ es un espacio de Hilbert, luego diremos que $f \in \mathcal{H}$ y $g \in \mathcal{H}$ son ortogonales si $< f, g >= 0$.

Teorema 6.8. *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 $h - Ph$ es ortogonal a todo M. Así definido, $P_M : \mathcal{H} \to \mathcal{H}$ 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 $Ker(P_M) = M^{\perp}$ e $Im(P_M) = M$.*

Definición 6.7. Bajo las mismas condiciones que las del teorema anterior, P_M será la proyección ortogonal sobre el subespacio M.

Teorema 6.9. *(Teorema de representación de Riesz) Sea $f \in \mathcal{H}^*$, luego existe un único vector $h_0 \in \mathcal{H}$ tal que $f(h) =< h, h_0 >$. Más aún, $\|f\| = \|h_0\|$.*

Definición 6.8. 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 6.2. *(Desigualdad de Bessel) Si $\{e_n : n \in \mathbb{N}\}$ es un conjunto ortonormal y $h \in \mathcal{H}$, luego*

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

Teorema 6.10. *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 $h = \sum_{e \in \varepsilon} < h, e > h$.*
- *$\|h\|^2 = \sum_{e \in \varepsilon} | < h, e > |^2$ (identidad de Parseval).*

Proposición 6.3. *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 6.11. *Sean $\mathcal{H}_1,\ \mathcal{H}_2$ espacios de Hilbert. Sea $A : \mathcal{H}_1 \to \mathcal{H}_2$ lineal y continuo. Luego, existe un único $B : \mathcal{H}_2 \to \mathcal{H}_1$ tal que $\forall h_1 \in \mathcal{H}_1,\ h_2 \in \mathcal{H}_2,\ < Ah_1, h_2 >=< h_1, Bh_2 >$. Notaremos $A^* = B$ y diremos que es el adjunto de A.*

Definición 6.9. Si $A : \mathcal{H} \to \mathcal{H}$ es un operador lineal y continuo, luego diremos que es autoadjunto si $A^* = A$.

Proposición 6.4. *Si A es autoadjunto, luego vale que*

$$\|A\| = \sup\{| < Ah, h > | : \|h\| = 1\}.$$

Definición 6.10. A se dice una isometría si $\|Ah\| = \|h\|\ \forall h \in \mathcal{H}$. Si además A es sobreyectiva diremos que es unitario.

Proposición 6.5. *Si $A : \mathcal{H} \to \mathcal{H}$ lineal y continuo, luego son equivalentes*

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

Definición 6.11. Una transformación lineal $T : \mathcal{H} \to \mathcal{H}$ 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 6.12. Sea $A : \mathcal{H} \to \mathcal{H}$ 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 6.12. *Sea T un operador compacto y autoadjunto en $\mathcal{H}$. Entonces vale que T tiene una cantidad numerable o finita de autovalores distintos. Si $\{\lambda_1, \lambda_2, \cdots\}$ son los autovalores no nulos de T y P_n es la proyección ortogonal en el autoespacio $ker(T - \lambda_n)$, entonces $P_nP_m = P_mP_n$ 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 6.6. *Si $T : \mathcal{H} \to \mathcal{H}$ compacto y M un subespacio invariante para T, luego $T|M$ es compacto.*

Definición 6.13. Un operador $T : \mathcal{H} \to \mathcal{H}$ 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 ortnormal $\{e_i, i \in I\}$ de $\mathcal{H}$ cumpliendo que

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

Referencias

[1] Conway J. B. (1990) *A Course in Functional Analysis*, Springer Graduate Texts in Mathematics

[2] Fava N. y Zo F.(1996). *Medida e Integral de Lebesgue*, Red Olímpica.

[3] Flury, B.A. (1990). Principal Points, *Biometrika*, **77**, 33-41.

[4] Flury B.A. (1993) Estimation of Principal Points, *Appl. Statist.*, **42**, 139-151.

[5] Flury, B. y Tarpey, T. (1993). Representing a Large Collection of Curves: A Case for Principal Points. *The American Statistician*, **47**, 304-306.

[6] Flury, B. y Tarpey, T. (1998). Principal Points. In: *Encyclopedia of Statistical Science*, Update Volume 2, Kotz, S., Read, C., and Banks, D. Eds. Wiley: New York, pp. 545-548.

[7] Flury, B. y Tarpey, T. (1999). Self–Consistency (Update). In: *Encyclopedia of Statistical Science*, Update Volume 3, Kotz, S., Read, C., and Banks, D. Eds. Wiley: New York, 655-659.

[8] Frahm, G. (2004). *Generalized Elliptical Distributions: Theory and Applications*. Tesis doctoral de la Universidad de Colonia, Alemania.

[9] Kelker, D. (1970). Distribution Theory of Spherical Distributions and a Location–Scale Parameter Generalization. *Sankhya A*, **32**, 419-430.

[10] Muirhead R. J. (1982). *Aspects of Multivariate Statistical Theory*, John Wiley and Sons Canada.

[11] Seber G.A.F. (1984). *Multivariate Observations*, John Wiley and Sons.

[12] Tarpey, T. (1994). Two Principal Points of Symmetric, Strongly Unimodal Distributions. *Statistics and Probability Letters*, **20**, 253-257.

[13] Tarpey, T. (1995). Principal Points and Self–Consistent Points of Symmetric Multivariate Distributions. *Journal of Multivariate Analysis*, **53**, 39-51.

[14] Tarpey, T. (1997). Estimating Principal Points of Univariate Distributions. *Journal of Applied Statistics*, **24**, 483-496.

[15] Tarpey, T. (1998). Self–Consistent Patterns for Symmetric Multivariate Distributions. *The Journal of Classification*, **15**, 57-79.

[16] Tarpey, T. (1999). Self–Consistency and Principal Component Analysis. *Journal of the American Statistical Association*, **94**, 456-467.

[17] Tarpey, T. (1999). Self–Consistency Algorithms. *Journal of Computational and Graphical Statistics*, **8**, 889-905.

[18] Tarpey T. y Flury B. (1996). Self-Consistency: A Fundamental Concept in Statistics, *Statistical Science*, **11**, 229-243.

[19] Tarpey T. y Kinateder K. (2003). Clustering Functional Data, *Journal of Classification*, **20**, 93-114.

[20] Tarpey T., Li L. y Flury B. (1995) Principal Points and Self-Consistent Points of Elliptical Distributions. *The Annals of Statistics*, **23**, 103-112.