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

Tesis de Licenciatura

Algunas Medidas de Profundidad para Datos Multivariados y Funcionales

Lucas Fernandez Piana

Director: Marcela Svarc

26 de marzo de 2013

Índice

Agradecimientos

Agradezco a mi familia por su apoyo incondicional. En especial a mis padres que siempre me ayudaron de todas las formas que pudieron y crearon, con mucho esfuerzo, las condiciones para que pueda estar en esta carrera. A mi hermano Matías porque siempre está presente cuando me quejo, cuando le pido algo y me cubre con las cuestiones domésticas, aunque yo abuse un poco de la situación. A mi tía Alejandra, siempre incondicional. A mis abuelas. A mis abuelos porque desde chico me generaron interes en la ciencia.

Agradezco a los amigos por la paciencia que me tienen. A mis amigos de toda la vida Fernando, Matías y Esteban con quienes compartí momentos inolvidables durante este tiempo. A mis amigos de handball, con los que hemos crecido juntos compartiendo un deporte, y siguen llamando aunque siempre les diga “tengo que estudiar”.

A los amigos que hice en la facultad, hemos pasado momentos terribles adentro del Pabellón 1 y también momento muy lindos. Esos recuerdos son para toda la vida. Le agradezco a mis amigas de exactas, por hacerme parte de sus vidas, gracias chicas.

A mis compñeros de trabajo por cubrirme cuando necesita estudiar. Gracias a todos los que sin conocerme, sin ser mis amigos me dieron una mano cuando se las pedí.

A Marcela Svarc, una gran persona, que me guió durante este trabajo. Siempre encontraste tiempo para mi, aunque no lo tuvieras. Gracias.

1. Introducción

Una herramienta fundamental del análisis multivariado es la definición de medidas de profundidad. Las mismas son de gran utilidad en diversas aplicaciones ya sea en técnicas descriptivas, en la estadística no paramétrica, como también como herramienta a la hora de definir técnicas de aprendizaje supervisado y no supervisado entre otras. Cuando consideramos datos univariados la definición de medidas de profundidad está estrechamente ligada a la de estadísticos de orden y esta es natural ya que se base en el orden de los números reales. Sin embargo, al considerar datos en dimensiones mayores ya no se cuenta con una noción de orden total lo cual dificulta la definición de los mismos. A partir de la década del 70 diferentes propuestas de medidas de profundidad fueron introducidas para el caso de datos multivariados, entre ellas podemos mencionar, la *Profundidad del Semi-espacio* (Tukey, 1975), *Profundidad Simplicial* (Liu,1990), *Simplicial Volume Depth*, *Profundidad Majorante*, (Singh, 1991), entre las más difundidas.

En los últimos tiempos la importancia de los datos funcionales en diversos ámbitos de la ciencia cobró suma importancia debido a los avances tecnológicos que permiten tener mediciones certeras en tiempo real. Podemos encontrar aplicaciones de los mismos en las ciencias de la salud, las finanzas y la ingenerería entre otras. Con el objetivo de poder darle un tratamiento estadísco adecuado a estos datos, las principales técnica del análisis multivariado fueron, y continuan en la actualidad, siendo estudiadas en este nuevo contexto.

Principalmente debido a que la geometría de los datos funcionales es muy distinta a la de los datos multivariados y además que, en general, las profundidades multivariadas son muy costosas computacionalmente en dimensiones altas, surgió la necesidad de definir nociones profundidad para los mismos. En este caso se busca estudiar el grado de centralidad de una función concreta respecto de un proceso estocástico, o de un conjunto de funciones. En las dos últimas décadas varias propuestas fueron introducidas, entre ellas podemos destacar, *Profundidad Integrada* (Fraiman y Muniz, 2001), la *Profundidad basada en nucleos* (Cuevas et al, 2007), la *Profundidad de Tukey integrada* (Cuesta-Albertos y Nieto-Reyes, 2008), la profundidad *Profundidad Dual Intregrada* (Cuevas y Fraiman, 2009) y la *Profundidad por Bandas* (Lopez-Pintado y Romo, 2009).

En esta monografía estudiaremos de manera detallada la noción de *Profundidad Dual Intregrada* basada en proyecciones al azar, introducida por Cuevas y Fraiman (2009), ya que la misma esta definida para datos en espacios de Banach en general, tiene buenas propiedades teóricas y además es accesible computacionalmente.

Nuestro trabajo se estructura del siguiente modo. En la Sección 2 introducimos la definición de medidas de profundidad y comentamos sus las propiedades deseables. En las Sección 3 introducimos algunas medidas de profundidad para datos multivaridos. Por último, en la Sección 4, damos diversas definiciones de profundidad para datos funcionales y en particular estudiamos en detalle la *Profundidad Dual Integrada* (Cuevas y Fraiman, 2009), definida en espacios generales. Algunas definiciones y resultados complementarios se encuentran en el Apéndice.

2. Definición y propiedades de una medida de profundidad

El concepto de *Profundidad* surgió en primero para datos multivariados, buscando dar una idea de centralidad en los datos. De este modo un dato con un valor de profundidad *alto* debería estar en el centro de la nube de puntos y a medida que se aleja del centro el valor de la profundidad debería disminuir. En el caso de datos univariados prodíamos pensar como punto más profundo la mediana y mediante los cuantiles asignarle profundidad al resto de los puntos y de este modo el concepto de profundidad estaría estrechamente ligado al orden natural de los números reales. Cuando pasamos a un contexto multivariado, no se cuenta con un orden total, sino únicamente con órdenes parciales y esto hace que surjan diversas definiciones de profundidad, cuya intrerpretación ya no va a ser inmmediata.

A grandes rasgos, para una distribución P en $\mathbb{R}^d$, una función de profundidad es una función $D(x, P)$ que provee un orden del centro hacia afuera para puntos $x \in \mathbb{R}^d$ basado en P.
Interpretar "orden del centro hacia afuera"sugiere dos cosas

- Una noción relevante de centro.
- Los puntos cerca del centro deberían tener mayor profundidad que los mas lejanos.

De estas consideraciones se desprende que el "centro"debe constar de los puntos que maximizen globalmente la función de profundidad.
Naturalemente surge una pregunta: existe una definición general de función de profundidad definida en $\mathbb{R}^d$ definida con respecto a una distribución arbitraria que puede ser continua o discreta? Qué propiedades son deseables para caracterizarlas?
Dada una distribución P en $\mathbb{R}^d$, nuestra atención quedará confinada a que medidas consideradas sean acotadas y no negativas.

Es deseable que cualquier definición de profundidad satisfaga las siguientes propiedades

P.1. Invarianza Afín
La profundidad de un punto $x \in \mathbb{R}^d$ no debería depender del sistema de coordenadas, en particular de la escala del sistema de medición.

P.2. Maximalidad al centro
Para una distribución que tenga un único *centro* definido, por ejemplo, un punto de simetría, la profundidad se maximizaría en dicho punto.

P.3. Se anula en el infinito
La profundidad de un punto x debería aproximarse a cero si $\|x\|$ es demasiado grande.

P.4. Monotonía respecto del punto de máxima profundidad
A medida que $x \in \mathbb{R}^d$ se aleja del punto *más profundo* a lo largo de cualquier rayo fijo la profundidad de x debería decrecer.

Existen medidas de profundidad que no satisfacen **P.4.** y que en su lugar cumplen una condición de **cuasi-concavidad como funcion de x**, es decir que el conjunto $\{x : D(x;P) \geq c\}$ es convexo para cualquier número real c.

Por otra parte, es bueno que las medidas de profundidad sean "continuas", es decir, que cumplan

P.5. Continuidad como función de x
Es esperable pensar que si dos puntos se encuentran lo suficientemente cerca las profundidades no sean demasiado distintas

P.6. Continuidad como función de P.
Análogo a la propiedad anterior para distribuciones.

A continuación daremos una definición formal de función de profundidad, dada por Zuo y Serfling (2000), para ello necesitamos introducir la siguiente notación.

Notación: Llamamos $\mathbb{P}$ a la clase de distribuciones sobre los conjuntos Borelianos de $\mathbb{R}^d$ y llamamos P_ξ la distribución dada por el vector aleatorio ξ.

Definición 1. *Sea $D(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{P} \to \mathbb{R}$ una función acotada y no negativa que satisface:*

1. *$D(Ax+b, P_{AX+b}) = D(x, P_X)$ para todo vector aleatorio $X \in \mathbb{R}^d$ vector aleatorio, para todo matriz $A \in R^{dxd}$ no singular y para cualquier vector $b \in \mathbb{R}^d$.*

2. *$D(\theta, P) = \sup_{x \in R^d} D(x, P)$, para cualquier $P \in \mathbb{P}$ con centro θ.*

3. *$\lim_{\|x\| \to +\infty} D(x, P) = 0$, para todo $P \in \mathbb{P}$.*

4. *Para toda $P \in \mathbb{P}$ con punto más profundo θ $D(x, P) \leq D(\theta + \alpha(x - \theta), P)$, para cualquier $\alpha \in [0, 1]$.*

Si además quisieramos que se cumplieran **P.5.** y **P.6.**, deberíamos pedir que

5. Fijada $P \in \mathbb{P}$ $D(\cdot, P) : \mathbb{R}^d \to \mathbb{R}$ es continua.

6. Fijado $x \in \mathbb{R}^d$ $D(x, \cdot) : \mathbb{P} \to \mathbb{R}$ es continua.

Anteriormente hicimos referencia a la palabra *centro* para resaltar un punto de simetría. Es posible tomar varias nociones de simetría multivariada. Sin embargo, las más usadas en la literatura son las siguientes.

Definición 2. *Sea $X \in \mathbb{R}^d$ un vector aleatorio decimos que:*

- *Es* ***centralemente simétrico*** *(C-simétrico) respecto de θ si la distribución de $X - \theta$ es la misma que la de $\theta - X$.*
- *Es* ***Angularmente simétrico*** *(A-simétrico) respecto de θ si $(X - \theta)/\|X - \theta\|$ es centralmente simétrico respecto del origen.*
- *Es* ***Semiespacialmente simétrico*** *(H-simétrico) respecto de θ si $P(X \in H) \geq \frac{1}{2}$ para todo H semiespacio cerrado tal que $\theta \in H$.*

Observación 1. *Es claro que C-simétrico $\Rightarrow$ A-simétrico $\Rightarrow$ H-simétrico.*

2.1. Relación entre medidas de profundidad y funciones de outlyingness

El concepto de profundidad es inversamente equivalente con el de outlyingness. Dada una función de outlyingness $O(x, P)$ con rango $[0, \infty)$ tenemos una profundidad asociada definida por

$$D(x, P) = \frac{1}{1 + O(x, P)}.$$

Y si $O(x, P)$ es acotada definimos

$$D(x, P) = 1 - \frac{O(x, P)}{\sup O(\cdot, P)}.$$

Esta relación la desarrollaremos de manera más profunda en la siguiente sección cuando demos ejemplos concretos de medidas de profundidad.

2.2. Relación entre funciones quantiles y medidas de profundidad

En las distribuciones univariadas los cuantiles delimitan una fracción superior e inferior de la población, y cada punto $x \in \mathbb{R}$ tiene una interpretación como cuantil ya que se puede escribir como $F^{-1}(p)$, para algún $p \in [0, 1]$. En el caso multivariado, debido a la falta de un orden natural, lo natural es *orientar hacia un centro*, que generalmente se adoptará alguna definición de mediana multidimensional. Este centro suele servir como punto de partida para definir cuantiles multivariados orientados a la mediana.

En el caso univariado una función de cuantiles orientada por la mediana $Q_F(u)$, con $u = 2p - 1$ y mediana $M = Q_F(0)$, se define por

$$Q_F(u) = F^{-1}(\frac{1 + u}{2}), \text{ donde } -1 < u < 1,$$

donde el signo de u indica la dirección desde la mediana. Esta función tiene inversa $Q_F^{-1}(x) = 2F(x) - 1,\ x \in \mathbb{R}$. La magnitud de $|Q_F^{-1}(x)| = |2F(x) - 1|$, es una medida estandar para medir outlyingness relativa a la distribución F. Como x satisface $x = Q_F(Q_F^{-1}(x))$, podemos pensar que los cuantiles $Q_F(u)$ están indexados por un parámetro direccional de outlyingness, u con magnitud $|u|$. Si $|u|$ es cercano a cero corresponde a un extremos, mientras que si es cercano a uno corresponde al "centro".

Una función *quantil orientada por la mediana* $Q_F(u)$ está definida por u en la bola unidad $\mathbb{B}^{d-1}$ con $M_F = Q_F(0)$ sería una versión de la mediana d-dimensional. La función de quantil Q_F tiene una inversa $Q_F^{-1}(x),\ x \in \mathbb{R}^d$, que satisface $x = Q_F(Q_F^{-1}(x))$. Además su norma $\|Q_F^{-1}(x)\|$ en $[0, 1)$ mide outlyingness de x con respecto a F. Asociamos una profundidad, $D(x, F) = 1 - \|Q_F^{-1}(x)\|$.

Sea $D(\cdot, F)$ una función de profundidad que se maximiza en el punto M_F que interpretaremos como mediana. Por conveniencia, supongamos que F es continua con soporte en $\mathbb{R}^d$. Para $x \in \mathbb{R}^d$, sea p el correspondiente índice a la región central con x en su frontera, y sea $u = pv$, para v un vector unitario en la dirección de x hacia M_F. Ajustando $Q_F(u) = x$, con $Q_F(0) = M_F$, los puntos $x \in \mathbb{R}^d$ generan una función quantil $Q_F(u),\ u \in \mathbb{B}^{d-1}$.

3. Medidas de Profundidad para Datos Multivariados

Las medidas de profundidad para datos multivariados surgieron en con el objetivo de extender la noción de los estadísticos de orden. Los mismos son sumamente útiles al emplear técnicas no paramétricas, sobre todo en casos donde la distribución de los datos no es gaussiana ni elíptica. Además, es importante contar con técnicas de "ordenación"para los puntos que tengan en cuenta la geometría de los datos. Por ejemplo si se busca definir la observación central, considerar la media o la mediana coordenada a coordenanda ignora la geometría de los datos. Por otra parte, cuando consideramos datos reales el orden natural de los mismos da un una noción de ranking entre las observaciones, al pasar a dimensiones mayores esto no se puede hacer de manera natural y por lo tanto, es conveniente poder definir un centro (que se puede hacer de muchas formas distintas) y a partir del mismo orientar las observaciones de adentro hacia afuera.

En esta sección estudiaremos algunos ejemplos de funciones de profundidad para datos multivariados y trataremos de dar una clasificación de las medidas de profundidad con respecto a las propiedad (**P.1.-P.4.**) En primer lugar definiremos las dos medidas de profundidad para datos multivariados más clásicas en la literatura.

3.1. Profundidad del Semi-espacio

La semilla de esta teória fue plantada por Tukey (1975) con la *Profundidad del Semi-espacio* (Halfspace Depth, HD).
La profundidad del semi-espacio de un punto $x \in \mathbb{R}^d$ con respecto a una medida de probabilidad P en $\mathbb{R}^d$ está definida por la mínima probabilidad de que un semi-espacio cerrado que contenga a x, esto es,

$$HD(x, P) = inf\{P(H) : H\ semi-espacio\ cerrado,\ x \in H\},\ x \in \mathbb{R}^d.$$

La versión muestral es la siguiente, sea $x \in \mathbb{R}^d$, la profundidad de x respecto de una muestra $X_1, X_2, \ldots, X_n$ también $\mathbb{R}^d$ es la menor fracción de puntos de la muestra que hay en un semiespacio cerrado que contenga a x, es decir,

$$HD_n(x) = \frac{\text{mín}_{u \in \mathbb{R}^d} \sharp\{i : \langle X_i, u\rangle\ \geq \langle x, u\rangle\}}{n},$$

donde $\langle \cdot, \cdot \rangle$ denota el producto interno.

Si consideramos P una medida de probabilidad en $\mathbb{R}$ y X una variable aletoria con función de distribución acumulada F, entonces

$$HD(P, x) = \text{mín}\{P(X \leq x), (1 - P(X < x))\} = \text{mín}\{F(x), (1 - F(x^-))\}.$$

El principal inconveniente que presenta esta medida de profundidad es que es computacionalmente muy costas en dimensiones moderadas y altas.

Teorema 1. *La profundidad del semi-espacio cumple las propiedades* ***P.1.-P.4.***

Demostración. Es claro que $H(x, P)$ es acotada, no negativa.

P.1. Sea, $X \in \mathbb{R}^d$ vector aleatorio, $A \in R^{dxd}$ una matriz no singular y $b \in \mathbb{R}^d$. Sea,

$$H : x \in H \Rightarrow Ax = b \in AH + b$$

y

$$P_{AX+b}(AH + b) = P(H).$$

Entonces $HD(Ax + b, P_{AX+b}) = HD(x, P)$.

P.2. Supongamos que P es H-simétrica respecto de un único punto $\theta \in \mathbb{R}^d$. Tenemos que, $P(H_\theta) \geq \frac{1}{2}$ para todo $H_\theta$tal que $\theta \in H_\theta$. Entonces, $HD(\theta, P) \geq \frac{1}{2}$.

Supongamos que existe $x_0 \in \mathbb{R}^d$ con $HD(x_0, P) \geq \frac{1}{2}$ entonces $P(H) \geq \frac{1}{2}$ para todo semi-espacio tal que $x_0 \in H$.

Por lo tanto, P es H-simétrica respecto de x_0. Esto es una contradicción.

Luego,

$$HD(\theta, P) = \sup_{x \in \mathbb{R}^d} HD(x, P).$$

P.3. Sea $0 < \alpha < 1$, debemos comparar $HD(x, P)\ y\ HD((1 - \alpha)\theta + \alpha x, P)$.

Por el Teorema de Separación del hiperplano sabemos que ,

$$\exists H_x\ tq\ H_x \subseteq H_{(1-\alpha)\theta+\alpha x}.$$

Entonces,

$$HD(x, P) \leq HD((1 - \alpha)\theta + \alpha x, P).$$

P.4. Sabemos que $P(\|X\| \geq \|x\|) \xrightarrow[\|x\| \to +\infty]{} 0$. Además, $H_x \subseteq \{\|X\| \geq \|x\|\}$.

Luego, $HD(x, P) \xrightarrow[\|x\| \to +\infty]{} 0$

$\square$

3.2. Profundidad del Simplicial

Otra noción importante fue planteada por Liu (1990). La *Profundidad Simplicial (SD)* de un punto $x \in \mathbb{R}^d$ con respecto a una medida de probabilidad P in $\mathbb{R}^d$ es la probabilidad de que x pertenezca a un simplex aleatorio en $\mathbb{R}^d$, i.e,

$$SD(x,P) = P(x \in S[X_1, \ldots, X_{d+1}]), \; x \in \mathbb{R}^d.$$

donde $X_1, \ldots, X_{d+1}$ es una muestra de vectores aleatorios con distribución P y $S[X_1, \ldots, X_{d+1}]$ es el simplex d-dimensional con vértices $X_1, \ldots, X_{d+1}$.

A continuación daremos la definición muestral de la profundidad simplicial. Sea $X_1, \ldots, X_n$ una muestra aleatoria en $\mathbb{R}^d$, la profundidad simplicial de un punto $x \in \mathbb{R}^d$ es

$$SD_n(x) = \binom{n}{d+1} \sum_{1 \leq i_1 < i_2 < \ldots < i_{d+1} \leq n} \mathcal{I}\{x \in S[X_{i_1}, \ldots, X_{i_{d+1}}]\},$$

donde $\mathcal{I}$ es la función indicadora.

En el caso univariado, la función de profundidad simplicial es $SD(x) = F(x)(1 - F(x^-))$, donde F es la función de distribución acumulada. Zuo y Serfling (2000) prueban que cumple las propiedades **P.1.-P.4.** para distribuciones continuas A-simétricas.

Sin embargo, estas propiedades no se cumplen en el caso general. Veamos el siguiente contraejemplo, consideramos una distribución discreta unidimensional donde la probabilidad puntual está dada por,

$$P(X = 0) = P(X = \pm 1) = P(X = \pm 2) = \frac{1}{5}.$$

Es claro que X es C-simétrico respecto de 0.

$$SD(\frac{1}{2}, P) = P(\frac{1}{2} \in S[X_1, X_2]) = \frac{12}{25}, \text{ pues}$$
$$\text{Si } X_1 = 0, -1 \text{ ó } 2 \rightarrow X_2 = 1 \text{ ó } 2.$$
$$SD(1, P) = P(1 \in S[X_1, X_2]) = \frac{15}{25}, \text{ pues}$$
$$\text{Si} X_1 = 1 \rightarrow X_2 = 0, 2, -1 \text{ ó } -2,$$
$$\text{Si} X_1 = -1 \text{ ó } -2 \rightarrow X_2 = 1 \text{ ó } 2.$$

Esto contradice **P.3.**

3.3. Estructura general para funciones de profundidad

Zou y Serfling (2000), proponen cuatro estructuras generales para la construcción de profundidades.

3.3.1. Funciones de profundidad tipo A

Sea $h(x, x_1, \ldots, x_r)$ una función acotada y no negativa que mide en algún sentido la cercania de x a los puntos $x_1, \ldots, x_r$. Una *función de profundidad tipo A* nos queda definida por la cercania esperada de x a una muestra aleatoria de tamano r

$$D(x, P) = Eh(x, X_1, \ldots, X_r),$$

donde $X_1, \ldots, X_r$ es una muestra de vectores aleatorios con distribución P.

Ejemplo 1. *Tomando $r = d + 1$ y $h(x, x_1, \ldots, x_{d+1}) = \mathcal{I}_{x \in S[x_1, \ldots, x_{d+1}]}$ obtenemos la profundidad simplicial.*

Ejemplo 2. *Dados $x_1, \ldots, x_d \in \mathbb{R}^d$ determinan un único hiperplano que los contiene, al cual le corresponden dos semi-espacios cerrados que lo contienen como frontera. Notamos $H^P_{x_1, \ldots, x_d}$ al que tenga probabilidad mayor o igual que $\frac{1}{2}$ bajo la distribución $P \in \mathbb{R}^d$. Definimos la Profundidad Mayorante como*

$$MJD(x, P) = P(x \in H^P_{x_1, \ldots, x_d}), \ x \in \mathbb{R}^d.$$

Es claro que es una profundidad tipo A tomando $r = d$ y $h(x, x_1, \ldots, x_d) = \mathcal{I}_{\{x \in H^P_{x_1, \ldots, x_d}\}}$.

3.3.2. Funciones de profundidad tipo B

Sea $h(x, x_1, \ldots, x_r)$ una función no acotada y no negativa que mide en algún sentido la distancia de x a los puntos $x_1, \ldots, x_r$. Una *función de profundidad tipo B* nos queda definida por

$$D(x, P) = (1 + Eh(x, X_1, \ldots, X_r))^{-1},$$

para $X_1, \ldots, X_r$ una muestra de vectores aleatorios de distribución P.

Ejemplo 3. *Tomando $h(x, x_1, \ldots, x_r) = \Delta^\alpha(S[x, x_1, \ldots, x_d])$ donde $\Delta(S[x, x_1, \ldots, x_d]$ es el volumen del simplex d-dimensional $S[x, x_1, \ldots, x_d]$ y $\alpha > 0$. h es una medida de dispersión de la nube de puntos $\{x, x_1, \ldots, x_d\}$ y de acuerdo a lo anterior*

$$(1 + E[\Delta^\alpha(S[x, x_1, \ldots, x_d])])^{-1},$$

define una función de profundidad tipo B. Sin embargo, no cumple con la propiedad de la invarianza afín porque para una matriz A no-singunlar y un vector b, ya que la ecuación

$$\Delta^\alpha(S[Ax + b, Ax_1 + b, \ldots, Ax_d + b]) = |\det(A)| \Delta^\alpha(S[x, x_1, \ldots, x_d]),$$

no necesariamente se verifica. Se cumple únicamente si $\det(A) = 1$.

Este problema puede ser rectificado y definimos la Profundidad Simplicial de Volumen

$$SVD^{\alpha} = (1 + E[(\frac{\Delta(S[x, X_1, \ldots, X_d])}{\sqrt{\det(\Sigma)}})^{\alpha}])^{-1},$$

donde Σ *es la matriz de covarianzas de* P*. Esta versión cumple* ***(P.1.)***

3.3.3. Funciones de profundidad tipo C

Sea $O(x, P)$ una medida de outlyingness para el punto $x \in \mathbb{R}^d$ con respecto al centro o al punto más profundo de la distribucón P. Nos queda definida una *función de profundidad tipo C* por

$$D(x, P) = (1 + O(x, P))^{-1}.$$

Ejemplo 4. *Sea* $x \in \mathbb{R}^d$*. Definimos la Profundidad de Proyección* $PD(x, P)$*, mediante la siguiente función de outlyingness*

$$O(x, P) = \sup_{u \in \mathbb{R}^d, \|u\|=1} \frac{u'x - Med(u'X)}{MAD(u'X)},$$

donde X *tiene distribución* P*,* Med *es la mediana univariada,* $MAD(Y) = Med(|Y - Med(Y)|)$ *y* $\|\cdot\|$ *es la norma Euclidea.*

3.3.4. Funciones de profundidad tipo D

Sea $\mathcal{C}$ una clase de cerrados de $\mathbb{R}^d$ y P una medida de probabilidad en $\mathbb{R}^d$. Definimos la correspondiente *función de profundidad tipo D* como

$$D(x; P, \mathcal{C}) = \inf_{C} \{P(C) : x \in C \in \mathcal{C}\}.$$

Hay una clase distinguida que releva nuestra atención y cumple dos propiedades

- Si $C \in \mathcal{C} \implies \overline{C}^c \in \mathcal{C}$.
- Para $C \in \mathcal{C}$ *y* $x \in C^o, \exists C_1$ tal que $x \in \partial C_1,\ C_1 \subseteq C^o$.

Observar que la clase de semi-espacios cerrados $\mathfrak{H}$ en $\mathbb{R}^d$ satisface ambas propiedades y por lo visto antes, la profundidad del semi-espacio es un típico ejemplo de este tipo de medidas de profundidad.

4. Medidas de Profudidad para Datos Funcionales

En los últimos años ha aumentado el interés por el estudio de los datos funcionales. Muchas técnicas del análisis multivariado pudieron ser extendidas de manera directa. En el caso particular de medidas de profundidad uevas propuesta fueron necesarias ya que el cómputo de las medidas de profundidad existente ya era inviable para espacios de dimensiones moderadas y altas, por lo tanto, impracticables en el casos de dimensión infinita.

Cada dato funcional se representa por una función $x_i(t)$, donde $t \in T$ un intervalo cerrado de números reales.

A continuación definiremos algunas medidas de profundidad para datos funcionales.

4.1. Profundidad de Fraiman y Muniz

Está fue la primer medida de profundidad propuesta para datos funcionales y fue introducida por Fraiman y Muniz (2001). Para cada $t \in [0, 1]$, sea $F_{n,t}$ la distribución empírica de la muestra $x_1(t), \ldots, x_n(t)$ y sea $Z_i(t)$ la profundidad univariada del dato $x_i(t)$ en esta muestra, dada por $D_i(t) = 1 - |\frac{1}{2} - F_{n,t}(x_i(t))|$. Para cada $1 \leq i \leq n$, sea

$$I_i = \int_0^1 D_i(t)dt,$$

luego construirmos un ranking para las observaciones $x_i(t)$ de acuerdo a los valores de I_i.

La adaptación al caso multivariado es directa, reemplazando la integral por la suma finita apropiada.

4.2. Profundidad basada en nucleos

De acuerdo con esta noción, la h-profundidad poblacional de un dato z está dado por

$$f_h(z) = E[K_h(\|z - X\|)], \tag{1}$$

donde X es una elemento aleatorio que describe la población, $\|.\|$ es una norma en el espacio respectivo (por ejemplo la norma L_2 en el caso funcional) y K_h es un núcleo re-escalado, $K_h(t) = \frac{1}{h}K(\frac{t}{h})$, siendo K una función núcleo. Definimos la h-moda de X como el valor más profundo que se obtiene máximizando f_h.

Dada una muestra aleatoria $X_1, \ldots, X_n$ de X, la versión muestral de la h-profundidad se define de modo natural reemplazando en (1) por la versión empírica, i.e.

$$\hat{f}_h(z) = \frac{1}{n}\sum_{1}^{n} K_h(\|z - X_i\|).$$

4.3. Profundidad de Tukey Aleatorizada

Vimos en la Sección 2 la Profundidad del Semiespacio, $HD(x,P)$, que también es conocida como Profunidad de Tukey. Un grave problema que presenta esta profundidad es que es altamente costosa computacionalmente, aun en dimensiones moderadas, ya que requiere calcular todas las posibles proyecciones univariadas.

Cuesta-Albertos y Nieto-Reyes (2008) definen la Profundidad de Tukey Aleatorizada, con el objetivo de superar este problema. A su vez esta definición toma gran importancia, ya que se puede extender a espacios de Hilbert separables.

Una definición equivalente de $HD(x,P)$ es la siguiente: dado $\upsilon \in \mathbb{R}^p$, sea Π_υ la proyección de $\mathbb{R}^p$ en el subespacio unidimensional generado por υ. Entonces, $P \circ \Pi_\upsilon^{-1}$ es la distribución marginal de P en ese subespacio y

$$HD(x,P) = inf\{D_1(\Pi_\upsilon(x), P \circ \Pi_\upsilon^{-1}) : \ x \in \mathbb{R}^p\},$$

donde $D_1(x,P) = \text{mín}\{F(x), 1 - F(x^-)\}$.

En primer término, Cuesta-Albertos y Nieto-Reyes (2008) definen la Profundidad de Tukey Aleatorizada para datos multivariados.

Definición 3. *Sea $P \in \mathbb{P}$. Sea $\upsilon \in \mathbb{P}$ absolutamente continua y sean $\upsilon_1, \ldots, \upsilon_k$ vectores aleatorios i.i.d. con distribución υ. La Profundidad de Tukey Aleatorizada de $x \in \mathbb{R}^p$ con respecto a P basada en k vectores aleatorios elegidos con υ es*

$$D_{T,k,\upsilon}(x,P) = min\{D_1(\Pi_{\upsilon_i}(x), P \circ \Pi_{\upsilon_i}^{-1}) : 1 \leq i \leq k\}, \ x \in \mathbb{R}^p.$$

Según prueban Cuesta-Albertos y Nieto-Reyes (2008) para datos multivariados esta profundidad cumple **P.1.-P.3.** y **P.4.** con convergencia en probabilidad.

Una posibilidad interesante que presenta la Profundidad de Tukey Aleatorizada es que puede extenderse a espacios de Hilbert separables. Sin embargo, en este caso se cumplen las propiedades **P.1.-P.3.**, pero no se cumple la propiedad **P.4.**. Veamoslo en el siguiente ejemplo.

Consideramos $\mathbb{H} = L_2[0,1]$ con el producto interno usual. Sea $\{\delta_n\} \subset \mathbb{R}^+$ con $\text{lím}_n \delta_n = 0$. Sea $\{x_n\} \in \mathbb{H}, \ x_n(t) = \frac{1}{\delta_n}\mathcal{I}_{[0,\delta_n)}(t)$. Tenemos, $\|x_n\| = \delta_n^{-\frac{1}{2}} \Rightarrow lim_n\|x_n\| = +\infty$. Tomemos υ igual a la distribución del movimiento Browniano y sea $P = \upsilon$. Si X es un elemento aleatorio con distribución υ, entonces $< x_n, X >$ converge a cero en probabilidad, pues

$$E| < x_n, X > | = E|\int_0^{\delta_n} X(t)\delta_n^{-1}dt| \leq \int_0^{\delta_n} E|X(t)|\delta_n^{-1}\, dt = \int_0^{\delta_n} (\frac{2t}{\Pi})^{\frac{1}{2}}\delta_n^{-1}\, dt \leq (\frac{2\delta_n}{\Pi})\frac{1}{2},$$

donde la última igualdad es valida porque la distribución de $X(t)$ es $N(0,t)$.

Luego, si $\upsilon_1, \ldots, \upsilon_k \in \mathbb{H}$ son elegidos aleatoriamente con distribución υ tenemos

$$lim_n D_1(< \upsilon_i, x_n >, P \circ \Pi_{\upsilon_i}^{-1}) = D_1(0, P \circ \Pi_{\upsilon_i}^{-1}) = max_x D_1(x, P \circ \Pi_{\upsilon_i}^{-1}) = 2^{-1},$$

porque $P \circ \Pi_{v_i}^{-1}$ porque la distribución es una normal centrada.

Cuesta-Albertos y Nieto-Reyes (2008), prueban que la Profundidad de Tukey Randomizada es no negativa, acotada y que satisface **P.1.-P.3.** Además obtienen resultados de consistencia fuerte respecto de P, condicional a haber elegido $v_1, \ldots, v_k$ direcciones al azar.

4.4. Profundidad Dual Integrada (PDI)

Cuevas y Fraiman (2009), presentan una definición de medidas de profundidad para datos en un espacio general de Banach separable, basada en proyecciones continuas unidimensionales.

Sea Ω un espacio de probabilidad y $\mathbb{E}$ un espacio de Banach separable con dual $\mathbb{E}'$ separable. Sea $X : \Omega \longrightarrow \mathbb{E}$ un elemento aleatorio en $\mathbb{E}$ con distribución P .

Dada una medida de probabilidad P^1 en $\mathbb{R}$ y $u \in \mathbb{R}$ sea $D(P^1, u)$ el valor que indica la profundidad de u con respecto a P^1, por ejemplo tomando la Profundidad Simplicial tenemos que,

$$SD(P^1, u) = F^1(u)(1 - F^1(u^-), \tag{2}$$

donde F^1 es la función de distribución acumulada de P^1. Esta es la profundidad unidimensional que consideraremos de aquí en adelante.

También se puede usar la Profundidad de Tukey

$$HD_T(P^1, u) = min(F^1(u), 1 - F^1(u^-)).$$

Definimos la *Profundidad Dual Integrada* (PDI) como

$$D(P, x) = \int D(P_f, f(x))dQ(f), \tag{3}$$

donde $f \in \mathbb{E}'$, $x \in \mathbb{E}$ y $P_f = P_{f(X)}$ es la distribución de una variable aleatoria $f(X)$ y Q es una medida de probabilidad en E'.

Observación 2. *Si $\mathbb{E}$ es un espacio infinito-dimensional Q se puede elegir como una medida Gaussiana no degenerada. En caso de que la dimensión sea finita tomamos Q como la Medida de Haar.*

Observación 3. *En el caso que $\mathbb{E}$ es un espacio de Hilbert $\mathbb{H}$, tenemos una isometría biyectiva $\Psi : \mathbb{H} \longrightarrow \mathbb{H}'$ dada por $\Psi(h) =< h, \cdot >$. Si $\varphi \in \mathbb{H}'$ y X un elemento aleatorio de $\mathbb{H}$, entonces $\varphi(X) =< h, X >$ para un único $h \in \mathbb{H}$ y $P_{\varphi(X)} = P_h$ es la distribución de $< h, X >$, que es la proyección del elemento aleatorio X en la dirección h. Luego, la ecuación (3) queda*

$$D(P, x) = \int D(P_h, < h, x >)dQ(h), \tag{4}$$

donde Q es una medida de probabilidad en $\mathbb{H}$.

Definición 4. *La* mediana *asociada con (3) será el punto más profundo*

$$x_0 = argmáx_{x\in\mathbb{E}} D(P, x).$$

Definición 5. *Sean $X_1, \ldots, X_n$ elementos aleatorios i.i.d en $\mathbb{E}$ con distribución P, sea P_n la districión empírica y sea $P_{n,f}$ la distribución empírica de $f(X_1), \ldots, f(X_n)$. Definimos la PDI Empírica como*

$$D(P_n, x) = \int D(P_{n,f}, f(x))dQ(f), \tag{5}$$

y la mediada muestral *como*

$$\hat{x}_{o,n} = argmáx_{x\in\mathbb{E}} D(P_n, x).$$

En la práctica para aproximar (5) consideramos $f_1, \ldots, f_n \in \mathbb{E}'$ con distribución Q independiente de $X_1, \ldots, X_n$ y calculamos

$$\tilde{D}(P_n, x) = \frac{1}{N}\sum_{j=1}^{N} D(P_{n,f_j}, f_j(x)). \tag{6}$$

Con esto podemos estimar (6) mirando el punto más profundo, $\tilde{x}_0$, definido por

$$máx_{1\leq j\leq n}\{\tilde{D}(P_n, X_j)\}.$$

4.4.1. Propiedades

Es bastante regular encontrar "objetos"definidos, en un primer momento, para espacios de dimensión finita que pierden algunas de sus características cuando se definen en espacios más generales. En nuestro caso, que consideramos espacios de Banach hay muchos ejemplos que muestran esto, como la bola unitaria que deja de ser compacta, entre otros. Las profundidades no deberían ser una excepción. Sería inocente de nuestra parte pensar que las mismas se conserven dado que el concepto está intrinsecamente relacionado con la distribución y la geometría de los datos. En lo siguiente, analizaremos las propiedades que mencionamos en la introducción para la profundidad PDI.

P.1. Invarianza Afín

Consideramos el caso finito dimensional. Si A es una transformación lineal no singular de $\mathbb{R}^d$ y P_{AX} es la distribución de AX, en el caso d-dimensional, entonces la propiedad de invarianza, i.e. $D(P_{AX}, Ax) = D(P_X, x)$, no se cumple la para cualquier medida y transformacion lineal biyectiva. Por lo tanto, buscamos una medida que sea invariante sobre algún subgrupo de transformaciones lineales no singulares de $\mathbb{R}^d$.

$$
\begin{aligned}
D(P_{AX}, Ax) &= \int D(P_{h(AX)}, < h, Ax >) dQ(h) \\
&= \int P(< h, AX > \leq < h, Ax >) \left(1 - P(< h, AX > \leq < h, Ax >)\right) dQ(h).
\end{aligned}
$$

Si Q es la medida de Haar, $\mathbb{D}^d = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$ y $\mathcal{I}_{\{h \in \mathbb{D}^d\}}$ la función indicadora sobre $\mathbb{D}^d$

$$
D(P_{AX}, Ax) =
$$

$$
= \int P(< hA^t, X > \leq < hA^t, x >) \big(1 - P(< hA^t, X > \leq < hA^t, x >)\big) \mathcal{I}_{\{h \in \mathbb{D}^d\}} dh.
$$

Sea $w = h^t A$, entonces,

$$
D(P_{AX}, Ax) =
$$

$$
= \int P(< w, X > \leq < w, x >) \left(1 - P(< w, X > \leq < w, x >)\right) |\det(A^{-1})| I_{\{w \in A^t(\mathbb{D}^d)\}} dw.
$$

Se puede observar que si $\det(A) = \pm 1$ y deja fijo a $\mathbb{D}^d$ se cumple $D(P_{AX}, Ax) = D(P_X, x)$

P.2. Maximalidad en el centro

Supongamos que P es centralmente simétrica respecto de θ. En particular, tenemos que para todo $f \in \mathbb{E}'$,

$$
P(f(X - \theta) \leq 0) = P(f(\theta - X) \leq 0) \Rightarrow P\left(f(X) \leq f(\theta)\right) = P\left(f(X) \geq f(\theta)\right) = \frac{1}{2}.
$$

Luego,

$$
D(P_f, f(\theta)) = P\left(f(X) \leq f(\theta)\right) P\left(f(X) \geq f(\theta)\right) = \frac{1}{4} \text{ para todo } f \in \mathbb{E}',
$$

tomando su valor máximo y hace que $D(P, \theta)$ se maximize.

P.3. Se anula en infinito

Teorema 2. *Supongamos que*

$$sup_{\|x\|=1} Q\{f : f(x) \leq \epsilon\} = O(\epsilon),$$

donde $O(\epsilon)$ *es una función tal que*

$$\lim_{\|x\| \to +\infty} O(\epsilon) = 0.$$

Entonces,

$$\lim_{\|x\| \to +\infty} D(P, x) = 0.$$

Demostración. Sea $\epsilon > 0$ y $M > 0$. Sea x tal que $\|x\| \geq M$.

$$\begin{aligned}
D(P,x) &= \int D(P_f, f(x)) dQ(f) \\
&= \int P(f(X) \leq f(x))(1 - P(f(X) \leq f(x))) dQ(f) \\
&= \int P(f(X) \leq f(x)) P(f(X) \geq f(x)) dQ(f) \\
&\leq \int P(f(X) \geq f(x)) dQ(f) \\
&= \int 1_{\{f : f(\frac{x}{\|x\|}) \leq \epsilon\}} P(f(X) \geq f(x)) dQ(f) + \int 1_{\{f : f(\frac{x}{\|x\|}) > \epsilon\}} P(f(X) \geq f(x)) dQ(f) \\
&\leq \int 1_{\{f : f(\frac{x}{\|x\|}) \leq \epsilon\}} P(f(X) \geq f(x)) dQ(f) + \int 1_{\{f : f(\frac{x}{\|x\|}) > \epsilon\}} P(f(X) \geq f(\frac{x}{\|x\|} \|x\|)) dQ(f) \\
&\leq \int 1_{\{f : f(\frac{x}{\|x\|}) \leq \epsilon\}} dQ(f) + \int 1_{\{f : f(\frac{x}{\|x\|}) > \epsilon\}} P(f(X) \geq \|x\| f(\frac{x}{\|x\|})) dQ(f) \\
&\leq O(\epsilon) + \int 1_{\{f : f(\frac{x}{\|x\|}) > \epsilon\}} P(f(X) \geq \|x\| f(\frac{x}{\|x\|})) dQ(f) \\
&\leq O(\epsilon) + \int P(f(X) \geq M\epsilon) dQ(f).
\end{aligned}$$

Usando el Teorema de la Convergencia Mayorada (Munoz y Blanco, 2002) dado que si llamamos

$$H_M(f) = P(f(X) \geq M\epsilon).$$

Por lo tanto, $H_M(f) \leq 1 \forall f \in \mathbb{E}'$.

Entonces,

$$\lim_{M\to+\infty}\int P(f(X)\geq M\epsilon)dQ(f)=\int\lim_{M\to+\infty}P(f(X)\geq M\epsilon)=0.$$

Luego, como ϵ es arbitrario $\lim_{\|x\|\to+\infty}\int D(P_f,f(x))dQ(f)\leq O(\epsilon)\xrightarrow[\epsilon\to 0]{}0.$

$\square$

P.4. **Monotonía respecto del punto más profundo**

Sea X un elemento aleatorio en $\mathbb{E}$ con función de probabilidad P con centro θ y $x\in\mathbb{E}$.

Sea $\alpha\in[0,1]$ y $f\in\mathbb{E}'$,

$$f(\theta-(1-\alpha)(x-\theta))=f(\theta)-(1-\alpha)(f(x)-f(\theta))$$

y P_f tiene centro $f(\theta)$.

Luego,

$$D(P_f,f(x))\leq D(P_f,f(\theta)-(1-\alpha)(f(x)-f(\theta))=D(P_f,f(\theta-(1-\alpha)(x-\theta))).$$

La propiedad se obtiene de la monotonía de la integral.

P.5. **Continuidad como función de x**

Si P_f tiene una distribución continua, entonces por construcción tenemos que $D(P_f,f(x))=P(f(X)\leq f(x))P(f(X)\geq f(x))$ es continua como función de x.

Es decir, $(x_n)_{n\in N}$ es una sucesión en $\mathbb{E}$ que converge a x.

Por lo tanto, $f(x_n)\to f(x)$. Luego, $D(P_f,f(x_n))\to D(P_f,f(x))$ para todo $f\in\mathbb{E}'$.

Luego, por consecuencia del Teorema de Convergencia Mayorada

$$\int D(P_f,f(x_n))dQ(f)\to\int D(P_f,f(x))dQ(f).$$

P.6. **Continuidad como función de P**

También se desprende del Teorema de Convergencia Mayorada y de la continuidad de $D(P_f,f(x))$ como funcional de P_f. Pues, si P^k tiende a P en sentido débil cuando $k\to\infty$, entonces $P_f^k\xrightarrow[k\to\infty]{}P_f$ para todo f.

4.4.2. Propiedades Asintóticas

Por otra parte, Cuevas y Fraiman (2008) estudiaron el comportamiento asintótico de las medidas de profundidad empíricas. A continuación presentamos dos teoremas que muestran, respectivamente, la convergencia uniforme de la profundidad empírica y la convergencia en probabilidad del punto más profundo muestral.

Teorema 3. *Sea* $\{X_j\}_{j\geq 1}$ *una sucesión de elementos aleatorios i.i.d en* $\mathbb{E}$*. Entonces,*

$$E(\sup_{P\in\mathbb{P}}\sup_{x\in\mathbb{E}}|D(P_n,x)-D(P,x)|)\xrightarrow[n\to+\infty]{}0,$$

donde $\mathbb{P}$ *es el espacio de funciones de distribución en* $\mathbb{E}$*.*

En particular, $\sup_{x\in\mathbb{E}}|D(P_n,x)-D(P,x)|\to 0$ *en probabilidad uniformemente en* $\mathbb{P}$*.*

Demostración. Recordemos que la medida de profundidad univariada de X está dada por

$$D(P_f,f(X))=P(f(X)\leq f(x))(1-P(f(X)\leq f(x)),$$

donde P_f es la distribución de la proyección $f(X)$ y F es la correspondiente distribución acumulada, es decir,

$$D(P_f,f(X))=F(f(x))(1-F(f(x)^-)).$$

La aproximamos empíricamente por

$$D(P_{n,f},f(x))=F_n(f(x))(1-F_n(f(x)^-)),$$

donde F_n es la distribución empírica asociada de $f(X_1),\ldots,f(X_n)$. Utilizando el resultado obtenido por Massart (1990) para la desigualdad de Dvoretzky-Kiefer-Wolfowitz (ver Apéndice) tenemos que,

$$\begin{aligned}P(\sup_{x\in E}|F_n(f(x))-F(f(x))|>\epsilon)=P(\sqrt{n}\sup_{x\in E}|F_n(f(x))-F(f(x))|>\sqrt{n}\epsilon)\\ \leq 2\exp\left(-2(\sqrt{n}\epsilon)^2\right)\\ =2\exp\left(-2n\epsilon^2\right),\end{aligned}$$

si $\exp(-2n\epsilon^2)\leq 1/2$.

Esta cota es uniforme en F.

Llamamos,

$$G=P(f(X)\leq f(x))\text{ y }G^-=P(f(X)<f(x)).$$

Y sus respectivas versiones empíricas,

$$G_n = F_n(f(x)) \text{ y } G_n^- = F_n(F(x)^-).$$

Entonces,

$$\begin{aligned}
|D(P_{n,f}, f(x)) - D(P_f, f(x))| \quad & & (7)\\
&= |G_n(1-G_n^-) - (G(1-G^-)| \\
&= |G_n(1-G_n^-) - (G(1-G^-) + G^-G_n - G^-G_n| \\
&= |G_n - G + G^-(G-G_n) + G_n(G^- - G_n^-)| \\
&\leq |G_n - G| + G^-|G-G_n| + G_n|G^- - G_n^-| \\
&\leq 3\sup_u |F_n(u)) - F(u)|.
\end{aligned}$$

En particular, la acotación no depende de la x.

De la inecuación (7) y por Borel Cantelli (James (2002)), se deduce que

$$\sup_{P\in\mathbb{P}} E(\sup_{x\in\mathbb{E}} |D(P_{n,f}, f(x)) - D(P_f, f(x))|).$$

La conclusión es directa por el teorema de convergencia mayorada.

□

Teorema 4. *Supongamos que* $\mathbb{E} = \mathbb{R}^d$ *y se satisface*

$$sup_{\|x\|=1} Q\{f : f(u) \leq \epsilon\} = O(\epsilon).$$

Entonces, si el punto más profundo, x_0*, es único y* $D(P,x)$ *es continua en* x_0*, entonces* $\hat{x}_{o,n} \xrightarrow[n\to+\infty]{} x_0$ *en probabilidad.*

Demostración. Queremos ver que $\hat{x}_{o,n}$ converge a x_0 en probabilidad. Por los Lema 1 y Lema 2 del Apéndice, basta ver que toda subsucesión tiene una subsucesión convergente casi seguramente.

Por el Teorema 3, tenemos $\lim\limits_{\|x\|\to+\infty} D(P,x) = 0$.

Podemos suponer que $\hat{x}_{o,n}$ está contenida en un compacto de $\mathbb{R}^d$. Sea $\hat{x}_{o,n_k}$ subsucesión, entonces tiene una subsucesión convergente c.s. a x_1 en el compacto. Por comodidad y sin perdida de generalidad la llamaremos del mismo modo.

Por el teorema anterior, $D(P_{n_k}, x) \to D(P,x)$ uniformente en x y P, entonces $|D(P_{n_k}, \hat{x}_{o,n_k}) - D(P, x_o)| \to 0$. Sumado a la continuidad de $D(P,x)$ en x_0 y que tiene un único máximo, entonces $x_o = x_1$.

□

El próximo resultado habla de la distribución asintótica de $D(P_n, x)$ como estimador de $D(P,x)$.

Teorema 5. *Sea* $\{X_j\}_{j\geq 1}$ *una sucesión de elementos aleatorios con valores en* $\mathbb{E}$*. Dado* $x \in \mathbb{E}$*, supongamos que para todo* f *la función de distribución de* $f(X)$ *es continua en* $f(x)$*.*

1. *Asumimos que el soporte de la distribución* P_X *de* X *es* $\mathbb{E}$ *y* $x \in \mathbb{E}$ *satisface que*

$$Q(\{f : P(f(X) \leq f(x)) = \frac{1}{2}\}) < 1.$$

Entonces,

$$\sigma^2 = \int\int (p_{ff'} - p_f p_{f'})(1 - 2p_f - 2p_{f'} + 4p_f p_{f'}) dQ(f) dQ(f') > 0,$$

donde, $p_{ff'} = P(f(X) \leq f(x), f'(X) \leq f'(x))$*,* $p_f = P(f(X) \leq f(x))$ *y* $p_{f'} = P(f'(X) \leq f'(x))$*.*

Además,

$$\sqrt{n}(D(P_n, x) - D(P, x)) \xrightarrow[n\to+\infty]{} N(o, \sigma^2),$$

débilmente cuando $n \to \infty$*, es decir, que la sucesión* $D_n = \sqrt{n}(D(P_n, x) - D(P, x))$ *es asintóticamente normal con media cero y varianza* σ^2*.*

Más aún, teniendo en cuenta la desigualdad de Berry-Esseen (ver Apéndice) se obtiene la siguiente cota,

$$\sup_t |P(D_n \leq t) - \Phi(t)| \leq C v \sigma^3 n^{-1/2}, \tag{8}$$

donde Φ *es la función de distribución acumulada para una distribución normal estándar,* v *es una constante que depende de* P *y* C *es una constante universal.*

2. *Si* x *es tal que* $\sigma^2 = 0$ *y* $Q(\{f : 0 < P(f(X) \leq f(x)) < 1\}) > 0$*, tenemos que* $n(D(P_n, x) - D(P, x))$ *converge débilmente a una variable aleatoria* ξ *que es la suma de variables aleatorias independientes* χ_1^2 *centradas. Más precisamente,*

$$\xi = \sum_{j=1}^{\infty} \lambda_j (Z_j^2 - 1),$$

donde $Z_j : j \geq 1$ *son* $N(0,1)$ *i.i.d. y los coeficientes* λ_j *son los autovalores del operador lineal* T *asociado con nucleo* $h(x, y)$ *especificado en la prueba.*

Demostración. Notamos F la distribución de $f(x)$ y F_n la correspondiente distribución empírica. Usaremos $F_n(f(x))(1 - F_n(f(x)))$ y $F(f(x))(1 - F(f(x)))$ en lugar de $F_n(f(x))(1 - F_n(f(x^-)))$ y $F(f(x))(1 - F(f(x^-)))$, respectivamente, en la definición de la profundidad simplicial dado que $f(X)$ es continua en $f(x)$ para casi toda f y la diferencia $D(P, x) - D(P_n, x)$ usando las dos notaciones coincide a.s.

1. Tenemos que

$$
\begin{aligned}
D(P_n, x) = & \\
= & \int D(P_{n,f}, f(x)) dQ(f) \\
= & \int (\frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}})(1 - \frac{1}{n} \sum_{j=1}^{n} \mathcal{I}_{\{f(X_j) < f(x)\}}) dQ(f).
\end{aligned}
$$

Observemos que

$$
\begin{aligned}
& (\frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}})(1 - \frac{1}{n} \sum_{j=1}^{n} \mathcal{I}_{\{f(X_j) < f(x)\}}) \\
&= \sum_{i=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}} - \frac{1}{n^2} (\sum_{i=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}})(\sum_{j=1}^{n} \mathcal{I}_{\{f(X_j) < f(x)\}}) \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)} - \frac{1}{n^2 \}} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}} \mathcal{I}_{\{f(X_j) < f(x)\}} \\
&= \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}} - \sum_{i=1}^{n} \sum_{j=1}^{n} \mathcal{I}_{\{f(X_i) < f(x)} \} \mathcal{I}_{\{f(X_j) < f(x)\}}.
\end{aligned}
$$

Si $i = j$

$$
\begin{aligned}
& \mathcal{I}_{\{f(X_i) < f(x)\}} - \mathcal{I}_{\{f(X_i) < f(x)\}} \mathcal{I}_{\{f(X_j) < f(x)\}} \\
&= \mathcal{I}_{\{f(X_i) < f(x)\}} - (\mathcal{I}_{\{f(X_i) < f(x)\}})^2 \\
&= \mathcal{I}_{\{f(X_i) < f(x)\}} - \mathcal{I}_{\{f(X_i) < f(x)\}} = 0.
\end{aligned}
$$

Tenemos

$$
\begin{aligned}
D(P_n, x) = & \\
= & \frac{1}{n^2} \sum_{i \neq j}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}} - \mathcal{I}_{\{f(X_i) < f(x)\}} \mathcal{I}_{\{f(X_j) < f(x)\}} \\
= & \frac{1}{n^2} \sum_{i \neq j}^{n} \mathcal{I}_{\{f(X_i) < f(x)\}} (1 - \mathcal{I}_{\{f(X_j) < f(x)\}}).
\end{aligned}
$$

Llamemos $h^* = \int \mathcal{I}_{\{f(u) \leq f(x)\}} (1 - \mathcal{I}_{\{f(v) \leq f(x)\}}) dQ(f)$ y observando que

$$
\frac{1}{n(n-1)} - \frac{1}{n^2(n-1)} = \frac{n-1}{n^2(n-1)} = \frac{1}{n^2}.
$$

Nos queda,

$$D(P_n, x) = \frac{1}{n(n-1)} \sum_{i \neq j}^{n} h^*(X_i, X_j) - \frac{1}{n^2(n-1)} \sum_{i \neq j}^{n} h^*(X_i, X_j).$$

Si definimos $h(u, v) = \frac{h^*(u,v)+h^*(v,u)}{2}$. Entonces,

$$\frac{1}{n(n-1)} \sum_{i \neq j} h(X_i, X_j),$$

es un U-estadístico con núcleo simétrico h (ver Apéndice). Además,

$$\sum_{i \neq j}^{n} h(X_i, X_j) = \frac{1}{2} \sum_{i \neq j}^{n} h^*(X_i, X_j) + h^*(X_j, X_i) = \frac{1}{2} (\sum_{i \neq j}^{n} h^*(X_i, X_j) + \sum_{i \neq j}^{n} h^*(X_j, X_i)),$$

hacemos un cambio de subíndices en la segunda suma, j por i, y nos queda

$$\sum_{i \neq j}^{n} h(X_i, X_j) = \frac{1}{2} \sum_{i \neq j}^{n} 2h^*(X_i, X_j) = \sum_{i \neq j}^{n} h^*(X_i, X_j).$$

Luego,

$$D(P_n, x) = \frac{1}{n(n-1)} \sum_{i \neq j}^{n} h(X_i, X_j) - \frac{1}{n^2(n-1)} \sum_{i \neq j}^{n} h^*(X_i, X_j).$$

Se ve rápido que

$$|h^*(u, v)| = \int \mathcal{I}_{\{f(u) \leq f(x)\}} (1 - \mathcal{I}_{\{f(v) \leq f(x)\}}) dQ(f) \leq 1.$$

Por lo tanto,

$$\sqrt{n} \frac{1}{n^2(n-1)} \sum_{i \neq j}^{n} h^*(X_i, X_j) \leq \sqrt{n} \frac{1}{n^2(n-1)} n^2 \leq \frac{\sqrt{n}}{n-1} \xrightarrow[n \to +\infty]{} 0.$$

Basta encontrar la distribución asintótica de

$$\sqrt{n} (\frac{1}{n(n-1)} \sum_{i \neq j}^{n} h(X_i, X_j) - D(P, x)),$$

Usaremos el Teorema Central del Límite para U-estadísticos no-degenaros, ver el **Teorema A** del Apéndice.

Sea

$$\begin{aligned} h_1(u) &= E(h(u, X_1)) \\ &= \frac{1}{2}\int \mathcal{I}_{\{f(u)\leq f(x)\}}(1 - P(f(X_1) \leq f(u)))dQ(f) \\ &+\frac{1}{2}\int P(f(X_1) \leq f(x))(1 - \mathcal{I}_{\{f(u)\leq f(x)\}})dQ(f). \end{aligned}$$

Veamos que se satisfacen las hipótesis

- $E(h^2(X_1, X_2)) \leq \infty$ dado que $h \leq 1$.
- $h_1(X)$ es insesgado ya que

$$E(h_1(X)) = \int P(f(X) \leq f(x))(1 - P(f(X) \leq f(x)))dQ(f) = D(P, x).$$

- Debemos probar que $\zeta_1 = Var(h_1(X))$ es estrictamente positiva.

$$\begin{aligned} h_1(X) - D(P, x) &= h_1(X) - E(h_1(X)) \\ &= \frac{1}{2}\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(1 - p_f) + p_f(p_f - \mathcal{I}_{\{f(X)\leq f(x)\}})dQ(f). \end{aligned}$$

Luego,

$$\begin{aligned} Var(h_1(X)) &= E([h_1(X) - E(h_1(X))]^2) \\ &= E(\frac{1}{4}[\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(1 - p_f) + p_f(p_f - \mathcal{I}_{\{f(X)\leq f(x)\}})dQ(f)]^2) \\ &= E(\frac{1}{4}\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(1 - p_f) + p_f(p_f - \mathcal{I}_{\{f(X)\leq f(x)\}})dQ(f) \\ &\times \int (\mathcal{I}_{\{f'(X)\leq f'(x)\}} - p_{f'})(1 - p'_f) + p_{f'}(p_{f'} - \mathcal{I}_{\{f'(X)\leq f'(x)\}})dQ(f') \end{aligned}$$

$$= \frac{1}{4}E(\int\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(\mathcal{I}_{\{f'(X)\leq f'(x)\}} - p'_f))(1 - p_f)(1 - p'_f))dQ(f)dQ(f') \tag{9}$$

$$= \frac{1}{4}E(\int\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(\mathcal{I}_{\{f'(X)\leq f'(x)\}} - p'_f))p_f(1 - p'_f))dQ(f)dQ(f') \tag{10}$$

$$= \frac{1}{4}E(\int\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(\mathcal{I}_{\{f'(X)\leq f'(x)\}} - p'_f))(1 - p_f)p'_f)dQ(f)dQ(f') \tag{11}$$

$$= \frac{1}{4}E(\int\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(\mathcal{I}_{\{f'(X)\leq f'(x)\}} - p'_f))p_f p'_f)dQ(f)dQ(f'). \tag{12}$$

A continuación desarrollaremos los términos (9), (10), (11) y (12).
El primer término (9) queda,

$$\begin{aligned}E(\mathcal{I}_{\{f(X)\leq f(x)\}}\mathcal{I}_{\{f'(X)\leq f'(x)\}} - \mathcal{I}_{\{f(X)\leq f(x)\}}p_{f'} - p_f\mathcal{I}_{\{f'(X)\leq f'(x)\}} + p_fp_{f'})(1-p_f)(1-p_{f'})\\ = (p_{ff'} - p_fp_{f'} - p_fp_{f'} + p_fp_{f'})(1-p_f)(1-p_{f'})\\ = (p_{ff'} - p_fp_{f'} - p)(1-p_f)(1-p_{f'}).\end{aligned}$$

El segundo término (10),

$$\begin{aligned}E((\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(p_{f'} - \mathcal{I}_{\{f'(X)\leq f'(x)\}})p_{f'}(1-p_f)\\ = (p_fp_{f'} - p_{f'}p_f - p_{ff'} + p_fp_{f'})p_{f'}(1-p_f)\\ = (p_fp_{f'} - p_{ff'})p_{f'}(1-p_f).\end{aligned}$$

El tercero término (11)

$$\begin{aligned}E((p_f - \mathcal{I}_{\{f(X)\leq f(x)\}})(\mathcal{I}_{\{f'(X)\leq f'(x)\}} - p_{f'})p_f(1-p_{f'})\\ = (p_fp_{f'} - p_{f'}p_f - p - ff' + p_fp_{f'})p_f(1-p_{f'})\\ = (p_fp_{f'} - p_{ff'})p_f(1-p_{f'}).\end{aligned}$$

Y por último tenemos (12)

$$\begin{aligned}E((p_f - \mathcal{I}_{\{f(X)\leq f(x)\}})(p_{f'} - \mathcal{I}_{\{f'(X)\leq f'(x)\}}))p_fp_{f'}\\ = (p_fp_{f'} - p_fp_{f'} - p_fp_{f'} + p_{ff'})p_fp_{f'}\\ = (p_{ff'} - p_fp_{f'})p_fp_{f'}.\end{aligned}$$

Sumando (9), (10), (11) y (12) queda

$$\begin{aligned}Var(h_1(X)) &= \\ &= (p_{ff'} - p_fp_{f'})[(1-p_f)(1-p_{f'}) - p_{f'}(1-p_f) - p_f(1-p_{f'}) + p_fp_{f'}]\\ &= (p_{ff'} - p_fp_{f'})[1 - p_{f'} - p_f + p_fp_{f'} - p_{f'} + p_{f'}p_f - p_f + p_fp_{f'}p_fp_{f'}]\\ &= (p_{ff'} - p_fp_{f'})[1 - 2p_f - 2p_{f'} + 4p_fp_{f'}]\\ &= (p_{ff'} - p_fp_{f'})(1-2p_f)(1-2p_{f'})\end{aligned}$$

Es decir,

$$\begin{aligned}&Var(h_1(X))\\ &\qquad = \frac{1}{4}\int\int(p_{ff'} - p_fp_{f'})(1-2p_f)(1-2p_{f'})dQ(f)dQ(f')\\ &\qquad = \frac{1}{4}E[(\int(\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(1-2p_f)dQ(f))^2] \geq 0.\end{aligned}$$

Esta cantidad es la que denotamos σ^2 en el enunciado del teorema, nos queda ver que es estrictamente positiva. Si tenemos que $\sigma^2 = 0$.

$$\int (\mathcal{I}_{\{f(X)\leq f(x)\}} - p_f)(1-2p_f)dQ(f) = 0 \text{ c.s.}$$

$$P_X(\{y \in E\colon \int \mathcal{I}_{\{f(y)\leq f(x)\}}(1-2p_f)dQ(f) = \int p_f(1-2p_f)dQ(f)\}) = 1.$$

Tomando $z = 2x - y$

$$P_X(\{z \in E\colon \int \mathcal{I}_{\{f(z)\geq f(x)\}}(1-2p_f)dQ(f) = \int p_f(1-2p_f)dQ(f)\}) = 1,$$

entonces

$$\int (1-2p_f)dQ(f) = 2\int p_f(1-2p_f)dQ(f) \Rightarrow \int (1-2p_f)^2 dQ(f) = 0.$$

Contradice que $p_f \neq \frac{1}{2}$ con probabilidad positiva.
Por lo tanto, σ^2 es estrictamente positivo y el resultado sigue del Teorema Central para U-estadísticos. La cota en (8) se sigue del **Teorema B** del Apéndice y de Callaert y Janssen (1978). Este resultado requiere que $\sigma > 0$ y además que $E|h|^3 < \infty$. Se verifica en este caso ya que h es acotada. La constante υ es precisamente $E|h|^3$.

2. Ahora, dado que $\sigma^2 = 0$, usaremos el **Teorema C** del Apéndice. Basta probar que $E(h^2(X,Y)) > 0$, donde X e Y son elementos aleatorios i.i.d en $\mathbb{E}$.

$$\begin{aligned}
E(h^2(X,Y)) = & \\
= \;& \frac{1}{4}E\int\int (\mathcal{I}_{\{f(X)\leq f(x)\}}\mathcal{I}_{\{f(Y)>f(x)\}} + \mathcal{I}_{\{f(Y)\leq f(x)\}}\mathcal{I}_{\{f(X)>f(x)\}}) \times \\
\times \;& (1_{\{f'(X)\leq f'(x)\}}\mathcal{I}_{\{f'(Y)>f'(x)\}} + \mathcal{I}_{\{f'(Y)\leq f'(x)\}}\mathcal{I}_{\{f'(X)>f'(x)\}})dQ(f)dQ(f') \\
\times \;& \frac{1}{4}\int\int E[(\mathcal{I}_{\{f(X)\leq f(x)\}}\mathcal{I}_{\{f(Y)>f(x)\}} + \mathcal{I}_{\{f(Y)\leq f(x)\}}\mathcal{I}_{\{f(X)>f(x)\}}) \times \\
\times \;& (\mathcal{I}_{\{f'(X)\leq f'(x)\}}\mathcal{I}_{\{f'(Y)>f'(x)\}} + \mathcal{I}_{\{f'(Y)\leq f'(x)\}}\mathcal{I}_{\{f'(X)>f'(x)\}})dQ(f)dQ(f')].
\end{aligned}$$

Sean $A_f = \{f(X) \leq f(x)\}$, $A_{f'} = \{f'(X) \leq f'(x)\}$ y notamos A^c el complemento del conjunto A. Podemos distribuir las indicadoras. Veamos el primer término de dicha suma, quedaría

$$\mathcal{I}_{\{f(X)\leq f(x)\}}\mathcal{I}_{\{f(Y)>f(x)\}}\mathcal{I}_{\{f'(X)\leq f'(x)\}}\mathcal{I}_{\{f'(Y)>f'(x)\}},$$

tomando esperanza y usando el hecho de que X e Y son independientes,

$$E[\mathcal{I}_{\{f'(X)\leq f'(x)\}}\mathcal{I}_{\{f'(X)\leq f'(x)\}}]E[\mathcal{I}_{\{f'(Y)>f'(x)\}}\mathcal{I}_{\{f(Y)>f(x)\}}] = P(A_f\cap A_{f'})P(A_f^c\cap A_{f'}^c).$$

Resolviendo análogamente los otros términos nos queda,

$$E(h^2(X,Y)) = \frac{1}{2}\int\int[P(A_f\cap A_{f'})P(A_f^c\cap A_{f'}^c)+P(A_f\cap A_{f'}^c)P(A_f^c\cap A_{f'})dQ(f)dQ(f').$$

Ahora, fijados $f, f' \in \{f : 0 < P(f(X) \leq f(x) < 1\}$:
Si $P(A_f \cap A_{f'}) = 0$ ó $P(A_f^c \cap A_{f'}^c) = 0 \Rightarrow P(A_f \cap A_{f'}^c) > 0$ y $P(A_f^c \cap A_{f'}) > 0$.
Luego, $E(h^2(X,Y)) > 0$.

□

Este último resultado provee una base teórica para el uso de proyecciones unidimensionales. Cuesta-Albertos, Fraiman y Ransford (2007) han mostrado un resultado similiar para espacios de Hilbert, algunas aplicaciones se pueden encontrar en problemas de bondad de ajuste, estimadores de posición y dispersión, etc. Esta es una generalización a espacios de Banach. En algún sentido, este resultado nos da una motivación para el uso del dual en la metodología con la que se tratan los espacios de Banach.

Teorema 6. *Sea $\mathbb{E}$ un espacio de Banach con dual $\mathbb{E}'$ separable y sea μ una medida Gaussiana no-degenerada en $\mathbb{E}'$ cuya distribución coincide con la de $\sum_i G_i f_i$ donde $f_i \in \mathbb{E}'$ con $\sum_i \|f_i\|^2 < \infty$ y G_i son variables aleatorias i.i.d normales estándar. Sean P y M dos medidas de probabilidad borelianas en $\mathbb{E}$ y se cumple:*

- *Los momentos absolutos, $m_n = \int \|x\|^n dP(x)$ son finitos con $\sum_{n\geq 1} m_n^{-\frac{1}{n}} = \infty$, esta última condición se conoce con el nombre de condición de Carleman.*
- *El conjunto $\varepsilon(P,M) = \{f \in \mathbb{E}' \ : \ P_f = M_f\}$ tiene medida positiva para μ.*

Entonces, $P = M$.

Demostración. Sea $W = \sum_i G_i f_i$ un proceso gaussiano no-degenerado con distribución μ, donde G_i normales estándar independientes y $f_i \in \mathbb{E}'$ (ver Proposición 3.6 de Ledoux y Talagrand (1991)). Para cada $n \geq 1$, sea $\mathfrak{F}_n =< f_1, \ldots, f_n >$, el mínimo subespacio cerrado que contiene a $\{f_1, \ldots, f_n\}$, y $\mathfrak{F}_n^+ =< f_j : \ j \geq n+1 >$.

Afirmo: $\bigcup_{n\geq 1} \mathfrak{F}_n$ es denso en $\mathbb{E}'$.

De otra forma tendríamos $\mathfrak{F} = \overline{\bigcup_{n\geq 1} \mathfrak{F}_n} \neq \mathbb{E}'$. Por el Teorema de Hahn-Banach, existe un elemento $\Psi \in \mathbb{E}''$ tal que tomaría el valor cero en $\mathfrak{F}$. Entonces, tendríamos $\Psi(W) = 0$, lo que contradice que W es no-degenerado. Ahora para cada $n \geq 1$ *y* $g \in \mathfrak{F}_n^+$, consideramos la g-sección $\varepsilon(P,M)^g$ de $\varepsilon(P,M)$.

$$\varepsilon(P,M)^g := \{f \in \mathfrak{F}_n : f + g \in \mathfrak{F}_n\}.$$

Fijado $n \geq 1$, por el Teorema de Fubini,

$$P(\sum_{i=1}^{\infty} G_i f_i \in \varepsilon(P,M)) = \int_{\mathfrak{F}_n^+} P(\sum_{i=1}^{n} G_i f_i \in \varepsilon(P,M)^g) d\nu_n(g),$$

donde ν_n es la distribución de $\sum_{i=n+1}^{\infty} G_i f_i$. Dado que esta probabilidad es positiva, existe $g \neq 0$ tal que

$$P(\sum_{i=1}^{n} G_i f_i \in \varepsilon(P,M)^g) > 0.$$

Además,

$$P(\sum_{i=1}^{n} G_i f_i \in \varepsilon(P,M)^g) = P((G_1, \ldots, G_n) \in \{y \in \mathbb{R}^n : \sum_{i=1}^{n} y_i f_i \in \varepsilon(P,M)^g\}).$$

Tenemos que la medida de Lebesgue λ_n del conjunto $\{y \in \mathbb{R}^n : \sum_{i=1}^{n} y_i f_i \in \varepsilon(P,M)^g\}$ es también positiva. Como $\varepsilon(P,M)$ es un cono, sigue que $\{y \in \mathbb{R}^n : \sum_{i=1}^{n} y_i f_i \in \varepsilon(P,M)^{tg}\}$ tiene medida de Lebesgue positiva para cada $t \in \mathbb{R} - \{0\}$. Por lo tanto, tenemos que,

$$\lambda_{n+1}\{y \in \mathbb{R}^{n+1} : \sum_{i=1}^{n} y_i f_i + y_{n+1} g \in \{f \in \mathbb{E}' : P_f = M_f\}\} > 0 \tag{13}$$

Finalmente, notando $U_1 = (f_1(X_1), \ldots, f_n(X_1), g(X_1))$ y $U_2 = (f_1(X_2), \ldots, f_n(X_2), g(X_2))$, donde X_1 , X_2 tienen distribución P y M respectivamente, la ecuación (13) se lee como

$$\lambda_{n+1}\{y \in \mathbb{R}^{n+1} : P_{y^t U_1} = P_{y^t U_2}\} > 0. \tag{14}$$

Usamos el Corolario 3.2 en Cuesta-Albertos, Fraiman y Ransford (2007) aplicado a U_1 y a U_2. Este resultado es análogo al que queremos probar en este teorema para distribuciones P and $M \in \mathbb{R}^d$ (con $d \geq 2$), que tiene dos requerimientos. El segundo lo chequeamos en (14) al ver que el conjunto tenía medida de Lebesgue positiva y el primero es ver que se satisface la Condición de Carleman para U_1, es decir que, los momentos absolutos de $m_n := \int \|x\|^n dP(x)$ son finitos y satisfacen que $\sum_{n\geq 1} m_n^{\frac{-1}{n}} = \infty$. Definimos $A = \sqrt{(\sum_i \|f_i\|^2)}$ y $C = \sqrt{(A^2 + \|g\|^2)}$ y tenemos

$$\|U_1\| = (\sum_{i=1}^{n} f_i^2(X_1) + g^2(X_1))^{\frac{1}{2}} \leq (\sum_{i=1}^{n} \|f_i\|^2 \|X_1\|^2 + \|g\|^2 \|X_1\|^2)^{\frac{1}{2}} = C\|X_1\|.$$

Luego, $(E\|U_1\|^p)^{\frac{1}{p}} \leq C(E\|X_1\|^p)^{\frac{1}{p}}$, por lo tanto U_1 satistafe la condición de Carleman.

Por lo tanto, tenemos que U_1 y U_2 tienen la misma distribución. En particular, $(f_1(X_1), \ldots, f_n(X_1))$ y $(f_1(X_2), \ldots, f_n(X_2))$ tienen la misma distribución . Además, como $\bigcup_{n\geq 1} \mathfrak{F}_n$ es denso en $\mathbb{E}'$ y los funcionales característicos de X_1 y X_2 son continuos en la topología del dual, tenemos que X_1 tiene la misma distribución de X_2, i.e. $P = M$. $\square$

5. Apéndice

5.1. Resultados previos

5.1.1. Desigualdad de Dvoretzky-Kiefer-Wolfowitz

Sea F_n la distribución empírica de una muestra de variables aleatorias i.i.d. con distribución F.

$$P(\sqrt{n}\ sup_x(F_n(x) - F(x)) > \lambda) \leq C\exp(-2\lambda^2),$$

donde C es una constante no especificada (Dvoretzky, Kiefer and Wolfowitz 1956). Masart (1990) prueba que se puede tomar $C = 1$ si $\exp(-2\lambda^2)$. En particular la inecuación es

$$P(\sqrt{n}\ sup_x(F_n(x) - F(x)) > \lambda) \leq 2\exp(-2\lambda^2).$$

5.1.2. Desigualdad de Berry-Essén

Sean $\{X_i\}$ variables aletorias i.i.d con esperanza μ y varianza $\sigma^2 > 0$,
sea

$$S_n = \frac{\sum_{i=1}^n X_i - E[\sum_{i=1}^n X_i]}{\sqrt{Var[\sum_{i=1}^n X_i]}},$$

y sea $G_n(t) = P(S_n \leq t)$. Entonces,

$$sup_t\ |G_n(t) - \Phi(t)| \leq \frac{33}{4}\frac{E|X_1 - \mu|^3}{\sigma^3 n^{\frac{1}{2}}},\ \forall\ n,$$

donde $\Phi(\cdot)$ es la función de distribución acumulada de una distribución normal estándar.

5.1.3. U-estadísticos

Los siguientes resultados se encuentran en Serfling (1980).

Definición 6. *Sean $X_1, X_2, \ldots$ variables aleatorias identicamente distribuídas con distribución F. Consideramos una "función paramétrica"$\theta = \theta(F)$ para la cual hay un estimador insesgado. Esto es, $\theta(F)$ puede ser representado como*

$$\theta(F) = E_F\{h(X_1, \ldots, X_m)\} = \int \ldots \int h(x_1, \ldots, x_m)dF(x_1)\ldots dF(x_m),$$

para alguna función $h = h(x_1, \ldots, x_m)$ que llamamos núcleo.

Observación: *Sin perdida de generalidad podemos asumir que h es simétrica. Si no, lo podemos reemplazar por un núcleo simétrico*

$$\frac{1}{m!}\sum_P h(x_{i_1}, \ldots, x_{i_m}),$$

donde $\sum_P$ denota la suma sobre todas las $m!$ permutaciones $(i_1, \ldots, i_m)$ de $(1, \ldots, m)$. Para cualquier núcleo h, el correspondiente U-estadístico *para estimar θ en base a la muestra $X_1, \ldots, X_n$ de tamano $n \geq m$ se define promediando el núcleo h simétrico sobre las observaciones:*

$$U_n = U(X_1, \ldots, X_n) = \frac{1}{\binom{n}{m}} \sum_c h(x_{i_1}, \ldots, x_{i_m}),$$

donde $\sum_c$ es la suma sobre las $\begin{pmatrix} n \\ m \end{pmatrix}$ combinaciones de m elementos $\{i_1, \ldots, i_m\}$ de $\{1, \ldots, n\}$.

Definición 7. *Consideremos un núcleo simétrico $h(x_1, \ldots, x_m)$ con $E_F|h(X_1, \ldots, X_m)| < \infty$. Definimos la función asociada*

$$h_c(x_1, \ldots, x_c) = E_F\{h(x_1, \ldots, x_c, X_{c+1}, \ldots, X_m)\}$$

para cada $c = 1, \ldots, m-1$ y $h_m = h$.

Definición 8. *Consideremos un núcleo simétrico $h(x_1, \ldots, x_m)$ satisfaciendo*

$$E_F\{h^2(X_1, \ldots, X_m)\} < \infty,$$

y h_c la función asociada. Definimos $\zeta_0 = 0$ y para cada $1 \leq c \leq m$,

$$\zeta_c = Var_F\{h_c(X_1, \ldots, X_c)\}.$$

Observación 4. $0 = \zeta_0 \leq \zeta_1 \leq \cdots \leq \zeta_m = Var_F\{h\} < \infty$.

Teorema A
Si $E_F h^2 < \infty$ y $\zeta_1 > 0$ entonces

$$n^{\frac{1}{2}}(U_n - \theta) \xrightarrow{d} N(0, m^2\zeta_1).$$

La convergencia asintótica de los U-estaísticos fue estudiada por varios autores, entre ellos, Callaert y Janssen (1978), que obtuvieron el siguiente resultado.

Teorema B
Si $\upsilon = E|h|^3 < \infty$ y $\zeta_1 > 0$, entonces

$$sup_t|P(\frac{\sqrt{n}(U_n - \theta)}{m\zeta_1^{\frac{1}{2}}} \leq t) - \Phi(t)| \leq C\upsilon(m^2\zeta_1)^{-\frac{3}{2}}n^{-\frac{1}{2}},$$

donde C es una constante absoluta.

Teorema C
Si $E_F(h^2) < \infty$ y $\zeta_1 = 0 < \zeta_2$, entonces

$$n(U_n - U) \xrightarrow{d} \frac{m(m-1)}{2} Y,$$

donde Y es una variable aleatoria de la forma $Y = \sum_{j=1}^{\infty} \lambda_j(\chi^2_{1j} - 1)$, y χ^2_{1j} son variable aleatorias independientes con distribución chi cuadrado.

5.2. Demostración de un Lema de medida

Lema 1. *Sea $\{X_n\}$ una sucesión tal que $X_n \xrightarrow[n\to+\infty]{} X$ en probabilidad, entonces existe una subsucesión $X_{n_k} \xrightarrow[k\to+\infty]{} X$ casi seguro.*

Demostración. Tenemos que $\forall m > 0$, $\lim\limits_{n\to+\infty} P(|X_n - X| > \frac{1}{m}) = 0$. Queremos que ver que existe una subsucesión X_{n_k} tal que $P(X_{n_k} \xrightarrow[k\to+\infty]{} X) = 1$. Lo que es equivalente a ver que

$$\forall m > 0,\ P(\bigcup_m \bigcap_N \bigcup_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}) = 1.$$

Buscamos n_k tal que $P(\Omega \backslash \bigcap_m \bigcup_N \bigcap_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}) = 0$.

$$\begin{aligned} P(\Omega\backslash \bigcap_m \bigcup_N \bigcap_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}) = & \\ &= P(\bigcup_m \bigcap_N \bigcup_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}) \\ &\leq \sum_m P(\bigcap_N \bigcup_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}). \end{aligned}$$

Los conjuntos $\bigcup_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}$ son decrecientes cuando N crece.

$$\sum_m P(\bigcap_N \bigcup_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}) = \sum_m \lim_{N\to+\infty} P(\bigcup_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\}).$$

Más aún,

$$\bigcap_{N=1}^{S} \bigcup_{k\geq N} \{|X_{n_k} - X| < \frac{1}{m}\} = \bigcup_{k\geq S} \{|X_{n_k} - X| < \frac{1}{m}\}.$$

Entonces,

$$\sum_m \lim_{N\to+\infty} P(\bigcup_{k\geq N}\{|X_{n_k}-X|<\frac{1}{m}\}) \leq$$

$$\leq \sum_m P(\bigcup_{k\geq N(m)}\{|X_{n_k}-X|<\frac{1}{m}\})$$
$$\leq \sum_m \sum_{k\geq N(m)} P(\{|X_{n_k}-X|<\frac{1}{m}\}.$$

Para algún N que depende de m, N(m). Sea $\epsilon > 0$ y n_k para que se cumpla que

$$P(\{|X_{n_k}-X|<\frac{1}{m}\}) \leq \frac{\epsilon}{2^k}$$

Podemos hacer esta elección porque X_n converge a X en probabilidad. Obtenemos

$$\sum_{k\geq N(m)} P(\{|X_{n_k}-X|<\frac{1}{m}\}) \leq \sum_{k\geq N(m)} \frac{\epsilon}{2^k} = \frac{1}{2^{N(m)}}\sum_{k=1}^{\infty}\frac{\epsilon}{2^k} = \frac{\epsilon}{2^{N(m)}}.$$

Luego,

$$\sum_m \sum_{k\geq N(m)} P(\{|X_{n_k}-X|<\frac{1}{m}\}) \leq \sum_m \frac{\epsilon}{2^{N(m)}} \leq \sum_m \frac{\epsilon}{2^m} = \epsilon.$$

Por lo tanto,

$$P(\Omega\backslash \bigcap_m \bigcup_N \bigcap_{k\geq N}\{|X_{n_k}-X|<\frac{1}{m}\}) \geq \epsilon.$$

Dado que la elección de n_k es independiente del ϵ arbitrario.

$$P(\Omega\backslash \bigcap_m \bigcup_N \bigcap_{k\geq N}\{|X_{n_k}-X|<\frac{1}{m}\}) = 0.$$

$\square$

Lema 2. *Sea* $\{X_n\}$ *una sucesión,* $X_n \xrightarrow[n\to+\infty]{} X$ *en probabilidad si y sólo si cada subsucesión* $\{X_{n_k}\}$ *contiene una subsucesión* $\{X_{n_{k_i}}\}$ *tal que* $X_{n_{k_i}} \xrightarrow[k\to+\infty]{} X$ *casi seguro.*

Demostración. Si $X_n \xrightarrow[n\to+\infty]{} X$ en probabilidad y $\{X_{n_k}\}$ es una subsucesión, entonces $X_{n_k} \xrightarrow[n\to+\infty]{} X$ en probabilidad y por el lema anterior existe una subsucesión $\{X_{n_{k_i}}\}$ que converge a X c.s.

Recíprocamente, supongamos que toda subsucesión tiene una subsucesión que converge a X c.s. y supongamos que X_n no converge a X en probabilidad para obtener una contradicción.

En ese caso, existe una subsucesión $\{X_{n_k}\}$, $\delta > 0$ y $\epsilon > 0$ tales que

$$P(|X_{n_k} - X| > \epsilon) \geq \delta.$$

Pero, esta subsucesión debería tener una subsucesión que converge a X c.s y por lo tanto en probabilidad. Llegamos a una contradicción.

$\square$

Bibliografía

- Callaerd, H. y Janssen, P. (1978). "The Berry-Esseen Theorem for U-Statistics", *Annals of Statistics*, **6 (2)**, 417-421.

- Cuesta-Albertos, J.A., Fraiman, R. and Ransford, T. (2007). "A sharp form of the Cramer–Wold theorem". *Journal of Theoretical Probability*, **20**, 201-209.

- Cuesta-Albertos, J. A. y Nieto-Reyes, A. (2008). "The random Tukey Depth". *Computational Statistics and Data Analysis*, **52**, 4979-4988.

- Cuevas A. y Fraiman R. (2009). "On Depth Measure and Dual Statistics. A Methodology for Dealing With General Data". *Journal of Multivariate Analysis*, **100**, 753-766.

- Cuevas A., Febrero, M. y Fraiman R. (2007). "Robust estimation and classification for functional data via projection-based depht notions". *Computational Statistics*.

- Dvoretzky, A., Kiefer, J. and Wolfowitz, J. (1956). "Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator". *Annals of Mathematical Statistics*. **27, 3**, 642-669.

- James B., (2002). "Probabilidade: Um Curso em Nível Intermediário (2da Ediçao)". *Associaçao Instituto Nacional de Matemática Pura e Aplicada*, 200-201.

- Fraiman, R. y Muniz, G. (2001). "Trimmed means for functional data". *Test*, **10**, 419-440.

- Ledoux, M. y Talagrand, M. (1991). "Probability in Banach Spaces". Springer, New York.

- Liu, R. (1990). "On a notion of data depth based on random simplices". *Annals of Statistics*, **18**, 405-414.

- Lopez-Pintado, S. and Romo, J. (2009) "On the concept of depth for functional data". *Journal of the American Statistical Association*, **104**, 486-503 .

- Massart, P. (1990). "The Tight Constant in the Dvoretzky-Kiefer-Wolfowitz Inequality", *Annals of Probability*, **18 (3)**, 1269-1283.

- Munoz M. y Blanco L. (2002). Ïntroducción a la Medida Avanzada de la Probabilidad". Unibiblos. 149

- Serfling, R. (1980). "Approximation Theorems of Mathematical Statistics". John Wiley and Sons, New York.

- Serfling, R. J. (2006). “Depth functions in nonparametric multivariate inference”, in: *Data Depth: Robust Multivariate analysis, Computational Geometry and Applications, in DIMAC Series in Discrete Mathematics and Theoretical Computer Science*, vol.72, 1-16.
- Tukey, J. W. (1975). “Mathematics and the picturing of data”. *Proc. Int. Congress Math.*, Vancouver 2, 523-531.
- Zuo, Y. y Serfling R. (2000). “General Notion of Statistical Depth Function”, in: *The Annals of Statistics*, vol 28, No 2, 461-482.