Tesis de Maestría

Estimación de Conjuntos y Longitudes

Una aproximación no paraétrica

Alejandro Cholaquidis

Orientador: Ricardo Fraiman

Maestría en Matemática
Facultad de Ciencias
Universidad de la República
Montevideo Uruguay

Resumen

Esta tesis está dividida en tres partes, en la primera se presentan algunos resultados referidos a la estimación de conjuntos para el caso en que podemos discernir si un punto de una muestra pertenece o no al conjunto. En particular estudiaremos los casos en que el conjunto es el soporte de una densidad o un conjunto de nivel de la misma. La segunda se centra en el cálculo del contenido de Minkowski de su superficie lateral; serán necesarias hipótesis de regularidad sobre el borde del conjunto. La tercera parte se centra en el estudio del caso en el que no se puede determinar si los puntos están en el conjunto a estimar o no. Presentamos una generalización de [4].

Palabras Claves: Contenido de Minkowski, Estimación no paramétrica. Conjunto α-convexo.

Abstract

This thesis is divided in three parts. In the first one we present some results referred to set estimation in the case we can discern if a point from a sample point is in the set or not. In particular, we will study the cases when the set is the support of a density or a level set. The second one is focused in the calculation of the Minkowski's content of its lateral surface; there will be necessary some hypothesis of regularity for the boundary of the set. Finally, the third part is centered in the study of the case when it is not possible to determine if the point from a sample is in the set to estimate or not. We present a generalization of the work of [4].

Keywords: Minkoski content, Non parametric estmation. α-convex set.

Índice general

Capítulo 1

Estimación de conjuntos

1.1. Introducción

El problema de estimar un conjunto S a partir de una cantidad finita de datos ha sido tratado en diferentes trabajos. En computación, por ejemplo, una construcción eficiente de la envolvente convexa a partir de una cantidad finita de puntos tiene importantes aplicaciones en procesamiento de imágenes y en análisis de clusters, entre otras. Este primer capítulo introductorio pretende presentar algunos conceptos elementales así como posibles aplicaciones.

El término estimación refiere al problema estadístico de estimar un conjunto desconocido S a partir de una muestra aleatoria de puntos $\aleph_n = \{X_1, \ldots, X_n\}$ cuya distribución está relacionada con S, S podría ser por ejemplo el soporte de una medida P_X absolutamente continua, a partir de una muestra de ésta. Otro ejemplo típico, y que se mencionara en el presente trabajo con detalle más adelante, es el caso en que S es un conjunto de nivel de una densidad.

Es claro que para estimar un conjunto por otro, tenemos que tener una medida de proximidad de conjuntos. En algunos casos nos interesará medir que tan próximos están los puntos de un conjunto a los de otro, en cuyo caso usaremos la distancia de Hausdorff, que más adelante definiremos con precisión, y en otros la medida de dichos puntos, en cuyo caso tener un punto muy lejos en el sentido de la distancia Euclideana, medido con la medida de Lebesgue, no sera relevante.

1.2. Notación y definiciones previas

De aquí en adelante, salvo que se especifique lo contrario, vamos a trabajar en $\mathbb{R}^d$, usaremos la notación usual $\langle \cdot, \cdot \rangle$ para el producto interno, y $\| \cdot \|$ para la norma Euclideana. $B(x, r)$ denotará la bola cerrada de centro en el punto x y radio r, y $B := B(0, 1)$. Si $A \subset \mathbb{R}^d$, entonces A^c, $\overline{A}$, $\mathring{A}$ (o $int(A)$), ∂A y $conv(A)$ denotan complemento, clausura, interior, borde, y envolvente convexa de A, respectivamente. Definimos $diam(A) := \sup_{x,y \in A} \|x - y\|$ el diámetro de A.

1.2.1. Distancia de Hausdorff

Definición 1.1. Sean A y C subconjuntos compactos, no vacíos de $\mathbb{R}^d$. La distancia de Hausdorff entre A y C se define como:

$$d_H(A,C) = \max\left\{\sup_{a\in A} d(a,C),\ \sup_{c\in C} d(c,A)\right\},$$

siendo

$$d(a,C) = \inf\left\{\|a-c\| : c \in C\right\}.$$

El hecho de elegir subconjuntos compactos de $\mathbb{R}^d$ nos asegura que d_H es una métrica, no obstante, si sólo pedimos que sean acotados y no vacios, d_H esta bien definida. Por otro lado, si no le pedimos que sean cerrados, podria darse el caso en que la distancia de Hausdorff sea 0 y que los conjuntos no sean iguales.

Definición 1.2. Sea A un subconjunto compacto, no vacío, de $\mathbb{R}^d$. El entorno abierto de radio $\varepsilon > 0$ de A, que denotaremos $\mathring{B}(A,\varepsilon)$ se define como:

$$\mathring{B}(A,\varepsilon) = \left\{x \in \mathbb{R}^d : d(x,A) < \varepsilon\right\},$$

análogamente definimos

Definición 1.3. Sea A un subconjunto compacto, no vacío, de $\mathbb{R}^d$. El entorno cerrado de radio $\varepsilon > 0$ de A, que denotaremos $B(A,\varepsilon)$ se define como:

$$B(A,\varepsilon) = \left\{x \in \mathbb{R}^d : d(x,A) \le \varepsilon\right\}.$$

teniendo en cuenta estas definiciones podemos definir:

Definición 1.4. Sean A y C subconjuntos compactos, no vacios de $\mathbb{R}^d$. La distancia de Hausdorff entre A y C se define como:

$$d_H(A,C) = \inf\left\{\varepsilon > 0 : A \subset \mathring{B}(C,\varepsilon) \quad y \quad C \subset \mathring{B}(A,\varepsilon)\right\}.$$

Definición 1.5. Sean A y C subconjuntos de $\mathbb{R}^d$. La suma de Minkowski, $\oplus$, se define como

$$A \oplus C = \left\{a + c : a \in A,\ c \in C\right\},$$

y la resta, $\ominus$, como

$$A \ominus C = \left\{x : \{x\} \oplus C \subset A\right\}.$$

Para $\lambda > 0$ definimos

$$\lambda C = \left\{\lambda c : c \in C\right\},$$

por lo tanto $\mathring{B}(A,\varepsilon) = A \oplus \varepsilon B$.

Es inmediato verificar que también es posible definir:

Definición 1.6. Sean A y C subconjuntos compactos, no vacios de $\mathbb{R}^d$. La distancia de Hausdorff entre A y C se define como;

$$d_H(A,C) = \inf\left\{\varepsilon > 0 : A \subset C \oplus \varepsilon\, \mathring{B} \quad y \quad C \subset A \oplus \varepsilon\, \mathring{B}\right\}.$$

Observemos que si bien la proximidad en distancia de Hausdorff nos asegura la proximidad de los puntos de los conjuntos, no implica que los conjuntos sean parecidos en algun sentido, basta pensar en un círculo y puntos dentro, tomados con distribución uniforme, el círculo y el conjunto de puntos pueden estar muy próximos en distancia de Hausdorff y no obstante su forma difiere enormemente. Este mismo ejemplo nos permite ver que la distancia de Hausdorff tampoco garantiza proximidad en el sentido de la medida de los conjuntos, ni tampoco de la longitud de su borde, así, por ejemplo, podemos tener conjuntos que distan lo mismo de otro dado, pero cuyo borde tiene una longitud bien distinta.

En el presente trabajo utilizaremos reiteradamente un resultado cuya demostración no incluiremos, que puede encontrarse en [20]. Para enunciarlo necesitaremos algunas definiciones previas.

Definición 1.7. Definimos la función $\lambda \mapsto \psi_\lambda(A)$, para $\lambda > 0$

$$\psi_\lambda(A) = (A \ominus \lambda B) \oplus \lambda B,$$

$$\psi_{-\lambda}(A) = (A \oplus \lambda B) \ominus \lambda B.$$

La función ψ_λ se llama *granulometría* de A.

Definición 1.8. Decimos que un conjunto compacto $A \subset \mathbb{R}^d$ pertenece al *modelo regular de Serra* si

$$A = (A \oplus \varepsilon B) \ominus \varepsilon B = (A \ominus \varepsilon B) \oplus \varepsilon B \quad \text{para algún } \varepsilon > 0.$$

Definición 1.9. Decimos que un conjunto $A \subset \mathbb{R}^d$ es *r-convexo* ($r > 0$) si $A = C_r(A)$, siendo

$$C_r(A) = \bigcap_{\{\mathring{B}(x,r):\mathring{B}(x,r)\cap A=\emptyset\}} \left(\mathring{B}(x,r)\right)^c.$$

Definición 1.10. Diremos que una bola de radio $r > 0$ rueda libremente por el complemento de S si para todo $x \in \partial S$ existe $c \in \mathbb{R}^d$ verificando

$$x \in B(c,r) \subset \overline{S^c}.$$

De forma análoga se define el concepto de rodamiento libre dentro de S. Más adelante veremos algunas propiedades equivalentes a la de rodamiento libre.

Teorema 1.11. *Sea $S \neq \emptyset$ un subconjunto compacto de $\mathbb{R}^d$ y $r_0 > 0$. Entonces son equivalentes:*

(i) $\psi_\lambda(S) = S$ *para* $\lambda \in (-r_0, r_0]$;

(ii) S *y* $\overline{S^c}$ *son r_0-convexos y* $int(S_i) \neq \emptyset$ *para cada componente conexa por caminos* $S_i \subset S$;

(iii) una bola de radio r rueda libremente dentro de cada componente conexa por caminos de S y $\overline{S^c}$ para todo $0 \leq r \leq r_0$;

(iv) ∂S *es una d-1 subvariedad de* $\mathbb{R}^d$ *de clase* C^1 *cuyo vector normal saliente* $n(s)$ *en* $s \in \partial S$ *satisface la siguiente condición de Lipschitz:*

$$\|n(s) - n(t)\| \leq \frac{1}{r_0}\|s - t\| \quad \textit{para todo} \quad s, t \in \partial S.$$

Más aún, para algun $r_0 > 0$*, las condiciones anteriores son equivalentes a*

(v) S *pertenece al modelo regular de Serra.*

1.2.2. Distancia en Medida

Una segunda noción de distancia entre conjuntos es la distancia en medida. Para nuestros propósitos es suficiente considerar el espacio de medida $(\mathbb{R}^d, \mathcal{B}, \mu)$ donde $\mathcal{B}$ denota la sigma álgebra de Borel en $\mathbb{R}^d$ y μ la medida de Lebesgue.

Definición 1.12. Sean A y C en $\mathcal{B}$. La distancia en medida entre A y C se define como

$$d_\mu(A, C) = \mu(A \triangle C),$$

donde $A \triangle C$ denota la diferencia simétrica entre A y C, esto es

$$A \triangle C = (A \setminus C) \cup (C \setminus A).$$

Observación 1.13. *Esta noción de distancia esta íntimamente relacionada con la distancia* $L_1(\mu)$ *entre funciones. De hecho, la distancia en medida es la distancia en el espacio métrico* $L_1(\mu)$ *entre las funciones indicadoras de cada conjunto esto es:*

$$d_\mu(A, C) = \int_{\mathbb{R}^d} |\mathbb{I}_A - \mathbb{I}_C| d\mu.$$

Es fácil ver que el hecho de que dos conjuntos estén próximos en distancia en medida no implica que lo estén en distancia de Hausdorff y recíprocamente, que lo estén en distancia de Hausdorff tampoco implica que lo estén en medida.

1.3. Estimación de un soporte sin restricciones de forma

Cuando estamos ante un conjunto S del cual no sabemos nada respecto de su forma debemos definir estimadores universales, que por lo tanto van a tener peores velocidades de convergencia que cuando tenemos información adicional sobre el conjunto. En (1980) Devroye y Wise,ver [9], propusieron un estimador muy intuitivo y general. Dada una sucesión de vectores aletorios d-dimensionales, $X_1, \ldots, X_n$, independientes y con distribución común P_X, la cual tiene por soporte el conjunto S, se define el estimador de S

$$\hat{S}_n = \bigcup_{i=1}^{n} B(X_i, \varepsilon_n).$$

Este estimador no es más que una versión suavizada de la muestra. Su sencillez hace que su estructura geométrica, al contrario de otros estimadores más sofisticados, sea muy fácil de analizar (incluso en dimensión arbitraria). Más adelante veremos que

bajo algunas restricciones de forma es posible demostrar que $\partial\hat{S}_n$ tiende en distancia de Hausdorff a ∂S y utilizaremos, en la sección de análisis de cluster, el estimador $\hat{S}_n$ para aproximar el número de componentes conexas de S.
Devroye y Wise (1980) utilizaron este estimador para detectar el comportamiento anormal de un sistema. Diremos que una observación Z es *anormal* si $Z \notin \hat{S}_n$. La probabilidad de cometer un error con este mecanismo, es decir decidir que una observación no está en S cuando realmente sí lo está o bien decidir que está en S cuando en realidad no lo ésta es

$$d_\mu(S, \hat{S}_n).$$

En el artículo citado se prueba, entre otras cosas, que el procedimiento anterior es consistente en probabilidad

$$d_\mu(S, \hat{S}_n) \xrightarrow{P} 0,$$

siempre que la restricción a S de la medida de probabilidad, μ, de la nueva observación Z, sea absolutamente continua respecto a la distribución de probabilidad de la muestra P_X, y el parámetro ε_n verifique

$$\varepsilon_n \to 0 \quad n\varepsilon_n^d \to \infty.$$

Este resultado nos asegura que, con carácter casi universal, el estimador de Devroye-Wise es consistente en medida. Nótese que las condiciones sobre ε_n son análogas a las que se imponen al parámetro h_n de un estimador tipo núcleo de la densidad para obtener resultados de consistencia en L_1. La consistencia del estimador de Devroye-Wise en distancia de Hausdorff es trivial dada su similitud con la propia muestra, bastaría tomar $\varepsilon_n \to 0$ para que

$$d_H(S, \hat{S}_n) \to 0 \quad c.s.$$

para cualquier conjunto compacto no vacío de $\mathbb{R}^d$.

1.4. Estimación de la frontera

Veamos ahora algunos resultados referentes a la estimación de un conjunto particular, ∂S. Este conjunto es de interés en varias áreas, por ejemplo en economía, en el sector de análisis de productividad, ∂S es el gráfico de una cierta función $g(x)$, y representa la máxima producción alcanzable por una empresa cuando el *imput* es x. S es entonces el hipografo de g, por lo tanto podríamos evaluar la productividad de una empresa con *imput* x_0 que produzca y_0 como $g(x_0) - y_0$.

Una pregunta que surge de forma natural es si el estimador universal definido por Devroye y Wise verifica que

$$d_H(\partial\hat{S}_n, \partial S) \to 0,$$

en probabilidad o c.s.. Es claro que el sólo hecho de pedir que $\varepsilon_n \to 0$ no garantiza esa convergencia ya que si ε_n tiende *muy rápido* a 0, es de esperar que nuestro estimador tenga muchos *agujeros* y no se verifique la convergencia antes mencionada. En [6] se hallan las condiciones que tiene que cumplir ε_n y S para que sea cierta. Las siguientes subsecciones serán resultados extraidos de ahí y de la tesis doctoral de Alberto Rodríguez Casal [18].

1.4.1. Una condición necesaria

Antes de ver las condiciones suficientes para que se de la convergencia de las fronteras para el estimador de Devroye y Wise, veamos una condición necesaria. El siguiente resultado se encuentra en [18].

Proposición 1.14. *Sea $S \subset \mathbb{R}^d$ compacto. Sean $X_1, X_2, \ldots, X_n, \ldots$ observaciones independientes de una distribución con soporte S. Sea $\hat{S}_n$ el estimador de Devroye y Wise. Supongamos que*

$$d_H\left(\partial S, \partial\hat{S}_n\right) \to 0, \quad c.s.. \tag{1.1}$$

Entonces

$$\varepsilon_n \to 0,$$

y, por lo tanto

$$d_H\left(S, \hat{S}_n\right) \to 0, \quad c.s.$$

Demostración. Por reducción al absurdo. Supongamos que el resultado no fuese cierto. En este caso, con probabilidad positiva, existiría $\varepsilon > 0$, y una subsucesión de ε_n, que denotaremos por simplicidad ε_n, tal que

$$\varepsilon_n \geq \varepsilon, \quad \forall n.$$

Ahora bien, para cualquier $y_n \in \partial\hat{S}_n$ se verifica que

$$\min_{i=1,\ldots,n} \|y_n - X_i\| = \varepsilon_n \geq \varepsilon.$$

Sea

$$s_n = P_S(y_n),$$

una proyección de y_n sobre el conjunto S (podría no ser única). Claramente

$$\varepsilon \leq \varepsilon_n = \min_{i=1,\ldots,n} \|y_n - X_i\| \leq \|y_n - s_n\| + \min_{i=1,\ldots,n} \|s_n - X_i\|. \tag{1.2}$$

Como S es compacto, existirá una subsucesión de $\{s_n\}$ convergente que (por simplicidad en la notación) seguiremos indexando en n. En este caso podemos escribir.

$$s_n \to s \in S.$$

Como

$$\min_{i=1,\ldots,n} \|s - X_i\| \to 0, \quad c.s.,$$

y

$$\min_{i=1,\ldots,n} \|s_n - X_i\| \leq \|s_n - s\| + \min_{i=1,\ldots,n} \|s - X_i\|,$$

entonces

$$\min_{i=1,\ldots,n} \|s_n - X_i\| \to 0, \quad c.s.,$$

y por lo tanto, utilizando 1.2 obtenemos que, con probabilidad positiva, para n suficientemente grande, se verifica

$$\|y_n - s_n\| \geq \frac{\varepsilon}{2},$$

en cuyo caso

$$B\left(y_n, \frac{\varepsilon}{4}\right) \subset S^c \Rightarrow d(y_n, \partial S) \geq \frac{\varepsilon}{4} \Rightarrow d_H\left(\partial S, \partial \hat{S}_n\right) \geq \frac{\varepsilon}{4},$$

lo cual contradice 1.1.
Para demostrar que si $\varepsilon_n \to 0$ casi seguro entonces, con probabilidad uno $d_H(S, \hat{S}_n)$ converge a cero nótese que se verifica

$$d_H(S, \hat{S}_n) \leq \max\left\{\varepsilon_n, d_H(\aleph_n, S)\right\}.$$

De la compacidad de S se sigue fácilmente que $d_H(\aleph_n, S) \to 0 \quad c.s.$ si $\varepsilon_n \to 0 \quad c.s.$ □

1.4.2. Estimación consistente de la frontera

Veamos una condición suficiente y universal, es decir sin restricciones sobre S que garantice la convergencia en distancia de Hausdorff, casi segura, de $\partial \hat{S}_n$ a ∂S.

Teorema 1.15. *Sea $S \subset \mathbb{R}^d$ compacto. Sean $X_1, X_2, \ldots, X_n, \ldots$ observaciones de una distribución, P_X, con soporte S. Sea $\hat{S}_n$ el estimador de Devroye y Wise. Supongamos que*

(i) $P\left(S \subset \hat{S}_n, \text{ a partir de un cierto } n\right) = 1,$

(ii) $\varepsilon_n \to 0, \quad c.s.$

Entonces

$$d_H(\partial S, \partial \hat{S}_n) \to 0, \quad c.s.$$

Demostración. Supongamos por absurdo que no se cumple la tesis, luego, tenemos dos casos:

Caso 1. Con probabilidad positiva, existe un $\varepsilon > 0$ y una sucesión $\{y_{n_k}\}$, que denotaremos $\{y_n\}$ por simplicidad, $y_n \in \hat{S}_n$ y tal que

$$d(y_n, \partial S) = \inf_{y \in \partial S} \|y_n - y\| > \varepsilon \quad \text{para todo } n. \tag{1.3}$$

Es claro que $y_n \notin \partial S$ y que $y_n \notin int(S)$ (ya que $int(S) \subset int(\hat{S}_n)$), por lo tanto $y_n \in S^c$. Como $y_n \in \hat{S}_n$ y $X_i \in S$ c.s. entonces $B(y_n, \varepsilon) \cap S \neq \emptyset$. Esto, junto con $y_n \in S^c$ implica (ya que $B(y_n, \varepsilon_n)$ es conexa) que $B(y_n, \varepsilon_n) \cap \partial S \neq \emptyset$. Luego $d(y_n, \partial S) \leq \varepsilon_n$, pero, como por hipótesis $\varepsilon_n \longrightarrow 0$ esto último contradice 1.3.

Caso 2. Con probabilidad positiva existe $\varepsilon > 0$ y una subsucesión, que denotaremos $\{y_n\}$, con $y_n \in \partial S$ tal que

$$d(y_n, \partial \hat{S}_n) = \inf_{y \in \partial \hat{S}_n} \|y_n - y\| > 2\varepsilon \quad \text{para todo } n.$$

Como $\{y_n\} \subset \partial S$ y ∂S es compacto, existe una subsucesión convergente, (que denotaremos $\{y_n\}$), tal que $y_n \longrightarrow y \in \partial S$. Por la desigualdad triangular, esto implica que, con probabilidad positiva se cumple que $d(y, \partial\hat{S}_n) > \varepsilon$ a partir de un cierto n. Entonces:

$$B(y,\varepsilon) \cap \partial\hat{S}_n = \emptyset \quad \text{a partir de un cierto } n, c.s.. \tag{1.4}$$

Como los sucesos $I_k = \{X_k \in B(y,\varepsilon)\}$ son independientes y $P(I_k) = P_X\big(B(y,\varepsilon)\big) > 0$, por el Lema de Borel-Cantelli tenemos que , con probabilidad 1 existe una subsucesión $X_{n_k} \in B(y,\varepsilon)$, por lo tanto $B(y,\varepsilon) \cap \hat{S}_{n_k} \neq \emptyset$. Como $B(y,\varepsilon)$ es conexa, por 1.4. tenemos que $B(y,\varepsilon) \subset int(\hat{S}_{n_k})$. Como $y \in \partial S$ existe $y_0 \in S^c$ tal que $\|y - y_0\| < \varepsilon/2$. Más aun, como S^c es abierto, existe un $r \in (0, \varepsilon/2)$ tal que $B(y_0, r) \subset S^c$ y $B(y_0, r) \subset B(y,\varepsilon) \subset int(\hat{S}_{n_k})$. Para k suficientemente grande, tal que $\varepsilon_{n_k} < r$ implica que $y_0 \notin \hat{S}_{n_k}$. Lo cual contradice que $B(y,r) \subset int(\hat{S}_{n_k})$.

En ambos casos llegamos a una contradicción. Observemos que la hipótesis $S \subset \hat{S}_n$ sólo se uso en 1. □

Observación 1.16. *Es claro que en la práctica el problema reside en elegir la sucesión de parámetros $\{\varepsilon_n\}$ que garanticen la condición (i). La sucesión que verifica esta condición y que comete menor sobresuavización es, obviamente $\varepsilon_n^0 = d_H(\aleph_n, S)$, siendo $\aleph_n = \{X_1, \ldots, X_n\}$.*

1.4.3. Restricciones de forma

En esta sección vamos a precisar las condiciones que se le deben pedir a S que nos permita obtener el orden de convergencia de ε_n^0 de modo de tener un criterio para elegir ε_n.

Definición 1.17. Se dice que un conjunto de Borel $S \subset \mathbb{R}^d$ es estándar con respecto a una medida de Borel μ si existe un $\lambda > 0$ y $\delta > 0$ verificando

$$\mu\big(B(x,\varepsilon) \cap S\big) \geq \delta\mu_L\big(B(x,\varepsilon)\big), \quad \forall x \in S, \quad 0 < \varepsilon \leq \lambda,$$

donde μ_L es la medida de Lebesgue d-dimensional.

Esta condición evita que el conjunto S tenga *picos infinitamente agudos*, lo cual intuitivamente produciría que en ése pico la probabilidad de tener un X de la muestra sea muy baja, y se tienda a subestimar el borde. Se puede demostrar la siguiente observación

Observación 1.18. *Si S es un conjunto no vacío tal que $\overline{S^c}$ es r-convexo y $S = \overline{int(S)}$. Entonces S es estándar con respecto a la medida μ_L restringida a S.*

Definición 1.19. Diremos que un conjunto compacto $S \subset \mathbb{R}^d$ es parcialmente expandible si existen constantes $r > 0$ y $C(S) \geq 1$ verificando

$$d_H\big(\partial S, \partial(S \oplus \varepsilon B)\big) \leq C(S)\varepsilon, \quad 0 \leq \varepsilon < r.$$

Si la condición anterior se verifica para todo $\varepsilon > 0$ diremos que el conjunto S es expandible.

Esta restricción de forma impide que el conjunto S se *enrosque* sobre sí mismo infinitamente o que tenga entrantes demasiado agudos. Los conjuntos parcialmente expandibles más regulares son aquellos para los cuales $r_0 = \infty$ y $C(S) = 1$. Dichos conjuntos son los conjuntos convexos.

Proposición 1.20. *$S \subset \mathbb{R}^d$ es parcialmente expandible con constante $C(S) = 1$ si y sólo si para algún $r > 0$ una bola de radio r rueda libremente por el complemento de S.*

Demostración. Probemos en primer lugar el siguiente resultado

$$\left.\begin{array}{c} x \in B(c,r) \subset \overline{S^c} \\ x \in \partial S \end{array}\right\} \Rightarrow d\big(x, \partial(S \oplus \varepsilon B)\big) \leq \varepsilon, \quad 0 \leq \varepsilon \leq r. \tag{1.5}$$

En efecto, sea

$$\{y\} = \partial B(x,\varepsilon) \cap [x,c],$$

donde $[x,c]$ denota el segmento que une los puntos x y c. Se verifica

$$x \in B(y,\varepsilon) \subset B(c,r) \subset \overline{S^c},$$

y por lo tanto, $d(y,S) = \varepsilon$. Sea $\{y_n\} = \partial B(x,\varepsilon_n) \cap [x,c]$, con $\varepsilon_n \to \varepsilon$ y $\varepsilon < \varepsilon_n < r$. Tendremos que

$$d(y_n, S) = \varepsilon_n > \varepsilon,$$

y, por lo tanto

$$y_n \in (S \oplus \varepsilon B)^c,$$

y, como $y_n \to y$, obtenemos que $y \in \overline{(S \oplus \varepsilon B)^c}$. Entonces $y \in \partial(S \oplus \varepsilon B)$ con $\|x - y\| = \varepsilon$.
Esto concluye la prueba de 1.5.

Ahora estamos en condiciones de probar la proposición. En primer lugar, si una bola de radio $r > 0$ rueda libremente por el complemento de S entonces, por 1.5, tendremos que

$$\forall x \in \partial S, \quad d\big(x, \partial(S \oplus \varepsilon B)\big) \leq \varepsilon, \quad 0 \leq \varepsilon < r.$$

Así

$$\sup_{x \in \partial S} d\big(x, \partial(S \oplus \varepsilon B)\big) \leq \varepsilon.$$

Por lo tanto, si $x \in \partial(S \oplus \varepsilon B)$ entonces $d(x,S) = \varepsilon$ y, en consecuencia,

$$\sup_{x \in \partial(S \oplus \varepsilon B)} d(x, \partial S) = \varepsilon.$$

Así

$$d_H\big(\partial S, \partial(S \oplus \varepsilon B)\big) = \max\left\{ \sup_{x \in \partial(S \oplus \varepsilon B)} d(x, \partial S),\ \sup_{x \in \partial S} d\big(x, \partial(S \oplus \varepsilon B)\big) \right\} = \varepsilon, \quad 0 \leq \varepsilon < r,$$

es decir, S es parcialmente expandible con constante $C(S) = 1$.

Recíprocamente si S es parcialmente expandible con constante $C(S) = 1$ entonces existe $r > 0$ tal que

$$d_H\big(\partial S, \partial(S \oplus \varepsilon B)\big) \leq \varepsilon, \quad 0 \leq \varepsilon < r.$$

Fijemos $0 < r_0 < r$ arbitrario. Si $x \in \partial S$ se verifica que $B(x, r_0) \cap \partial(S \oplus r_0 B) \neq \emptyset$. Sea $y \in B(x, r_0) \cap \partial(S \oplus r_0 B)$. Como $d(y, S) = r_0$ entonces

$$x \in B(y, r_0) \subset \overline{S^c},$$

y, por lo tanto, una bola de radio r_0 rueda libremente por el complemento de S. □

1.4.4. Tasas de convergencia

Teorema 1.21. *Sea $S \subset \mathbb{R}^d$ un conjunto compacto, parcialmente expandible con constante $C(S)$ para algún $r > 0$. Sean $X_1, \ldots, X_n, \ldots$ observaciones de una distribución con soporte S. Si la sucesión de estimadores*

$$\hat{S}_n = \bigcup_{i=1}^{n} B(X_i, \varepsilon_n)$$

verifica

(i) $\varepsilon_n \to 0$, *c.s.*

(ii) $P\left(S \subset \hat{S}_n,\ \text{a partir de un cierto } n\right) = 1.$

Entonces, con probabilidad uno, existe n_0, tal que, si $n \geq n_0$

$$d_H(\partial S, \partial \hat{S}_n) \leq C(S)\varepsilon_n. \tag{1.6}$$

Demostración. Por las hipótesis (i) y (ii) se tiene que, con probabilidad uno, existe n_0 tal que para $n \geq n_0$

$$\aleph_n \subset S \subset \hat{S}_n, \qquad 0 \leq \varepsilon_n < r.$$

Recordemos que

$$d_H(\partial S, \partial \hat{S}_n) = \max\left\{ \sup_{x \in \partial \hat{S}_n} d(x, \partial S),\ \sup_{x \in \partial S} d(x, \hat{S}_n) \right\}.$$

A continuación probaremos que, si $n \geq n_0$, se verifica

$$\sup_{x \in \partial \hat{S}_n} d(x, \partial S) \leq \varepsilon_n, \tag{1.7}$$

$$\sup_{x \in \partial S} d(x, \partial \hat{S}_n) \leq C(S)\varepsilon_n \tag{1.8}$$

En efecto, si $x \in \partial \hat{S}_n$ entonces, como $B(x, \varepsilon_n) \cap S \neq \emptyset$, necesariamente

$$B(x, \varepsilon_n) \cap \partial S \neq \emptyset,$$

ya que, en caso contrario tendríamos, por la conexión de $B(x, \varepsilon_n)$, que

$$x \in B(x, \varepsilon_n) \subset int(S) \subset int(\hat{S}_n).$$

Así

$$d(x, \partial S) \leq \varepsilon_n, \quad \forall x \in \partial \hat{S}_n,$$

esto prueba 1.7.

Para probar 1.8 nótese que si para algún $x \in \partial S$ se verificase que

$$d(x, \partial \hat{S}_n) > C(S)\varepsilon_n$$

entonces, como $x \in \partial S \subset \hat{S}_n = \aleph_n \oplus \varepsilon_n B$, la conexión de $B(x, C(S)\varepsilon_n)$ implicaría que

$$B\big(x, C(S)\varepsilon_n\big) \subset int(\aleph_n \oplus \varepsilon_n B) \subset int(S \oplus \varepsilon_n B),$$

y, por tanto, existiría algún $x \in \partial S$ para el cual

$$d\big(x, \partial(S \oplus \varepsilon_n B)\big) > C(S)\varepsilon_n,$$

lo cual contradice la condición de que el conjunto sea parcialmente expandible. Tenemos entonces que

$$\forall x \in \partial S, \quad d(x, \partial \hat{S}_n) \leq C(S)\varepsilon_n,$$

lo cual prueba 1.8.

Teniendo en cuenta que $C(S) \geq 1$, las desigualdades 1.7 y 1.8 prueban el teorema. $\square$

El valor más pequeño de ε_n que garantiza $S \subset \hat{S}_n$ es, claramente, $\varepsilon_n^n = d_H(\aleph_n, S)$. Un estudio del orden de convergencia de esta sucesión nos permitiría, utilizando el teorema anterior, determinar el orden de convergencia de $d_H(\partial S, \partial \hat{S}_n)$. Esto es precisamente lo que se hace en el siguiente teorema, bajo la hipótesis de que S es un conjunto estándar.

Teorema 1.22. *Sea $X_1, X_2, \ldots$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos en $\mathbb{R}^d$ con distribución común P_X. Sea S el soporte de P_X. Supongamos que S es compacto y estándar con respecto a P_X. Se verifica que*

$$\limsup_{n\to\infty} \left(\frac{n}{\log n}\right)^{\frac{1}{d}} d_H(\aleph_n, S) \leq \left(\frac{2}{\delta\omega_d}\right)^{\frac{1}{d}} \quad c.s. \tag{1.9}$$

donde $\aleph_n = \{X_1, X_2, \ldots, X_n\}$, ω_d es el volumen de la bola unidad d-dimensional y δ es la constante de estandaridad.

Demostración. Como, con probabilidad uno, $\aleph_n \subset S$, entonces

$$d_H(\aleph_n, S) = \sup_{x\in S} \min_{i=1,\ldots,n} \|x - X_i\|, \quad c.s.$$

Sea $\Delta > 0$ y S_Δ un conjunto (con cardinal $N(\Delta)$) de centros del conjunto S correspondientes a un cubrimiento minimal de S formado por bolas cerradas de radio Δ. Es decir, estamos considerando una clase de conjuntos $\big\{B(s,\Delta), \quad s \in S_\Delta \subset S\big\}$ verificando

$$S \subset \bigcup_{s\in S_\Delta} B(s, \Delta). \tag{1.10}$$

Para $x \in S, s_0 \in S_\Delta$ arbitrarios

$$\min_{i=1,\ldots,n} \|x - X_i\| \leq \|x - X_j\| \leq \|x - s_0\| + \|s_0 - X_j\|, \quad j = 1, \ldots, n.$$

Así

$$\min_{i=1,\dots,n} \|x - X_i\| \leq \|x - s_0\| + \min_{j=1,\dots,n} \|s_0 - X_j\| \leq \|x - s_0\| + \max_{s \in S_\Delta} \min_{j=1,\dots,n} \|s - X_j\|,$$

y, como $x \in S$, por 1.10 tenemos que

$$\min_{i=1,\dots,n} \|x - X_i\| \leq \Delta + \max_{s \in S_\Delta} \min_{i=1,\dots,n} \|s - X_i\|.$$

Sea $\varepsilon > 0$ arbitrario. De la desigualdad anterior se sigue que

$$P\left(\sup_{x \in S} \min_{i=1,\dots,n} \|x - X_i\| > \varepsilon\right) \leq P\left(\max_{s \in S_\Delta} \min_{i=1,\dots,n} \|s - X_i\| > \varepsilon - \Delta\right).$$

Tomando $\Delta = (1-\nu)\varepsilon, \quad 0 < \nu < 1$, de tiene que

$$P\big(d_H(\aleph_n, S) > \varepsilon\big) \leq P\left(\max_{s \in S_{(1-\nu)\varepsilon}} \min_{i=1,\dots,n} \|s - X_i\| > \nu\varepsilon\right). \tag{1.11}$$

Definamos los sucesos $I_{s,i} = \big\{X_i \in B(s, \nu\varepsilon)\big\}$, con $s \in S_{(1-\nu)\varepsilon}, i = 1, \dots, n$. Tenemos que

$$\begin{aligned} P\left(\max_{s \in S_{(1-\nu)\varepsilon}} \min_{i=1,\dots,n} \|s - X_i\| > \nu\varepsilon\right) &= P\left(\bigcup_{s \in S_{(1-\nu)\varepsilon}} \bigcap_{i=1}^n I_{s,i}^c\right) \\ &\leq \sum_{s \in S_{(1-\nu)\varepsilon}} P\left(\bigcap_{i=1}^n I_{s,i}^c\right) = \sum_{s \in S_{(1-\nu)\varepsilon}} \prod_{i=1}^n P(I_{s,i}^c) = \sum_{s \in S_{(1-\nu)\varepsilon}} \prod_{i=1}^n \big(1 - P(I_{s,i})\big) \end{aligned}$$

Como S es un conjunto estándar, para ε suficientemente pequeño

$$P(I_{s,i}) = P\big(X_i \in B(s, \nu\varepsilon)\big) = P_X\big(B(s, \nu\varepsilon)\big) \geq \delta\omega_d(\nu\varepsilon)^d.$$

Así

$$\begin{aligned} P&\left(\max_{s \in S_{(1-\nu)\varepsilon}} \min_{i=1,\dots,n} \|s - X_i\| > \nu\varepsilon\right) \\ &\leq \sum_{s \in S_{(1-\nu)\varepsilon}} \big(1 - \delta\omega_d(\nu\varepsilon)^d\big)^n = N\big((1-\nu)\varepsilon\big)\big(1 - \delta\omega_d(\nu\varepsilon)^d\big)^n. \end{aligned}$$

Como $N\big((1-\nu)\varepsilon\big) \leq A\big((1-\nu)\varepsilon\big)^{-d}$, donde A es una constante que depende de la dimensión d y del diámetro del conjunto S, se tiene

$$P\left(\max_{s \in S_{(1-\nu)\varepsilon}} \min_{i=1,\dots,n} \|s - X_i\| > \nu\varepsilon\right) \leq A\big((1-\nu)\varepsilon\big)^{-d}\big(1 - \delta\omega_d(\nu\varepsilon)^d\big)^n.$$

Aplicando la desigualdad $(1-x)^n \leq e^{-nx}$, para $0 \leq x \leq 1$, obtenemos que, para ε suficientemente pequeño,

$$P\left(\max_{s \in S_{(1-\nu)\varepsilon}} \min_{i=1,\dots,n} \|s - X_i\| > \nu\varepsilon\right) \leq A\big((1-\nu)\varepsilon\big)^{-d} e^{-n\delta\omega_d(\nu\varepsilon)^d}.$$

Utilizando 1.11, se tiene

$$P\big(d_H(\aleph_n, S) > \varepsilon\big) \leq A\big((1-\nu)\varepsilon\big)^{-d} e^{-n\delta\omega_d(\nu\varepsilon)^d}. \tag{1.12}$$

Sea $l > \left(\frac{2}{\delta\omega_d}\right)^{\frac{1}{d}}$. Para n suficientemente grande

$$\begin{aligned} P\left(\left(\frac{n}{\log n}\right)^{\frac{1}{d}} d_H(\aleph_n, S) > l\right) &\leq A\big((1-\nu)l\big)^{-d} \frac{n}{\log n} \exp\big\{-\delta\omega_d\nu^d l^d \log n\big\} \\ &= A\big((1-\nu)l\big)^{-d} \frac{n^{1-\delta\omega_d\nu^d l^d}}{\log n}. \end{aligned}$$

Para que converja la serie $\sum_n \frac{n^{1-\delta\omega_d\nu^d l^d}}{\log n}$ es suficiente que $1-\delta\omega_d\nu^d l^d < -1$, para lo cual basta tomar

$$\left(\frac{2}{\delta\omega_d l^d}\right) < \nu < 1,$$

que es siempre posible, ya que, por hipótesis,

$$l > \left(\frac{2}{\delta\omega_d}\right)^{\frac{1}{d}}.$$

Por lo tanto

$$\sum_n P\left(\left(\frac{n}{\log n}\right)^{\frac{1}{d}} d_H(\aleph_n, S) > l\right) < \infty,$$

y, por el primer Lema de Borel-Cantelli

$$\limsup_{n\to\infty} \left(\frac{n}{\log n}\right)^{\frac{1}{d}} d_H(\aleph_n, S) \leq l, \quad c.s., \quad l > \left(\frac{2}{\delta\omega_d}\right)^{\frac{1}{d}},$$

lo cual concluye la demostración del teorema. □

Una consecuencia inmediata de éste teorema es el siguiente

Teorema 1.23. *Sea $X_1, X_2, \dots$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos en $\mathbb{R}^d$ con distribución común P_X. Sea S el soporte de P_X. Supongamos que S es un conjunto compacto parcialmente expandible con constante $C(S)$, para algún $r > 0$, y estándar con respecto a P_X. Consideremos el estimador de Devroye y Wise*

$$\hat{S}_n(\varepsilon_n) = \bigcup_{i=1}^{n} B(X_i, \varepsilon_n),$$

donde

$$\varepsilon_n = C\left(\frac{\log n}{n}\right)^{\frac{1}{d}}, \quad C > \left(\frac{2}{\delta\omega_d}\right)^{\frac{1}{d}}.$$

Entonces, con probabilidad uno, existe n_0 verificando

$$d_H(\partial S, \partial \hat{S}_n) \leq C(S)\varepsilon_n, \quad n \geq n_0,$$

$$d_H(S, \hat{S}_n) \leq \varepsilon_n, \quad n \leq n_0.$$

Demostración. Como $C > \left(\frac{2}{\delta\omega_d}\right)^{\frac{1}{d}}$, por 1.9, existirá, con probabilidad uno, n_1 tal que si $n \geq n_1$, entonces

$$d_H(\aleph_n, S) \leq C\left(\frac{\log n}{n}\right)^{\frac{1}{d}} = \varepsilon_n$$

y por lo tanto

$$S \subset \hat{S}_n.$$

Como, evidentemente, $\varepsilon_n \to 0$, estamos en condiciones de aplicar el Teorema 1.21 y concluir que, con probabilidad uno, existe un n_0 verificando

$$d_H(\partial S, \partial \hat{S}_n) \leq C(S)\varepsilon_n, \quad n \geq n_0.$$

Finalmente, la desigualdad $d_H(S, \hat{S}_n) \leq \varepsilon_n$, es trivial, ya que, $S \subset \hat{S}_n$ y $\aleph_n \subset S$. □

Veamos ahora que este resultado no puede mejorarse usando el estimador de Devroye y Wise, esto es:

Teorema 1.24. *Si $d \geq 2$ existen conjuntos en las condiciones del Teorema anterior, tales que para cualquier sucesión de radios $\{\varepsilon_n\}$*

$$\liminf_{n\to\infty} \left(\frac{n}{\log n}\right)^{\frac{1}{d}} d_H(S, \hat{S}_n) \geq \left(\frac{1}{2^d\omega_d}\right)^{\frac{1}{d}}, \quad c.s.$$

Demostración. Para demostrar el teorema serán necesarios dos lemas

Lema 1.25. *Sea $S = [0,1]^d$. Si denotamos mediante $\hat{S}_n$ el estimador de Devroye y Wise asociado a una muestra $X_1, \ldots, X_n \in S$ entonces, para cualquier $\varepsilon_n > 0$*

$$d_H(S, \hat{S}_n) = \max\left\{\varepsilon_n - \varepsilon_{min}, d_H(\aleph_n, S) - \varepsilon_n\right\}, \tag{1.13}$$

donde

$$\varepsilon_{min} = \min_{i=1,\ldots,n} \min_{x\in\partial S} \|X_i - x\|.$$

Demostración. Distinguiremos tres situaciones

1 $\varepsilon_n \leq \varepsilon_{min}$.
Entonces

$$\hat{S}_n \subset S,$$

y, por lo tanto

$$d_H(S, \hat{S}_n) = d_H(\aleph_n, S) - \varepsilon_n = \max\left\{\varepsilon_n - \varepsilon_{min}, d_H(\aleph_n, S) - \varepsilon_n\right\}.$$

2 $\varepsilon \geq d_H(\aleph_n, S)$.
Entonces
$$S \subset \hat{S}_n,$$
y por lo tanto, por tratarse de $S = [0,1]^d$,
$$d_H(S, \hat{S}_n) = \varepsilon_n - \varepsilon_{min} = \max\left\{\varepsilon_n - \varepsilon_{min}, d_H(\aleph_n, S) - \varepsilon_n\right\}.$$

3 $\varepsilon_{min} < \varepsilon < d_H(\aleph_n, S)$.
Recordemos que
$$d_H(S, \hat{S}_n) = \min\left\{\varepsilon > 0 : S \subset B(\hat{S}_n, \varepsilon), \quad \hat{S}_n \subset B(S, \varepsilon)\right\}.$$

Ahora bien
$$S \subset B(\hat{S}_n, \varepsilon) \Rightarrow S \subset (\aleph_n \oplus \varepsilon_n B) \oplus \varepsilon B = \aleph_n \oplus (\varepsilon_n + \varepsilon)B$$
$$\Rightarrow d_H(\aleph_n, S) \leq \varepsilon_n + \varepsilon \Rightarrow \varepsilon \geq \tilde{\varepsilon}_0 = d_H(\aleph_n, S) - \varepsilon_n.$$

Además
$$S \subset B(\hat{S}_n, \tilde{\varepsilon}_0).$$
Por otra parte, por tratarse de $S = [0,1]^d$,
$$\hat{S}_n = \aleph_n \oplus \varepsilon_n B \subset B(S, \varepsilon) \Rightarrow \varepsilon \geq \tilde{\varepsilon}_1 = \varepsilon_n - \varepsilon_{min}.$$

Además
$$\hat{S}_n \subset B(S, \tilde{\varepsilon}_1).$$
Por lo tanto
$$d_H(S, \hat{S}_n) = \max\{\tilde{\varepsilon}_0, \tilde{\varepsilon}_1\} = \max\left\{\varepsilon_n - \varepsilon_{min}, d_H(\aleph_n, S) - \varepsilon_n\right\}.$$

□

El lema anterior nos permite obtener el radio ε_n^* que minimiza $d_H(S, \hat{S}_n)$, que en este caso será
$$\varepsilon_n^* = \frac{d_H(\aleph_n, S) - \varepsilon_{min}}{2}.$$
Así, para cualquier $\varepsilon_n > 0$ se verifica
$$d_H(S, \hat{S}_n) \geq \frac{d_H(\aleph_n, S) - \varepsilon_{min}}{2}. \tag{1.14}$$

En el lema anterior los razonamientos no dependen de un modo crucial de que $S = [0,1]^d$. De hecho, es sencillo demostrar, empleando técnicas similares a las utilizadas en el lema anterior, que con probabilidad 1, para n suficientemente grande, es válida para cualquier conjunto S para el cual una bola de radio $r > 0$ ruede libremente por el complemento de S.

A la vista de este resultado debemos analizar el orden de convergencia de ε_{min} para conseguir acotar inferiormente $d_H(S, \hat{S}_n)$. Eso es lo que se realiza en el siguiente lema.

Lema 1.26. *Sean $X_1, X_2, \ldots$ observaciones independientes con distribución uniforme en $S = [0,1]^d$. Entonces se verifica*

$$\limsup_{n\to\infty} \frac{n}{\log n}\varepsilon_{min} \leq \frac{1}{2d}, \quad c.s.$$

Demostración. Sea $\varepsilon > \frac{1}{2d}$. Para n suficientemente grande $\varepsilon \dfrac{\log n}{n} < \dfrac{1}{2}$ y entonces

$$\left\{\varepsilon_{min} > \varepsilon\frac{\log n}{n}\right\} = \left\{\aleph_n \subset \left[\varepsilon\frac{\log n}{n}, 1-\varepsilon\frac{\log n}{n}\right]^d\right\}.$$

Por lo tanto

$$P\left(\frac{n}{\log n}\varepsilon_{min} > \varepsilon\right) = P\left(\varepsilon_{min} > \varepsilon\frac{\log n}{n}\right) = \left(1-2\varepsilon\frac{\log n}{n}\right)^{dn} \leq e^{-2\varepsilon\frac{\log n}{n}dn} = e^{-2\varepsilon d\log n}.$$

Como $\varepsilon > \frac{1}{2d}$, entonces

$$\sum_n P\left(\frac{n}{\log n}\varepsilon_{min} > \varepsilon\right) < \infty,$$

y, entonces, aplicando el Lema de Borel-Cantell se sigue la tesis del lema □

Veamos ahora la demostración del teorema. Tomemos $S = [0,1]^d$. Para demostrar este teorema usaremos un resultado de Janson de 1987, relativo al radio máximo de la bola contenida en S que no incluye ningun punto de la muestra. Formalmente, se define como

$$\Delta_n = \sup\{r : \quad \exists x \text{ con } B(x,r) \subset S \setminus \aleph_n\}.$$

De los resultados del citado artículo, se puede concluir que

$$\lim_{n\to\infty}\left(\frac{n}{\log n}\right)^{\frac{1}{d}}\Delta_n = \left(\frac{1}{\omega_d}\right)^{\frac{1}{d}}, \quad c.s. \tag{1.15}$$

A partir de la definición de Δ_n es fácil comprobar que

$$d_H(\aleph_n, S) \geq \Delta_n. \tag{1.16}$$

Teniendo en cuenta la ecuación 1.14 se obtiene que, para cualquier ε_n,

$$\left(\frac{n}{\log n}\right)^{\frac{1}{d}} d_H(S,\hat{S}_n) \geq \frac{1}{2}\left\{\left(\frac{n}{\log n}\right)^{\frac{1}{d}}\Delta_n - \left(\frac{n}{\log n}\right)^{\frac{1}{d}}\varepsilon_{min}\right\}. \tag{1.17}$$

Aplicando el Lema 1.26 y teniendo en cuenta que $d \geq 2$ se tiene que

$$\left(\frac{n}{\log n}\right)^{\frac{1}{d}}\varepsilon_{min} = \left(\frac{\log n}{n}\right)^{\frac{d-1}{d}}\left(\frac{n}{\log n}\right)\varepsilon_{min} \to 0, \quad c.s.,$$

que implica, por 1.15 y 1.16, que

$$\liminf_{n\to\infty}\left(\frac{n}{\log n}\right)^{\frac{1}{d}} d_H(S,\hat{S}_n) \geq \frac{1}{2}\left(\frac{1}{\omega_d}\right)^{\frac{1}{d}},$$

lo cual concluye la prueba. □

1.5. Estimación *plug-in* de conjuntos de nivel

En la presente sección abordaremos el tema de la estimación de conjuntos de nivel $\{f \geq c\}$. Veremos el enfoque propuesto por Rodríguez Casal en [18], en el cual la función f no necesariamente es una densidad, y su dominio es un espacio métrico cualquiera. Como allí se menciona, hay ejemplos de interés práctico donde se hace necesario ese enfoque. En la siguiente sección veremos el trabajo de Cuevas y Fraiman [7]. Allí también se proponen estimadores basados en núcleos y se estudian las convergencias en distancia de Hausdorff y en medida, pero se hace el estudio del caso particular en que $c = 0$ es decir, del soporte propiamente dicho de la densidad. Veamos ahora en detalle el primer enfoque propuesto por Rodriguez Casal.

1.5.1. Estimación consistente de la frontera

En esta sección se darán condiciones que permitan garantizar que

$$\lim_{n\to\infty} d_H\big(\partial\{f \geq c\}, \partial\{f_n \geq c\}\big) = 0, \quad c.s., \tag{1.18}$$

donde se supone que f y f_n son funciones definidas en el mismo espacio métrico M, y toman valores reales. Además, mientras que f es una función fija, desconocida, las funciones f_n serán aleatorias y estarán definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Para que tenga sentido calcular la distancia de Hausdorff entre conjuntos compactos debemos tener definida en M una métrica que denotaremos por ρ siendo por tanto (M, ρ) un espacio métrico. Para poder establecer el resultado de convergencia antes mencionado, debemos imponer la siguiente restricción a los espacios métricos considerados.

(M1) Existe $r_0 > 0$ tal que

$$\forall x \in M \quad B(x,r) = \big\{y : \rho(x,y) \leq r\big\}, \quad 0 < r < r_0,$$

es conexa.

Esta condición es claro que se cumple en cualquier ejemplo de interés práctico.

(f0) $\partial\{f \geq c\} \neq \emptyset$.

(f1) f no tiene zonas planas de nivel c. De forma precisa, para todo $x \in M$, con $f(x) = c$, existen sucesiones $\{u_n\}$ y $\{l_n\}$ verificando

$$u_n \to x, \quad f(u_n) > c, \quad y \quad l_n \to x, \quad f(l_n) < c.$$

Esta condición implica, en el caso de que f sea continua, que $\partial\{f \geq c\} = \{f = c\}$. Además evita situaciones de máximos y mínimos locales en el nivel c. Podría darse el caso de que f_n convergiera uniformemente a f y todas sus derivadas también y no se verifique la convergencia en 1.18.

(f2) Existe $c^- < c$ tal que el conjunto $\{f \geq c^-\}$ es compacto. Esta condición, además de garantizar la compacidad de $\{f \geq c\}$ se impone para evitar situaciones de asíntotas horizontales en el valor c.

Teorema 1.27. *Se (M,ρ) un espacio métrico verificando (M1) y $f: M \to \mathbb{R}$, una función continua verificando (f0), (f1), (f2). Sean*

$$f_n : M \to \mathbb{R}, \quad n = 1, 2, \ldots$$

funciones aleatorias definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. Supongamos que, para cada n, con probabilidad uno, f_n es continua, y además

$$\|f - f_n\|_\infty = \sup_{x \in M} |f(x) - f_n(x)| \to 0, \quad c.s.$$

Entonces

$$\lim_{n\to\infty} d_H\big(\partial\{f \geq c\}, \partial\{f_n \geq c\}\big) = 0, \quad c.s.$$

Demostración. Fijemos $\varepsilon > 0$ arbitrario. Debemos demostrar que, con probabilidad uno, existe n_0 tal que si $n \geq n_0$ entonces

$$\begin{aligned} (i) \quad \partial\{f \geq c\} \ &\subset B\big(\partial\{f_n \geq c\}, \varepsilon\big), \\ (ii) \quad \partial\{f_n \geq c\} \ &\subset B\big(\partial\{f \geq c\}, \varepsilon\big). \end{aligned}$$

Veremos entonces que se cumplen las condiciones (i) y (ii).

(i) Como $\partial\{f \geq c\}$ es un subconjunto cerrado del conjunto compacto $\{f \geq c^-\}$, entonces también $\partial\{f \geq c\}$ es un conjunto compacto. Así existe

$$\{x^1, \ldots, x^m\} \subset \partial\{f \geq c\},$$

tal que

$$\partial\{f \geq c\} \subset \bigcup_{i=1}^m B\left(x^i, \frac{\varepsilon}{2}\right). \tag{1.19}$$

Ahora bien, la continuidad de f, implica que $\partial\{f \geq c\} \subset \{f = c\}$. Entonces, por $(f1)$, existen sucesiones $\{u_k^i\}_k$, $\{l_k^i\}_k$ verificando

$$\lim_{k\to\infty} u_k^i = x^i, \quad f(u_k^i) > c; \quad \lim_{k\to\infty} l_k^i = x^i, \quad f(l_k^i) < c, \quad i = 1, \ldots, m.$$

Como $u_k^i \to x^i$ existe k_1^i tal que si $k \geq k_1^i$, se verifica

$$\rho(u_k^i, x^i) \leq \frac{\varepsilon}{4}, \quad i = 1, \ldots, m. \tag{1.20}$$

Sea $r_k^i = \rho(u_k^i, l_k^i) > 0$. Como $\lim_{k\to\infty} r_k^i = 0$ entonces existe k_2^i tal que, si $k \geq k_2^i$, se verifica

$$0 < r_k^i < r_0, \quad r_k^i \leq \frac{\varepsilon}{4}, \quad i = 1, \ldots, m. \tag{1.21}$$

Además como

$$\lim_{n\to\infty} f_n(u_k^i) = f(u_k^i) > c, \quad c.s.; \quad \lim_{n\to\infty} f_n(l_k^i) = f(l_k^i) < c, \quad c.s.,$$

se tiene que, con probabilidad uno, existe n_k^i tal que, si $n \geq n_k^i$, se verifica

$$f_n(u_k^i) > c, \quad f_n(l_k^i) < c, \quad i = 1, \ldots, m. \tag{1.22}$$

Sea $k^i = \max\{k_1^i, k_2^i\}$. Si $n \geq n_{k^i}^i = n^i$ entonces por 1.22,

$$u^i \in B(u^i, r^i) \cap \{f_n \geq c\}, \quad l^i \in B(u^i, r^i) \cap \{f_n < c\},$$

donde $u^i = u_{k^i}^i$, $l^i = l_{k^i}^i$ y $r^i = r_{k^i}^i$. Como $k^i \geq k_2^i$, por 1.21, $r^i < r_0$, y, por lo tanto, según *(M1)*, $B(u^i, r^i)$ es conexa. Así $B(u^i, r^i) \cap \partial\{f_n \geq c\} \neq \emptyset$. Sea $z_n^i \in B(u^i, r^i) \cap \partial\{f_n \geq c\}$. Entonces, utilizando 1.20 y 1.21,

$$\rho(x^i, z_n^i) \leq \rho(x^i, u^i) + \rho(u^i, z_n^i) \leq \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \frac{\varepsilon}{2}, \quad n \geq n^i.$$

Sea ahora $x \in \partial\{f \geq c\}$. Por 1.19 existirá x^i tal que $\rho(x, x^i) \leq \frac{\varepsilon}{2}$. Así

$$d\big(x, \partial\{f_n \geq c\}\big) \leq \rho(x, z_n^i) \leq \rho(x, x^i) + \rho(x^i, z_n^i) \leq \varepsilon, \quad n \geq \max\{n^1, \ldots, n^m\}.$$

Entonces

$$\partial\{f \geq c\} \subset B\big(\partial\{f_n \geq c\}, \varepsilon\big), \quad n \geq \max\{n^1, \ldots, n^m\},$$

lo cual concluiría la prueba de (i).

(ii) Por reducción al absurdo. Supongamos que no se verificase (ii) entonces, con probabilidad positiva, existiría una subsucesión, que denotaremos por $\{x_n\}$, verificando

$$x_n \in \partial\{f_n \geq c\}, \quad d\big(x_n, \partial\{f \geq c\}\big) > \varepsilon.$$

Recordemos que, según *(f2)*, existe $c^- < c$, tal que el conjunto $\{f \geq c^-\}$ es compacto. Sea $\alpha = c - c^- > 0$. Dado que

$$\|f - f_n\|_\infty = \sup_{x \in M} |f_n(x) - f(x)| \to 0, \quad c.s.$$

Entonces, con probabilidad uno, para n suficientemente grande, se verifica

$$\|f - f_n\|_\infty \leq \alpha.$$

Así, para todo $x \in \{f_n \geq c\}$

$$c \leq f_n(x) \leq f(x) + \alpha \Rightarrow f(x) \geq c - \alpha = c^-.$$

Entonces, con probabilidad uno, para n suficientemente grande, se verifica $\{f_n \geq c\} \subset \{f \geq c^-\}$. Como, por hipótesis, $\{f \geq c^-\}$ es compacto, y $x_n \in \partial\{f_n \geq c\} \subset \{f \geq c^-\}$, entonces existirá una subsucesión convergente, que por simplicidad en la notación seguiremos indicando con el índice n,

$$x_n \to x \in \partial\{f \geq c^-\}.$$

Ahora bien como, con probabilidad uno, f_n es continua,

$$x_n \in \partial\{f_n \geq c\} \subset \{f_n = c\}.$$

Así

$$|f(x) - c| = |f(x) - f_n(x_n)| \leq |f(x) - f(x_n)| + |f(x_n) - f_n(x_n)| \leq$$
$$|f(x) - f(x_n)| + \|f - f_n\|_\infty.$$

Por lo tanto, por la continuidad de f y la convergencia uniforme casi segura de f_n a f, se tiene que

$$f(x) = c,$$

y, por *(f1)*, $x \in \partial\{f \geq c\}$. Con lo cual

$$d\big(x_n, \partial\{f \geq c\}\big) \leq \rho(x_n, x) \to 0.$$

Pero, por hipótesis,

$$d\big(x_n, \partial\{f \geq c\}\big) > \varepsilon,$$

obteniendo la contradicción que buscábamos.

□

La convergencia uniforme en todo M puede necesitar, en algunos casos, condiciones relativamente restrictivas sobre la función f. Por ejemplo, en el espacio euclídeo d-dimensional si f es una densidad uniformemente continua entonces existen estimadores f_n que convergen uniformemente de forma casi segura a f, véase por ejemplo, Devroye y Penrod (1986). No obstante cuando la función no cumple esas condiciones de regularidad se puede demostrar resultados análogos a los del teorema anterior, restringiéndonos a un compacto K y pidiendo convergencia uniforme en K. Dichos resultados que aquí omitiremos se pueden encontrar en la tesis doctoral de Rodriguez Casal.

1.5.2. Consistencia en medida

Sean $L(c)$ y $L_n(c)$ los conjuntos de nivel poblacional y estimado respectivamente, es decir,

$$L(c) = \{f \geq c\}, \tag{1.23}$$

$$L_n(c) = \{f_n \geq c\}, \tag{1.24}$$

que en los sucesivo, cuando no exista confusión, denotaremos por L y L_n, la pregunta es entonces si $d_\mu(L_n, L) \to 0$ en probabilidad o c.s. Antes de responder a esta pregunta, vamos a definir un espacio de medida $(M, \mathfrak{F}, \mu)$, además, para que la diferencia simétrica $L \triangle L_n$, sea medible será suficiente que tanto f como f_n sean medibles. Parece claro que si deseamos que L_n aproxime el conjunto L en medida entonces los conjuntos construidos a partir de pequeñas desviaciones de f del valor c deberían tener una medida pequeña. Esto se garantiza con las condiciones

$$\mu\big(\{f = c\}\big) = 0, \tag{1.25}$$

y, para algún $\varepsilon_0 > 0$

$$\mu\big(\{c - \varepsilon_0 \leq f < c + \varepsilon_0\}\big) < \infty. \tag{1.26}$$

Bajo estas condiciones será suficiente que las funciones f_n converjan a f, de forma casi segura, en algún $L_p(\mu)$ tal como se demuestra en el siguiente teorema.

Teorema 1.28. *Sea* $(M, \mathfrak{F}, \mu)$ *un espacio de medida y* $f : M \to \mathbb{R}$, *una función medible verificando las condiciones 1.25, y 1.26. Sean*

$$f_n : M \to \mathbb{R}, \quad n = 1, 2 \ldots,$$

funciones aleatorias definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. *Supongamos que, para cada* n, *con probabilidad uno,* f_n *es medible. Además supongamos que, para algún* $1 \leq p \leq \infty$

$$f, f_n \in L_p(\mu) \quad y \quad \|f - f_n\|_p \to 0, \quad c.s.$$

Entonces

$$d_\mu(L, L_n) \to 0, \quad c.s.$$

Demostración. Sea $\varepsilon > 0$ arbitrario. Para cada $n \in \mathbb{N}$ definamos el conjunto

$$E_{n,\varepsilon} = \big\{x : \ |f(x) - f_n(x)| \leq \varepsilon\big\}.$$

Es claro que, con probabilidad uno, todos los conjuntos de las familias

$$\{E_{n,\varepsilon}, \ n = 1, 2, \ldots\}, \quad \{L \triangle L_n, \ n = 1, 2, \ldots\},$$

son medibles. En este caso

$$\mu(L \triangle L_n) = \mu(L \triangle L_n \cap E_{n,\varepsilon}) + \mu(L \triangle L_n \cap E^c_{n,\varepsilon}).$$

Como

$$L \triangle L_n \cap E_{n,\varepsilon} \subset \{c - \varepsilon \leq f < c + \varepsilon\},$$

entonces

$$\mu(L \triangle L_n) \leq \mu\big(\{c - \varepsilon \leq f < c + \varepsilon\}\big) + \mu(E^c_{n,\varepsilon}). \tag{1.27}$$

Se verifica que

$$\lim_{n\to\infty} \mu(E^c_{n,\varepsilon}) = 0, \quad c.s. \tag{1.28}$$

En efecto, si $p = \infty$, como

$$\mu(E^c_{n,\varepsilon}) > 0 \Leftrightarrow \|f - f_n\|_\infty \geq \varepsilon,$$

entonces, con probabilidad uno, existe n_0 tal que, si $n \geq n_0$,

$$\mu(E^c_{n,\varepsilon}) = 0.$$

Si $1 \leq p < \infty$

$$\mu(E^c_{n,\varepsilon}) = \int \mathbb{I}_{E^c_{n,\varepsilon}} d\mu \leq \frac{1}{\varepsilon^p} \int |f(x) - f_n(x)|^p \mu(dx) \to 0, \quad c.s.$$

Esto concluye la prueba de 1.28. Por tanto por 1.27 y 1.28, obtenemos que, para cualquier $\varepsilon > 0$,

$$\limsup_{n\to\infty} \mu(L \triangle L_n) \leq \mu\big(\{c - \varepsilon \leq f < c + \varepsilon\}\big), \quad c.s.$$

Finalmente, por 1.25 y 1.26,

$$\lim_{\varepsilon\to 0} \mu\big(\{c-\varepsilon \leq f < c+\varepsilon\}\big) = 0.$$

con lo cual

$$\lim_{n\to\infty} \mu(L \triangle L_n) = 0, \quad c.s.$$

que es lo que queriamos demostrar. □

Veamos un corolario para el caso en que f es una densidad.

Corolario 1.29. *Sea $(M, \mathfrak{F}, \mu)$ un espacio de medida donde μ es una medida σ-finita. Sea ν una medida de probabilidad en $(M, \mathfrak{F})$ absolutamente continua respecto a μ donde f es su función de densidad respecto a μ, esto es*

$$\nu(A) = \int_A f d\mu, \quad A \in \mathfrak{F}.$$

Sea $(\Omega, \mathcal{A}, P)$ un espacio de probabilidad y

$$f_n : \Omega \times M \to \mathbb{R}, \quad n = 1, 2, \ldots$$

una sucesión de funciones tales que, para cada n, f_n es una función medible en el espacio de medida producto $(\Omega \times M, \mathcal{A} \times \mathfrak{F}, P \times \mu)$. Supongamos que, para cada n y $\omega \in \Omega$, $f_n(\omega, \cdot)$ es una funcion de densidad con respecto a μ, es decir, existe una medida de probabilidad ν_n en $(M, \mathfrak{F})$ verificando

$$v_n(A) = \int_A f_n(\omega, x)\mu(dx), \quad A \in \mathfrak{F}.$$

Si

$$f_n(x) \to f(x), \quad c.s.(P),$$

salvo, a lo sumo, en un conjunto de medida μ cero y se verifican las condiciones 1.59 y 1.59 entonces

$$d_\mu(L, L_n) \to 0, \quad c.s.(P).$$

Demostración. El Teorema de Scheffé nos asegura que, para una sucesión de densidades fija, es suficiente la convergencia puntual de dicha sucesión a una densidad para garantizar la convergencia en L_1. En el Teorema de Glick (véase, por ejemplo, Devroye y Györfi 1985) se prueba que, para densidades con respecto a la medida de Lebesgue en el espacio euclídeo, la convergencia puntual casi segura es suficiente para garantizar al convergencia en L_1 de forma casi segura. En este corolario lo que se hace es aplicar esas ideas al contexto mas general.

Por hipótesis, salvo un conjunto de medida μ cero se verifica

$$P\big(\{\omega : f_n(\omega, x) \nrightarrow f(x)\}\big) = 0$$

Así

$$\int_\Omega \mathbb{I}_{\{(\omega,x):\ f_n(\omega,x)\nrightarrow f(x)\}} P(d\omega) = 0, \quad c.s(\mu),$$

y, por lo tanto
$$\int_M \int_\Omega \mathbb{I}_{\{(\omega,x):\ f_n(\omega,x)\nrightarrow f(x)\}} P(d\omega)\mu(dx) = 0.$$
Por el Teorema de Fubini
$$\int_M \int_\Omega \mathbb{I}_{\{(\omega,x):\ f_n(\omega,x)\nrightarrow f(x)\}} P(d\omega)\mu(dx) = \int_\Omega \int_M \mathbb{I}_{\{(\omega,x):\ f_n(\omega,x)\nrightarrow f(x)\}} \mu(dx)P(d\omega)$$
Así
$$\int_M \mathbb{I}_{\{(\omega,x):\ f_n(\omega,x)\nrightarrow f(x)\}} \mu(dx), \quad c.s(P).$$
Entonces existe $\Omega' \subset \Omega$ con $P(\Omega') = 1$ para el cual si $\omega \in \Omega'$ se verifica
$$f_n(\omega, x) \to f(x), \quad c.s(\mu),$$
y, por el Lema de Scheffé
$$\int_M |f(x) - f_n(\omega, x)|\mu(dx) \to 0.$$
Por lo tanto estamos en condiciones de aplicar el teorema anterior para el caso $p = 1$, de donde se sigue el corolario. □

1.5.3. Tasas de convergencia

Obtendremos, a partir de imponer condiciones sobre f, tasas de convergencia casi segura para
$$d_H\big(\partial\{f \geq c\}, \partial\{f_n \geq c\}\big).$$
Estas tasas de convergencia serán del orden de
$$\|f - f_n\|_\infty = \sup_{x\in M} |f(x) - f_n(x)|.$$
La condición es

(T) Existe $\gamma > 0$ y $L > 0$ tal que si $|t - c| \leq \gamma$ entonces
$$d_H\big(\{f = c\}, \{f = t\}\big) \leq L|t - c|.$$

Esta condición implica, en particular, la condición *(f1)*. Así, también en este caso, cuando f es continua, tendremos que $\partial\{f \geq c\} = \{f = c\}$. Además es fácil ver que se excluye la posibilidad de que f posea asíntotas horizontales. Cuando $\{f = c\}$ es claro que, bajo *(T)*, f verificará la condición *(f2)*. Cuando la función f verifica la condición *(T)* podremos calcular las tasas de convergencia de $\partial\{f_n \geq c\}$ como se demuestra en el siguiente teorema.

Teorema 1.30. *Sea* (M, ρ) *un espacio métrico verificando (M1) y*
$$f : M \to \mathbb{R},$$

una función verificando la condicion (T). Sean

$$f_n : M \to \mathbb{R}, \quad n = 1, 2, \ldots$$

funciones aleatorias definidas en el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P)$. *Supongamos que, para cada* n*, con probabilidad uno,* f_n *es continua y además*

$$\|f - f_n\|_\infty = \sup_{x \in M} |f_n(x) - f(x)| \to 0, \quad c.s.$$

Entonces, si $\partial\{f \geq c\} \neq \emptyset$*, se verifica*

$$d_H\big(\partial\{f \geq c\}, \partial\{f_n \geq c\}\big) = \mathcal{O}(\|f - f_n\|_\infty), \quad c.s.$$

Demostración. En primer lugar acotaremos

$$\sup_{x \in \partial\{f \geq c\}} d\big(x, \partial L_n(c)\big).$$

Sea $x \in \partial\{f \geq c\}$. Definamos $\varepsilon_n = 2\|f - f_n\|_\infty$. Como, casi seguro, $\varepsilon_n \to 0$, se tiene que, con probabilidad uno, existe n_0 (independiente de x) tal que, si $n \geq n_0$, $\varepsilon_n \leq \gamma$. En este caso, por la propiedad *(T)*, existen $u_n \equiv u_x^{\varepsilon_n}$ y $l_n \equiv l_x^{\varepsilon_n}$ verificando

$$f(u_n) = c + \varepsilon_n, \quad \rho(x, u_n) \leq L\varepsilon_n; \quad f(l_n) = c - \varepsilon_n, \quad \rho(x, l_n) \leq L\varepsilon_n.$$

No es restrictivo, para el propósito de acotar $\rho(x, \partial L_n(c))$, suponer que $\|f - f_n\|_\infty > 0$. En este caso

$$\begin{aligned} f_n(u_n) = f(u_n) + f_n(u_n) - f(u_n) = c + \varepsilon_n + f_n(u_n) - f(u_n) &\geq c + \varepsilon_n - \|f - f_n\|_\infty \\ &= c + 2\|f - f_n\|_\infty - \|f - f_n\|_\infty > c. \end{aligned}$$

Análogamente se prueba que $f_n(l_n) < c$. En virtud de *(M1)* y como, con probabilidad uno $\varepsilon_n \to 0$, podemos suponer que $B\big(u_n, \rho(u_n, l_n)\big)$ es conexa, en cuyo caso existirá $z_n \in \partial\{f_n \geq c\}$ tal que $z_n \in B\big(u_n, \rho(u_n, l_n)\big)$ y, por lo tanto,

$$\begin{aligned} \rho(z_n, x) \leq \rho(z_n, u_n) + \rho(u_n, x) \leq \rho(u_n, l_n) + \rho(u_n, x) &\leq \rho(u_n, x) + \rho(x, l_n) + \rho(u_n, x) \\ &\leq 3L\varepsilon_n = 6L\|f - f_n\|_\infty. \end{aligned}$$

Así, si $n \geq n_0$,

$$\sup_{x \in \partial\{f \geq c\}} d\big(x, \partial L_n(c)\big) \leq 6L\|f - f_n\|_\infty. \tag{1.29}$$

En segundo lugar acotaremos

$$\sup_{x \in \partial L_n(c)} d(x, \partial L(c)).$$

Sea $x \in \partial\{f_n \geq c\}$. Como, con probabilidad uno, f_n es continua entonces

$$f_n(x) = c, \quad c.s.,$$

y, por tanto,

$$|f(x) - c| = |f(x) - f_n(x)| \leq \|f - f_n\|_\infty, \quad c.s.$$

Recordemos que, si $n \geq n_0$, se verifica $\varepsilon_n \leq \gamma$. Entonces, por *(T)*,

$$d\big(x, \partial L(c)\big) \leq L|f(x) - c| \leq L\|f - f_n\|_\infty, \quad n \geq n_0.$$

Así, si $n \geq n_0$,

$$\sup_{x \in \partial L_n(c)} d\big(x, \partial L(x)\big) \leq L\|f - f_n\|_\infty. \tag{1.30}$$

Utilizando 1.29 y 1.30 obtenemos

$$d_H\big(\partial\{f \geq c\}, \partial\{f_n \geq c\}\big) = \mathcal{O}(\|f - f_n\|_\infty), \quad c.s.,$$

que es lo que queríamos demostrar. □

1.6. Estimación *plug-in* del soporte de una densidad

En las secciones anteriores estudiamos la estimación de conjuntos de nivel $\{f \geq c\}$ donde c era un número real cualquiera. El objetivo de esta sección es presentar los resultados de Cuevas y Fraiman, ver [5], en dicho trabajo se propone estimar $\{f \geq 0\}$ por $\{\hat{f}_n \geq \alpha_n\}$ siendo $\hat{f}_n$ un estimador de f y α_n una sucesión de números reales, y hacer tener α_n a 0. Para la consistencia en medida del estimador no será necesario que $\hat{f}_n$ sea un estimador basado en núcleos. Sí lo será para la convergencia en distancia de Hausdorff como veremos en las siguientes secciones.

1.6.1. Consistencia en Medida

Los resultados de esta sección, referentes a la consistencia en medida del estimador, serán universales, significando esto que no serán necesarias hipótesis sobre S, excepto algunas muy elementales para evitar casos patológicos. A f y f_n le pediremos las siguientes hipótesis:

(S1) Si $E_0 = \{x \in S : f(x) = 0\}$ entonces la medida de Lebesgue $\mu_L(E_0) = 0$. Esta hipótesis es solamente para evitar casos patológicos, ejemplo de esto sería el caso en que tuviéramos un $A \in [0,1]$ abierto y denso en $[0,1]$ tal que $0 < \mu_L(A) < 1$. Si tomaramos f la densidad de una variable con distribución uniforme en A y nula en A^c entonces el soporte de f sería todo el intervalo $[0,1]$ y $\mu(E_0) > 0$.

(R1) $\alpha_n^{-1} \int |f_n - f| \to 0 \quad c.s.$ (resp., en probabilidad).

(R2) $\rho_n \int |f_n - f| \to 0$ en probabilidad, donde ρ_n es una sucesión tal que

$$\rho_n \to \infty \quad y \quad \rho_n \alpha_n a_n \to 0 \quad \text{donde} \quad a_n = \mu_L(\{f \leq 2\alpha_n\} \cap S).$$

Veamos ahora una definición que nos permitirá clasificar las densidades según la *velocidad* en que decrecen a 0. Mas precisamente:

Definición 1.31. Sea f una densidad en $\mathbb{R}^d$ con soporte compacto S, y $\gamma > 0$. Denotemos $f^*(t) = \mu_L(\{f < t\} \cap S)$. Decimos que f tiene orden γ de decaimiento si $f^*(t)$ tiene el mismo orden (cuando $t \to 0^+$) que t^γ, esto es,

$$0 < \liminf_{t \to 0^+} \frac{f^*(t)}{t^\gamma} \leq \limsup_{t \to 0^+} \frac{f^*(t)}{t^\gamma} = c,$$

para alguna constante c.

Denotaremos como $\mathscr{S}_\gamma(S)$ el espacio de las densidades con soporte S y orden de decaimiento γ. Denotemos:

$$\mathscr{S}_\infty(S) = \{f : f^*(t) = o(t^\gamma) \quad \text{cuando} \quad t \to 0^+, \qquad \text{para todo } \gamma > 0\}.$$

Teorema 1.32. *Sea f una densidad en $\mathbb{R}^d$ con soporte compacto S. Dada una sucesión $\{f_n\}_{n\geq 1}$ de estimadores de densidad, definimos una sucesión de estimadores $S_n = \{f_n > \alpha_n\}$, donde $\alpha_n \to 0$.*

(a) Si se verifican (S1) y (R1), entonces $d_\mu(S_n, S) \to 0$ c.s. (resp, en probabilidad).

(b) Si se verifica (R2), entonces $\beta_n d_\mu(S_n, S) \to 0$, en probabilidad, donde

$$\beta_n = \frac{1}{a_n + (\rho_n \alpha_n)^{-1}}. \tag{1.31}$$

(c) Supongamos que $f \in \mathscr{S}_\gamma(S)$ y que (R2) se verifica con $\rho_n = n^\rho$ $(\rho > 0)$, tomemos $\alpha_n = n^{-\alpha}$, donde $0 < \alpha < \rho$ y $\rho - \alpha < \alpha\gamma$. Entonces

$$n^{\rho-\alpha} d_\mu(S_n, S) \to 0 \quad \textit{en probabilidad.}$$

Si $f \in \mathscr{S}_\infty(S)$, entonces la estimación de S tiene orden n^β para $\beta < \rho$.

Demostración. Definamos $A_n = \big\{x : |f_n(x) - f(x)| \geq \alpha_n\big\}$. Considerando una partición adecuada de $S_n \triangle S$ y tomando en cuenta que $\mu_L(S_n \cap S^c \cap A_n^c) = 0$ y $S_n^c \cap S \cap A_n^c \subset \{f \leq 2\alpha_n\} \cap S$, obtenemos que

$$d_\mu(S_n, S) \leq \mu_L(A_n) + \mu_L(S_n^c \cap S \cap A_n^c) \leq \mu_L(A_n) + a_n.$$

De *(S1)* $a_n \downarrow 0$, ya que $\{f \leq 2\alpha_n\} \cap S \downarrow E_0$. Sabemos también que $\mu_L(A_n) \to 0, a.s.$: esto se sigue inmediatamente de *(R1)*, ya que

$$\alpha_n^{-1} \int |f_n - f| \geq \mu_L(A_n) + \alpha_n^{-1} \int_{A_n^c} |f_n - f|.$$

Esto concluye la prueba de la parte (a). Para demostrar (b), observemos que para todo $\delta > 0$ y n suficientemente grande

$$P\big(\beta_n d_\mu(S_n, S) > \delta\big) \leq P\left(\mu_L(A_n) + a_n > \frac{\delta}{\beta_n}\right) \leq P\left(\frac{1}{\alpha_n\left(\delta/\beta_n - a_n\right)} \int |f_n - f| > 1\right),$$

donde hemos usado que $\delta/\beta_n - a_n$ es positiva para n suficientemente grande, que se sigue de que $\rho_n a_n \alpha_n \to 0$. Luego, de *(R2)* se sigue la convergencia a 0 de la última expresión de la desigualdad anterior.

La parte c) se sigue de 1.31, ya que $\alpha_n = n^{-\alpha}$, la hipótesis $f \in \mathscr{S}_\gamma(S)$ implica que la sucesión a_n es de orden $n^{\alpha\gamma}$ y, por lo tanto, β_n es de orden $n^{\rho-\alpha}$. $\square$

Observación 1.33. *Es claro que la sucesión a_n en (R2) depende de la forma en que f decrece a 0. $f(x) > a > 0$ para todo x y $a_n = 0$ a partir de cierto n, será el mejor* *caso a la hora de estimar S. En general cuanto mas lento decrezca a_n a 0 peor será el orden de convergencia β_n que se puede obtener de acuerdo al teorema anterior. Esto es intuitivo si pensamos que a_n decrecerá lento a 0 cuando f tenga zonas de medida muy grande según μ_L pero de probabilidad muy baja. Observemos, respecto a la parte (c), que, si bien al tener $\beta = n^{\rho-\alpha}$ es claro que el orden mejorara cuanto mas pequeño sea α, la cota $\alpha\gamma$ va en la dirección opuesta. Por lo tanto es también de esperar que en el caso en que $f \in \mathscr{S}_\infty(S)$ vamos a poder estimar el soporte a un orden arbitrariamente cerca de n^β.*

órdenes de convergencia para la distancia en medida

En esta subsección vamos a presentar un resultado referente a la velocidad de convergencia en medida, del estimador propuesto. Para eso será necesario imponer restricciones de forma a S. Consideremos $S^h = S \oplus hB$, y la función $\Delta(S,h) = \mu_L(S^h) - \mu_L(S)$. Esta función nos da alguna idea de la forma de ∂S, cuanto mas simple sea, mas pequeña será $\Delta(S,h)$. En el caso en que S es una unión de conjuntos convexos se demuestra que $\Delta(S,h) = \mathcal{O}(h)$ (ver [3]). Para esta sección f_n serán estimadores de tipo núcleo de la forma.

$$\hat{f}_n = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x-X_i}{h}\right) \equiv \frac{1}{n}\sum_{i=1}^{n} K_h(x-X_j),$$

donde $h = h_n$, K es la función núcleo y $K_h = (1/h^d)K(t/h)$.
Vamos a pedir la siguiente condición sobre el núcleo K:

(K) $c_1\mathbb{I}_{B(0,r_1)}(t) \le K(t) \le c_2\mathbb{I}_{B(0,r_2)}(t)$ siendo c_1 y c_2 constantes positivas y $0 < r_1 < r_2$. Donde $\mathbb{I}_A$ denota la función indicatriz del conjunto A.

Notación: Denotemos, para $\varepsilon > 0$, $R(S,\varepsilon)$ el número mínimo de bolas de radio ε, centradas en puntos de S, necesarias para cubrir S.

Teorema 1.34. *Sea $S_n = \{\hat{f}_n > \alpha_n\}$, donde $\hat{f}_n$ es un estimador basado en núcleos, cuyo núcleo K verifica (K1). Supongamos que existe $a > 0$ tal que $f > a > 0$ en S, supongamos además que S es estándar, recordemos que, esto significa que para todo $\lambda > 0$, existe $\delta \in (0,1)$ tal que*

$$\mu_L\big(B(x,\varepsilon)\cap S\big) \ge \delta\mu_L\big(B(x,\varepsilon)\big) \quad \forall x \in S, \quad 0 < \varepsilon \le \lambda. \tag{1.32}$$

Entonces

$$E\big(d_\mu(S_n,S)\big) \le c_3 h_n^d R(S, r_1 h_n/2)\exp(-c_4 n h_n^d) + \Delta(S, r_2 h_n),$$

a partir de cierto n, donde c_3 y c_4 son constantes positivas. Como una consecuencia directa, si asumimos que $\Delta(S,h_n) = \mathcal{O}(h_n)$ cuando h_n tiende a 0, tenemos que

$$E\big(d_\mu(S_n,S)\big) = \mathcal{O}\big(\exp(-c_4 n h_n^d) + h_n\big).$$

Por lo tanto si se elige h_n adecuadamente, se puede obtener un orden de tipo $E\big(d_\mu(S_n,S)\big) = \mathcal{O}(n^{-s})$ con $0 < s < 1/d$.

Demostración. Tenemos que

$$\mu_L(S_n \triangle S) = \mu_L\big(\{\hat{f}_n > \alpha_n, f = 0\}\big) + \mu_L\big(\{f > 0, \hat{f}_n \leq \alpha_n\}\big). \tag{1.33}$$

Por *(K1)*, si $x \in (S^{r_2 h})^c$, entonces $\hat{f}_n(x) = 0$ c.s. Por lo tanto

$$\mu_L\big(\{\hat{f}_n > \alpha_n, f = 0\}\big) \leq \mu_L(S^{r_2 h_n}) - \mu_L(S) = \Delta(S; r_2 h_n) \quad c.s. \tag{1.34}$$

Para acotar el segundo término de 1.33 consideremos un cubrimiento minimal de S con bolas $B(Z_j, r_1 h_n/2)$, $Z_j \in S, j = 1, \ldots, R(S; r_1 h_n/2)$. Denotemos $B_j := B(Z_j, r_1 h_n/2)$ y $R := R(S, r_1 h_n/2)$. Entonces

$$\mu_L(S \cap S_n^c) \leq \mu_L\left(\Big(\bigcup_{j=1}^{R} B_j\Big) \cap S_n^c\right) \leq \sum_{j=1}^{R} \mu_L(B_j \cap S_n^c). \tag{1.35}$$

Sea

$$A_{n,j} = \left\{\frac{1}{n h_n^d} \sum_{i=1}^{n} \mathbb{I}_{B_j}(X_i) > \frac{\alpha_n}{c_1}\right\}.$$

Obsérvese que si $\omega \in A_{n,j}$, entonces $B_j \cap S_n^c(\omega) = \emptyset$, para ver esto sea $t \in B_j$, entonces

$$\begin{aligned}
\frac{1}{n h_n^d} \sum_{i=1}^{n} K\left(\frac{t - X_i}{h_n}\right) &\geq \frac{1}{n h_n^d} \sum_{i=1}^{n} c_1 \mathbb{I}_{B(t, r_1 h_n)}(X_i) \\
&= \frac{c_1}{n h_n^d} \#\big\{i : X_i \in B(t, r_1 h_n)\big\} \\
&\geq \frac{c_1}{n h_n^d} \#\big\{i : X_i \in B_j\big\} > \alpha_n.
\end{aligned}$$

Luego, anotando $\lambda_1 = \mu_L\big(B(0,1)\big)$, tenemos que

$$\begin{aligned}
E\left(\sum_{j=1}^{R} \mu_L(B_j \cap S_n^c)\right) &\leq E\left(\sum_{j=1}^{R} \mathbb{I}_{A_{n,j}^c} \lambda_1 \left(\frac{r_1 h_n}{2}\right)^d\right) \\
&= \sum_{j=1}^{R} h_n^d \lambda_1 \left(\frac{r_1}{2}\right)^d P(A_{n,j}^c).
\end{aligned}$$

Lo que haremos ahora será acotar $P(A_{n,j}^c)$. Si δ es la constante de estandaridad, tomamos $\lambda \geq \sup_n r_1 h_n/2$; y $f > a > 0$ en S tenemos que

$$p_{j,n} := p_j = E\big(\mathbb{I}_{B_j}(X_i)\big) = P(X_i \in B_j) > a\delta \left(\frac{r_1 h_n}{2}\right)^d \lambda_1 := a_1 h_n^d,$$

de donde $p_j/2 - \alpha_n h_n^d/c_1 > a_1 h_n^d/2 - \alpha_n h_n^d/c_1$, luego, ya que $\alpha_n \to 0$ y $a_1 > 0$, existe n_0 tal que $p_j/2 - \alpha_n h_n^d/c_1 > 0$, para todo $n \geq n_0$ y

$$\begin{aligned}
P(A_{n,j}^c) &= P\left(\sum_{i=1}^{n} \big(p_j - \mathbb{I}_{B_j}(X_i)\big) \geq n p_j - \frac{n h_n^d \alpha_n}{c_1}\right) \\
&\leq P\left(\sum_{i=1}^{n} \big(p_j - \mathbb{I}_{B_j}(X_i)\big) \geq \frac{n p_j}{2}\right) \quad \text{para } n \text{ suficientemente grande.}
\end{aligned}$$

Si usamos una desigualdad de tipo Bernstein, obtenemos que

$$P(A^c_{n,j}) \leq 2\exp\left(-\frac{3np_j}{28}\right). \tag{1.36}$$

Entonces, si $p_j > a_1 h^d$, de 1.36 y 1.34, anotando $c_3 = 2\lambda_1 r_1^d 2^{-d}$, $c_4 = 3a_1/28$ se sigue la tesis. □

1.6.2. Consistencia en distancia de Hausdorff

En esta sección veremos resultados referentes a la consistencia en distancia de Hausdorff del estimador propuesto anteriormente, para el caso en que $\hat{f}_n$ es de tipo núcleo. Serán necesarias las siguientes hipótesis sobre el núcleo.

(K2) K es una densidad acotada, uniformemente Lipschitz en $\mathbb{R}^d$, tal que $\|t\|^d K(t)$ es acotado y existen constantes positivas c_1, r_1 tal que

$$c_1 \mathbb{I}_{B(0,r_1)}(u) \leq K(u), \tag{1.37}$$

donde $\mathbb{I}_A$ denota la función indicatriz del conjunto A.

(K3) $K(u)$ es una función decreciente de $\|u\|$ tal que $\|u\|^{d+1}K(u) \to 0$ cuando $\|u\| \to \infty$.

(H1) $h_n \to 0$ y $nh_n^d/\log n \to \infty$, cuando $n \to \infty$.

Usaremos también la hipótesis de que K tiene soporte compacto, significando esto que es 0 fuera de un conjunto compacto.

Teorema 1.35. *Supongamos que el soporte S es compacto y estándar, y que f es acotada y existe $a > 0$ tal que $f > a > 0$ en S. Si suponemos (K2),(K3) y (H1), entonces*

$$\beta_n d_H(S_n, S) \to 0 \quad c.s. \tag{1.38}$$

para toda sucesión $\beta_n \to \infty$ tal que $\beta_n h_n \to 0$ y $\{\beta_n^{d+1} h_n/\alpha_n\}$ es acotada. Esta conclusión también se cumple si (K3) se remplaza por la hipótesis de que K tiene soporte compacto. En este caso la hipótesis de que $\{\beta_n^{d+1} h_n/\alpha_n\}$ sea acotada no es necesaria y se obtiene cualquier orden β_n de tipo $o\big((n/\log n)^{1/d}\big)$ tomando h_n adecuadamente.

Demostración. Veamos que

$$\exists a_0 \text{ tal que } \inf_{x\in S} \hat{f}_n(x) > a_0 \text{ a partir de un cierto } n \quad c.s. \tag{1.39}$$

Usando que S es estándar , que $f > a > 0$ en S y *(K2)*, tenemos que, para todo $x \in S$ y $h_n < r_1$,

$$\begin{aligned} E\big(\hat{f}_n(x)\big) &= \int K_h(x-t)f(t)dt \geq a\int_S K_h(x-t)dt \\ &\geq c_1 a \int_S \frac{1}{h_n^d}\mathbb{I}_{B(0,r_1)}\left(\frac{x-t}{h_n}\right)dt \geq \frac{c_1 a}{h_n^d}\delta\mu_L\big(B(x, r_1 h_n)\big) \\ &= \frac{c_1 a}{h_n^d}\delta\mu_L\big(B(0,1)\big) r_1^d h_n^d := 2a_0 > 0. \end{aligned}$$

Luego 1.39 se sigue si demostramos

$$\sup_x \left| \hat{f}_n(x) - E\big(\hat{f}_n(x)\big) \right| \to 0 \quad c.s., \text{ cuando } n \to \infty.$$

La demostración de éste resultado, bajo la hipótesis *(H1)* y *(K2)* puede encontrarse en Prakasa Rao [(1983), páginas 185-187], ver [17].
Para demostrar 1.38, sea $\varepsilon > 0$, veremos que $S_n \subset S^{\varepsilon/\beta_n}$ c.s. a partir de un cierto n. Para ver esto, observemos que para todo $x \neq 0$,

$$\frac{1}{h_n^d} K\left(\frac{x}{\|x\|}\frac{\varepsilon}{\beta_n h_n}\right) = \frac{\beta_n^{d+1} h_n}{\varepsilon^{d+1}} K\left(\frac{x}{\|x\|}\frac{\varepsilon}{\beta_n h_n}\right)\left(\frac{\varepsilon}{\beta_n h_n}\right)^{d+1}.$$

Dado que $\beta_n h_n \to 0$, $\{\beta_n^{d+1} h_n/\alpha_n\}$ es acotado, y *(K3)* tenemos que

$$\forall \varepsilon > 0 \quad \exists n_1 : \quad K_h\left(\frac{x}{\|x\|}\frac{\varepsilon}{\beta_n}\right) < \alpha_n \quad \forall n > n_1, \quad \forall x \neq 0 \tag{1.40}$$

Nótese que, si K tiene soporte compacto, 1.40 se verifica, sin necesidad de pedir que $\{\beta_n^{d+1} h_n/\alpha_n\}$ sea acotado. Si $x \in S$ y $n > n_1$, 1.40 y la hipótesis de que K es una función decreciente de $\|x\|$ implica que existe X_j ($\in S \quad c.s.$) tal que

$$K_h(x - X_j) > \alpha_n \quad y \quad \|x - X_j\| < \frac{\varepsilon}{\beta_n},$$

lo cual implica que $S_n \subset S^{\varepsilon/\beta_n} \quad c.s.$ para $n > n_1$. Por otro lado, como $\alpha_n \to 0$, y 1.39 implica que $S \subset S_n$, y por supuesto $S \subset S_n^{\varepsilon/\beta_n}$, para n suficientemente grande, *c.s.*.
Obtenemos entonces que

$$S \subset S_n^{\varepsilon/\beta_n} \quad y \quad S_n \subset S^{\varepsilon/\beta_n} \quad \text{a partir de un cierto } n c.s., \varepsilon > 0$$

Observemos que se puede obtener cualquier orden β_n de tipo $o(n/\log n)^{\frac{1}{d}}$ tomando $h_n = o(\beta_n^{-1})$ y $o(h_n) = (\log n/n)^{\frac{1}{d}}$. Se puede demostrar, ver [7], que el resultado anterior no puede mejorarse, la demostración es similar a la del Lema 1.26. □

1.7. Estimación de conjuntos α-convexos

En esta sección veremos algunos resultados referentes a la estimación de conjuntos α-convexos, resultados que se encuentran en [19]. Como se dijo, esta hipótesis es una generalización intuitiva de la idea de convexidad, y es equivalente a pedir que ruede libremente una bola por el complemento del conjunto. Si bien la hipótesis es menos restrictiva que la convexidad, se obtienen los mismos órdenes de convergencia.

1.7.1. El estimador

Si observamos las equivalencias del teorema 1.11 es natural estimar el soporte S de una distribución P_X a partir de una muestra $\aleph_n = \{X_1, \ldots, X_n\}$, para el caso en que S es α-convexo, por su capsula α-convexa:

$$C_\alpha(\aleph_n) = (\aleph_n \oplus \alpha \overset{\circ}{B}) \ominus \alpha \overset{\circ}{B} . \tag{1.41}$$

Observemos que para el caso en que S es α-convexo se cumple que $C_\alpha(\aleph_n) \subset S$. Implícitamente estamos suponiendo que el valor de α es conocido pero ésto no es problema, ya que si S es α-convexo, es λ-convexo para todo $\lambda \leq \alpha$. De modo que para el caso en que α es desconocido podemos tomar $\alpha_n \to 0$ y considerar $C_{\alpha_n}(\aleph_n)$. Este caso será estudiado más adelante. Análogamente a lo que pasaba con el estimador de Devroye y Wise, es de esperarse que α_n deba decrecer a 0 pero no demasiado rápido. Veamos ahora algunos resultados respecto de la consistencia del estimador.

Teorema 1.36. *Sea $X_1, X_2, \ldots$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos en $\mathbb{R}^d$ con distribución común P_X y soporte S. Supongamos que S es compacto y estándar respecto de P_X. Sea $\hat{S}_n$ un estimador de S, tal que para n suficientemente grande*

$$\aleph_n \subset \hat{S}_n \subset S,$$

entonces, con probabilidad uno

$$d_H(S, \hat{S}_n) = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{1/d}\right).$$

Demostración. Por hipótesis, para n suficientemente grande $\aleph_n \subset \hat{S}_n \subset S$ y por lo tanto $d_H(S, \hat{S}_n) \leq d_H(\aleph_n, S)$, luego, el teorema es una consecuencia inmediata del Teorema 1.22. □

Corolario 1.37. *Como corolario inmediato obtenemos que, bajo las hipótesis del teorema anterior $d_H\big(S, C_r(\aleph_n)\big) = \mathcal{O}\big((\log /n)^{1/d}\big)$.*

Si además pedimos que S esté en las hipótesis del Teorema 1.11 el resultado anterior se puede mejorar:

Teorema 1.38. *Sea $X_1, X_2, \ldots$ una sucesión de vectores aleatorios independientes e idénticamente distribuidos en $\mathbb{R}^d$ con distribución común P_X absolutamente continua respecto de la medida de Lebesgue, μ_L, sea f la densidad de P_X y S su soporte. Supongamos que f está en las hipótesis del teorema 1.11. Supongamos además que existe $a > 0$ tal que $f > a > 0$ en S, entonces, con probabilidad 1*

$$d_H\big(S, C_r(\aleph_n)\big) = \mathcal{O}\left(\left(\frac{\log n}{n}\right)^{2/(d+1)}\right).$$

Los mismos órdenes se verifican para $d_H\big(\partial S, \partial C_r(\aleph_n)\big)$ y $d_\mu\big(S, C_r(\aleph_n)\big)$.

Demostración. La demostración será consecuencia de 2 proposiciones, veamos la primera.

Proposición 1.39. *Sea S un conjunto en las hipótesis del Teorema 1.11, sea $\varepsilon_n \to 0$ tal que*

$$P\big(\partial S \subset B(Z_n, 2\varepsilon_n),\ \text{a partir de un cierto } n\big) = 1,$$

donde $Z_n = \big\{X_i \in \aleph_n : d(X_i, \partial S) \leq \varepsilon_n^2\big\}$. Entonces, con probabilidad uno, $d_H\big(S, C_r(\aleph_n)\big)$, $d_H\big(\partial S, \partial C_r(\aleph_n)\big)$ y $d_\mu\big(S, C_r(\aleph_n)\big)$ son de orden ε_n^2.

Demostración. La demostración de esta proposición se basa en 3 lemas, el primero de ellos es muy útil desde el punto de vista geométrico.

Lema 1.40. *Sea S un conjunto en las hipótesis del Teorema 1.11. Sea $y \in S^c$, tal que $d(y,S) = r - \delta$, para algún $0 \leq \delta < r$. Si $x \in S$ verifica $d(x, \partial S) \leq \delta/2$ y $\|x - y\| \geq r$, entonces, la proyección s, de y sobre S, verifica que*

$$\|x - s\| \geq \sqrt{\frac{r\delta}{2}}.$$

Demostración. Sea η el vector normal saliente de S en el punto $s \in \partial S$, es fácil ver que

$$y = s + (r - \delta)\eta,$$

y

$$B(t,s) \subset S \text{ con } t = s - r\eta.$$

Observemos que, por las hipótesis sobre x

$$\begin{aligned} r^2 &\leq \|x - y\|^2 = \|x - s - (r-\delta)\eta\|^2 = \|x - s\|^2 - 2(r-\delta)\langle x - s, \eta\rangle + (r-\delta)^2, \\ \left(r - \frac{\delta}{2}\right)^2 &\leq \|x - t\|^2 = \|x - s + r\eta\|^2 = \|x - s\|^2 + 2r\langle x - s, \eta\rangle + r^2. \end{aligned}$$

La segunda desigualdad se sigue de que $d(x, \partial S) \leq \delta/2$ y $B(t,r) \subset S$. Estas dos últimas desigualdades pueden rescribirse como

$$\begin{aligned} \|x - s\|^2 - 2(r-\delta)\langle x - s, \eta\rangle &\geq 2r\delta - \delta^2, \\ \|x - s\|^2 + 2r\langle x - s, \eta\rangle &\geq -r\delta + \frac{\delta^2}{4}. \end{aligned}$$

Si multiplicamos la primera desigualdad por r, la segunda por $r - \delta$ las sumamos, y dividimos todo por $2r - \delta$ obtenemos que

$$\|x - s\|^2 \geq \frac{2r^2\delta - r\delta^2 - r(r-\delta)\delta + (r-\delta)\delta^2/4}{2r - \delta} = \frac{r^2\delta + (r-\delta)\delta^2/4}{2r - \delta} \geq \frac{r\delta}{2}.$$

lo que concluye la demostración del lema. □

Veamos ahora otro lema, que surge como una aplicación del anterior, y que también tiene una interpretación geométrica clara y es de importancia en la demostración del Teorema. Intuitivamente nos da una cota para la distancia que puede *meterse* dentro del conjunto, una bola centrada en el exterior, sin intersecar puntos de la muestra.

Lema 1.41. *Sea $\partial S \subset B(Z_n, 2\varepsilon_n)$. Si $y \in S^c$ cumple que $d(y,S) = r - \delta$ con $\max(2, 8/r)\varepsilon_n^2 < \delta < r$ entonces*

$$\mathring{B}(y,r) \cap \aleph_n \neq \emptyset.$$

Demostración. Sea s la proyección de y sobre S. La hipótesis $\partial S \subset B(Z_n, 2\varepsilon_n)$ implica que existe $X_i \in Z_n$ tal que $\|X_i - s\| \leq 2\varepsilon_n$. Recordemos que $d(X_i, \partial S) \leq \varepsilon_n^2 < \delta/2$. Por lo tanto si $\|X_i - y\| \geq r$, por el Lema anterior, tenemos que

$$\|X_i - s\| \geq \sqrt{\frac{r\delta}{2}} \geq 2\varepsilon_n,$$

lo cual es absurdo. Por lo tanto hemos probado que $\|X_i - r\| < r$ lo cual concluye la prueba. □

El siguiente lema nos da una *cota inferior* para $C_r(\aleph_n)$.

Lema 1.42. *Supongamos que $\partial S \subset B(Z_n, 2\varepsilon_n)$ y $d_H(\aleph_n, S) < r$. Entonces*

$$S \ominus (L\varepsilon_n^2 B) \subset C_r(\aleph_n), \quad L \geq \max(2, 8/r).$$

Demostración. Supongamos que existe $x \in S \ominus (L\varepsilon_n^2 B)$ que no pertenezca a $C_r(\aleph_n)$. En este caso, existe un $y \in S^c$ tal que $\mathring{B}(y,r) \cap \aleph_n = \emptyset$ y $x \in \mathring{B}(y,r)$. Por el Lema 1.41, sabemos que $d(y,S) \geq r - \text{máx}(2, 8/r)\varepsilon_n^2$. Sea $z \in [x,y] \cap \partial S$ donde $[x,y]$ denota el segmento entre los puntos x e y.
Tenemos que

$$\begin{aligned} &\|x - z\| \geq L\varepsilon_n^2, \\ &\|z - y\| \geq r - \text{máx}(2, 8/r)\varepsilon_n^2, \end{aligned}$$

lo cual es absurdo ya que

$$r > \|x - y\| = \|x - z\| + \|z - y\| \geq r - \text{máx}(2, 8/r)\varepsilon_n^2 + L\varepsilon_n \geq r.$$

Por lo tanto, todo $x \in S \ominus (L\varepsilon_n^2 B)$ está en $C_r(\aleph_n)$. □

Veamos ahora la demostración de la proposición. Por el Lema 1.42 tenemos que, con probabilidad 1, para n suficientemente grande

$$S \ominus L\varepsilon_n^2 B \subset C_r(\aleph_n) \subset S, \quad L \geq \text{máx}(2, 8/r). \tag{1.42}$$

Por el Teorema 1.11, el conjunto S cumple, $(\varepsilon \leq r)$ la condición

$$S = (S \ominus \varepsilon B) \oplus \varepsilon B. \tag{1.43}$$

Si $L\varepsilon_n^2 \leq r$

$$S = (S \ominus L\varepsilon_n^2 B) \oplus L\varepsilon_n^2 B \subset C_r(\aleph_n) \oplus L\varepsilon_n^2 B.$$

Luego, dado que S es r-convexo, $C_r(\aleph_n) \subset S$, y

$$d_H\big(S, C_r(\aleph_n)\big) \leq L\varepsilon_n^2.$$

La distancia en medida puede acotarse de forma similar. De 1.43 tenemos que

$$\mu\big(C_r(\aleph_n) \triangle S\big) \leq \mu\big(S \setminus (S \ominus L\varepsilon_n^2 B)\big).$$

Como $\mu(S) - \mu(S \ominus \varepsilon B) = \mathcal{O}(\varepsilon)$, (ver Ec. (6), en [21]), obtenemos que $\mu\big(C_r(\aleph_n) \triangle S\big) = \mathcal{O}(\varepsilon_n^2)$. Las cotas para los bordes se siguen de 1.42 y 1.43. □

Proposición 1.43. *Para $c>0$ suficientemente grande.*

$$P\big(\partial S \subset B(Z_n, 2\varepsilon_n),\ \text{a partir de un cierto } n\big) = 1, \tag{1.44}$$

donde $\varepsilon_n = (c \log n/n)^{\frac{1}{d+1}}$ *y* $Z_n = \big\{X_i \in \aleph_n : d(X_i, \partial S) \leq \varepsilon_n^2\big\}$.

Demostración. Se puede demostrar (ver [11]) que

$$P\big(\partial S \not\subseteq B(Z_n, 2\varepsilon_n)\big) \leq D(S, \varepsilon_n)\Pi(S, Z_n, \varepsilon_n), \quad \varepsilon > 0,$$

donde $D(S,\varepsilon) = \text{máx}\{\#F : F \subset \partial S, \quad \|x-y\| > \varepsilon, \text{ para } x,\ y \in F, \text{ diferentes}\}$ y $\Pi(S, Z_n, \varepsilon) = \sup_{s\in\partial S} P\big(B(s,\varepsilon) \cap Z_n = \emptyset\big)$ donde $\#F$ denota el cardinal del conjunto F. Es fácil demostrar que $D(S,\varepsilon)$ es de orden ε^{-d}. Luego para ver 1.44 es suficiente demostrar que

$$\sum_{n=1}^{\infty} \varepsilon_n^{-d} \Pi(S, Z_n, \varepsilon_n) < \infty \tag{1.45}$$

Para cada $s \in \partial S$ tenemos que

$$\begin{aligned} P\big(B(s,\varepsilon_n) \cap Z_n = \emptyset\big) &= P\left(\bigcap_{i=1}^{n} \big\{X_i \notin B(\partial S, \varepsilon_n^2) \cap B(s, \varepsilon_n)\big\}\right) \\ &= \Big(1 - P_X\big(B(\partial S, \varepsilon_n^2) \cap B(s,\varepsilon_n)\big)\Big)^n \\ &\leq \exp\Big(-nP_X\big(B(\partial S, \varepsilon_n^2) \cap B(s,\varepsilon_n)\big)\Big), \end{aligned}$$

la última igualdad sale de usar que $(1-x)^n \leq \exp(-nx)$, para $0 \leq x \leq 1$. Por lo tanto como f es acotada por abajo en S, tenemos que si

$$\mu\big(S \cap B(\partial S, \varepsilon^2) \cap B(s,\varepsilon)\big) \geq \alpha\varepsilon^{d+1}, \quad \forall \varepsilon \in [0,\beta], \quad \forall s \in \partial S,$$

para algunas constantes $\alpha, \beta > 0$, se seguiría inmediatamente 1.45, y por lo tanto también 1.44. Queda entonces demostrar esto último.

Lema 1.44. *Sea S un conjunto en las hipótesis del Teorema 1.11. Entonces, existen constantes $\alpha, \beta > 0$ tal que*

$$\mu\big(S \cap B(\partial S, \varepsilon^2) \cap B(s,\varepsilon)\big) \geq \alpha\varepsilon^{d+1}, \quad \forall \varepsilon \in [0,\beta], \quad \forall s \in \partial S.$$

Demostración. Sea $s \in \partial S$. Para $0 < h < r$ definimos $\mathcal{C}(h) = \mathcal{R}(h) \cap B(s - r\eta(s), r)$ donde $\mathcal{R}(h)$ es el *rectángulo* $\big\{x \in \mathbb{R}^d : -h \leq \langle x - s, \eta(s)\rangle \leq 0\big\}$ y $\eta(s)$ es el vector normal saliente de S en s. Veremos que existe una constante $A > 0$ que depende sólo de d y r, tal que, para h suficientemente pequeño

$$\mu\big(\mathcal{C}(h)\big) \geq Ah^{(d+1)/2}. \tag{1.46}$$

Probaremos también que, si $h_\varepsilon = \text{mín}\{\varepsilon^2/2, \varepsilon^2/2r\} = A'\varepsilon^2$ es menor que r, (es decir $\varepsilon < \sqrt{r/A'}$) entonces

$$\mathcal{C}(h_\varepsilon) \subset B(s,\varepsilon) \cap S \cap B(\partial S, \varepsilon^2). \tag{1.47}$$

La demostración del Lema es entonces una consecuencia inmediata de 1.46 y 1.47, ya que, en este caso, para ε suficientemente pequeño tenemos que

$$\mu\big(B(s,\varepsilon)\cap S\cap B(\partial S,\varepsilon^2)\big)\geq \mu\big(\mathcal{C}(h_\varepsilon)\big)\geq Ah_\varepsilon^{(d+1)/2}=A(A')^{(d+1)/2}\varepsilon^{d+1}=\alpha\varepsilon^{d+1}.$$

Para demostrar 1.46, observemos que existe $\mathcal{T}$ que transforma $\mathcal{C}(h)$ en el conjunto

$$\mathcal{C}_0(h)=\mathcal{T}\big(\mathcal{C}(h)\big)=\big\{x=(x_1,\ldots,x_d)\in\mathbb{R}^d:\ 0\leq x_d\leq h\big\}\cap B\big(-(r-h)e_d,r\big),$$

donde $e_d=(0,0,\ldots,1)^T$. Por la invarianza de la medida de Lebesgue bajo rotaciones y traslaciones, tenemos que $\mu\big(\mathcal{C}(h)\big)=\mu\big(\mathcal{C}_0(h)\big)$.

Es claro que 1.46 vale para el caso $d=1$. Para demostrarlo para $d\geq 2$ sea, $(0\leq l\leq h)$, $\mathcal{C}_0(h,l)=\big\{x=(x_1,\ldots,x_{d-1},l)\in\mathbb{R}^d: x\in\mathcal{C}_0(h)\big\}$.

$$\mu\big(\mathcal{C}_0(h)\big)=\int_0^h\mu_{d-1}\big(\mathcal{C}_0(h,l)\big)dl,$$

donde μ_{d-1} denota la medida de Lebesgue $(d-1)$-dimensional. Obsérvese que $\mathcal{C}_0(h,l)$ es una esfera $d-1$-dimensional, cuyo radio $r(l)$ satisface la ecuación

$$r(l)=\sqrt{r^2-(r-h+l)^2}=\sqrt{2r(h-l)-(h-l)^2}.$$

Luego

$$\mu(\mathcal{C}_0(h))=\omega_{d-1}\int_0^h r(l)^{d-1}dl\geq\omega_{d-1}r^{d-1}\int_0^h l^{(d-1)/2}dl=\frac{2\omega_{d-1}r^{d-1}}{d+1}h^{(d+1)/2},$$

donde ω_{d-1} es el volumen de la bola $d-1$-dimensional. Esto concluye la prueba de 1.46. Solo resta demostrar 1.47. Para $x\in\mathcal{C}(\varepsilon^2/2)$ tenemos que

$$d(x,\partial S)\leq d\big(x,B(s+r\eta(s),r)\big)\leq 2\|x-P_{H_s}(x)\|=2|\langle x-s,\eta(s)\rangle|\leq\frac{2\varepsilon^2}{2}=\varepsilon^2,$$

donde $P_{H_s}(x)$ es la proyección de x sobre $H_s=\big\{x:\ \langle x-s,\eta(s)\rangle=0\big\}$.Obsérvese que H_s no es otra cosa que el plano tangente a ∂S en s. La primera desigualdad se sigue de que $\mathring{B}\big(s+r\eta(s),r\big)\subset S^c$ (ver 1.40). Tenemos que

$$\mathcal{C}\left(\frac{\varepsilon^2}{2}\right)\subset S\cap B(\partial S,\varepsilon^2).\tag{1.48}$$

Por lo tanto, para demostrar 1.47 basta con demostrar que

$$\mathcal{C}\left(\frac{\varepsilon^2}{2r}\right)\subset B(s,\varepsilon).\tag{1.49}$$

Si $x\in\mathcal{C}\big(\varepsilon^2/(2r)\big)$ entonces $x\in B\big(s-r\eta(s),r\big)$ y

$$\begin{aligned}r^2\ \geq\ \|x-(s-r\eta(s))\|^2\ &=\ \|x-s+P_{H_s}(x)-P_{H_s}(x)+r\eta(s)\|^2\\ =\|P_{H_s}(x)-s\|^2+\|x-P_{H_s}(x)+r\eta(s)\|^2\ &=\ \|P_{H_s}(x)-s\|^2+\big(r+\langle x-s,\eta(s)\rangle\big)^2.\end{aligned}$$

Como $\langle x-s,\eta(s)\rangle \geq -\varepsilon^2/(2r) > -r$, tenemos que $\big(r+\langle x-s,\eta(s)\rangle\big)^2 \geq \big(r-\varepsilon^2/(2r)\big)^2$ y, por lo tanto

$$\|P_{H_s}(x)-s\|^2 \leq r^2 - \left(r-\frac{\varepsilon^2}{2r}\right)^2.$$

Luego

$$\begin{aligned}\|x-s\|^2 = \|x-P_{H_s}(x)+P_{H_s}(x)-s\|^2 &= \|P_{H_s}(x)-s\|^2 + \langle x-s,\eta(s)\rangle^2 \\ &\leq r^2 - \left(r-\frac{\varepsilon^2}{2r}\right)^2 + \left(\frac{\varepsilon^2}{2r}\right)^2 = \varepsilon^2,\end{aligned}$$

es decir, $x \in B(s,\varepsilon)$, lo cual concluye la prueba de 1.48 y la demostración del Lema. □

La proposición 1.43 es consecuencia de estos dos lemas. □

Claramente también el Teorema 1.38 queda demostrado a patir de las proposiciones anteriores. □

Teorema 1.45. *Bajo las mismas hipótesis del teorema anterior, sea $\{r_n\}$ una sucesión de números reales positivos que convergen a 0 tal que $r_n \gg (\log n/n)^{1/d}$. Entonces, con probabilidad 1*

$$d_H\big(S, C_{r_n}(\aleph_n)\big) = \mathcal{O}\left(r_n^{-1}\left(\frac{\log n}{n}\right)^{2/(d+1)}\right).$$

Los mismos órdenes se verifican para $d_H\big(\partial S, \partial C_{r_n}(\aleph_n)\big)$ y $d_\mu\big(S, C_{r_n}(\aleph_n)\big)$.

Demostración. La demostración es muy similar a la del caso r constante, de hecho, la Proposición 1.43 solo depende de la muestra y de las propiedades del soporte, por lo tanto, es independiente del estimador. La Proposición 1.39 se puede demostrar en éste caso, cambiando los lemas correspondientes, de una forma obvia, sustituyendo el Lema 1.42 por el siguiente Lema.

Lema 1.46. *Sea $\partial S \subset B(Z_n, 2\varepsilon_n)$ y $d_H(\aleph_n, S) < r_n \leq r$. Entonces*

$$S \ominus (L_n\varepsilon_n^2 B) \subset C_{r_n}(\aleph_n), \quad L_n \geq \text{máx}(2, 8/r_n).$$

La demostración del teorema sigue los mismos argumentos que el teorema anterior. Observemos que $r_n \gg (\log n/n)^{\frac{1}{d}}$ asegura que

$$P\big(d_H(\aleph_n, S) < r_n \text{ a partir de un cierto } n\big) = 1,$$

que es necesario para aplicar 1.46. □

1.7.2. $E\Big(d_\mu\big(S, C_{r_n}(\aleph_n)\big)\Big)$

Si bien en 1.41 se tomó la bola abierta para definir el estimador, y es fácil ver que $(\aleph_n \oplus \alpha \overset{\circ}{B}) \ominus \alpha \overset{\circ}{B}$ no necesariamente coincide con $(\aleph_n \oplus \alpha B) \ominus \alpha B$, se puede ver que si lo hacen, con probabilidad 1. En lo que sigue estudiaremos el orden de convergencia de $E\Big(d_\mu\big(S, C_{r_n}(\aleph_n)\big)\Big)$ donde $C_{r_n}(\aleph_n) = (\aleph_n \oplus \alpha B) \ominus \alpha B$, resultados extraidos de [16].

Teorema 1.47. *Sea $S \subset \mathbb{R}^d$ un conjunto compacto, no vacío, y α-convexo con $\alpha > 0$. Sea X una variable aleatoria con distribución P_X y densidad f cuyo soporte es S. Sean $\aleph_n = \{X_1, \ldots, X_n\}$ observaciones de X, y $\{r_n\}$ una sucesión de números reales positivos, tales que $r_n \leq \alpha$. Entonces*

$$\lim_{n\to\infty} E\Big(d_\mu\big(S, C_{r_n}(\aleph_n)\big)\Big) = 0$$

si y solo si $\lim_{n\to\infty} nr_n^d = \infty$.

Demostración.

$$\begin{aligned}
E\Big(d_\mu\big(S, C_{r_n}(\aleph_n)\big)\Big) &= E\Big(\mu\big(S \setminus C_{r_n}(\aleph_n)\big)\Big) = E\Big(\mu\big\{x \in S : x \notin C_{r_n}(\aleph_n)\big\}\Big) \\
&= E\int_S \mathbb{I}_{\big\{x\notin C_{r_n}(\aleph_n)\big\}}\mu(dx) = \int_S P\big(x \notin C_{r_n}(\aleph_n)\big)\mu(dx) \\
&= \int_S P\big(\exists y \in B(x, r_n) : B(y, r_n) \cap \aleph_n = \emptyset\big)\mu(dx).
\end{aligned}$$

El resto de la demostración consiste en acotar la probabilidad dentro de la integral anterior, y para eso serán necesarias algunas definiciones y lemas.

Definición 1.48. Sea $x \in \mathbb{R}^d$, $r > 0$ y $\mathcal{E}_{x,r} = \big\{B(y, r) : y \in B(x, r)\big\}$. La familia $\mathcal{U}_{x,r}$ es inevitable para $\mathcal{E}_{x,r}$, si para toda $B(y, r) \subset \mathcal{E}_{x,r}$, existe $U \in \mathcal{U}_{x,r}$ tal que $U \subset B(y, r)$.

Como consecuencia de la definición anterior, si $\mathcal{U}_{x,r_n}$ es una familia inevitable de conjuntos para $\mathcal{E}_{x,r_n}$ entonces

$$\big\{\exists y \in B(x, r_n) : B(y, r_n) \cap \aleph_n = \emptyset\big\} \subset \big\{\exists U \in \mathcal{U}_{x,r_n} : U \cap \aleph_n = \emptyset\big\},$$

y por lo tanto

$$P\big(\exists y \in B(x, r_n) : B(y, r_n) \cap \aleph_n = \emptyset\big) \leq P\big(\exists U \in \mathcal{U}_{x,r_n} : U \cap \aleph_n = \emptyset\big).$$

Si suponemos que $\mathcal{U}_{x,r_n}$ es una familia finita

$$\begin{aligned}
P\big(\exists y \in B(x, r_n) : B(y, r_n) \cap \aleph_n = \emptyset\big) &\leq \sum_{U\in\mathcal{U}_{x,r_n}} P(U \cap \aleph_n = \emptyset) \\
&= \sum_{U\in\mathcal{U}_{x,r_n}} P(\forall X_j, j = 1, \ldots, n, X_j \notin U) \\
&= \sum_{U\in\mathcal{U}_{x,r_n}} \big(1 - P_X(U)\big)^n. \qquad (1.50)
\end{aligned}$$

Luego si para cada $x \in S$ definimos una familia inevitable, $\mathcal{U}_{x,r_n}$, finita, usando 1.50, obtenemos

$$\begin{aligned} E\Big(d_\mu\big(S, C_{r_n}(\aleph_n)\big)\Big) &= \int_S P\big(\exists y \in B(x,r_n) : B(y,r_n) \cap \aleph_n = \emptyset\big)\mu(dx) \\ &\leq \int_S \sum_{U \in \mathcal{U}_{x,r_n}} \big(1 - P_X(U)\big)^n. \end{aligned}$$

De la desigualdad anterior vemos que el problema de encontrar una cota superior para $E\Big(d_\mu\big(S, C_{r_n}(\aleph_n)\big)\Big)$ se reduce al problema de encontrar una cota inferior para $P_X(U)$ para toda $U \in \mathcal{U}_{x,r_n}$. Como sólo estamos interesados en presentar brevemente los resultados de la tesis doctoral de Pateiro,(para mas detalle ver [16]), omitiremos la construcción de familias inevitables en $\mathbb{R}^d$ y enunciaremos los resultados referentes a ellas, para el caso general.

Proposición 1.49. *Sea S un subconjunto compacto, no vacío, de $\mathbb{R}^d$, tal que una bola de radio $\alpha > 0$ rueda libremente en S y $\overline{S^c}$. Sea X una variable aleatoria con distribución P_X y soporte S. Supongamos que P_X verifica*

$$P_X(C) \geq \delta\mu(C \cap S),$$

para todo boreliano C.
Entonces, para todo $x \in S$ tal que $d(x, \partial S) > r_n/2$, existe una familia finita $\mathcal{U}_{x,r_n}$, con m_1 elementos, inevitable para $\mathcal{E}_{x,r_n}$, que verifica

$$P_X(U) \geq L_1 r_n^d, \quad U \in \mathcal{U}_{x,r_n},$$

donde la constante L_1 es independiente de x.

Veamos ahora el caso en que $d(x, \partial S) \leq r_n/2$

Proposición 1.50. *Sea S un subconjunto compacto, de $\mathbb{R}^d$ tal que una bola de radio $\alpha > 0$ rueda libremente en S y $\overline{S^c}$. Sea X una variable aleatoria con distribución P_X y soporte S. Supongamos que P_X verifica*

$$P_X(C) \geq \delta\mu(C \cap S),$$

para todo boreliano C.
Entonces, para todo $x \in S$ tal que $d(x, \partial S) \leq r_n/2$, existe una familia finita $\mathcal{U}_{x,r_n}$, con m_2 elementos, inevitable para $\mathcal{E}_{x,r_n}$, que verifica

$$P_X(U) \geq L_2 r_n^{\frac{d-1}{2}} d(x, \partial S)^{\frac{d+1}{2}}, \quad U \in \mathcal{U}_{x,r_n},$$

donde la constante $L_2 > 0$ es independiente de x.

Como se demuestra en [16], se pueden tomar cada $U \equiv U^u_{x,r_n} = \{x\} \oplus C_{u,r_n}$ la traslación según vectores $u \in \mathcal{W}$ del conjunto C_{u,r_n}. Algo muy intuitivo si pensamos en la simetría del problema en cuestión. Veamos ahora la demostración del teorema. Supongamso primero que $nr_n^d = \infty$, observemos que es suficiente demostrar que para todo $x \in S$

$$\lim_{n\to\infty} P\big(\exists y \in B(x,r_n) : B(y,r_n) \cap \aleph_n = \emptyset\big) = 0. \tag{1.51}$$

ya que por el teorema de convergencia dominada tenemos que

$$\begin{aligned}
\lim_{n\to\infty} E(d_\mu(S, C_{r_n}(\aleph_n))) &= \lim_{n\to\infty} \int_S P\big(\exists y \in B(x,r_n) : B(y,r_n)\cap \aleph_n = \emptyset\big)\mu(dx) \\
&= \int_S \lim_{n\to\infty} P\big(\exists y \in B(x,r_n) : B(y,r_n)\cap \aleph_n = \emptyset\big)\mu(dx) \\
&= 0. \qquad (1.52)
\end{aligned}$$

Para cada $x \in S$ consideremos una familia $\mathcal{U}_{x,r_n} = \{U^u_{x,r_n}, u \in \mathcal{W}\}$ inevitable para $\mathcal{E}_{x,r_n}$, donde $\mathcal{W}$ es un conjunto de m elementos.

$$P\big(\exists y \in B(x,r_n) : B(y,r_n)\cap \aleph_n = \emptyset\big) \leq \sum_{u\in\mathcal{W}} \big(1 - P_X(U^u_{x,r_n})\big)^n. \qquad (1.53)$$

Para acotar $P_X(U^u_{x,r_n})$ en 1.53 usaremos el siguiente lema. Ver Devroye (1983).

Lema 1.51. *Si f es una densidad en $\mathbb{R}^d$ y A es un subconjunto compacto de $\mathbb{R}^d$, con $\mu(A) > 0$, entonces*

$$\lim_{h\to 0} \frac{1}{\mu(hA)} \int_{\{x\}\oplus hA} f(y)dy = f(x), \qquad \textit{para casi todo } x.$$

Luego, usando éste lema y la forma de los U_{x,r_n} obtenemos que para cada x existe h_x tal que si $h \geq h_x$

$$P_X(U^u_{x,h}) = \int_{U^u_{x,h}} f(y)dy \geq \frac{f(x)}{2}\mu(C_{u,h}). \qquad (1.54)$$

Para cada $n \in \mathbb{N}$ sea $h_n \equiv h_{n,x} = \min(r_n, h_x)$. Entonces $U^u_{x,h_n} \subset U^u_{x,r_n}$, y aplicando 1.54

$$P_X(U^u_{x,r_n}) \geq P_X(U^u_{x,h_n}) \geq \frac{f(x)}{2}\mu(C_{u,h_n}) \geq \frac{f(x)}{2}\frac{\mu\big(B(0,h_n)\big)}{m} = \frac{f(x)}{2}\frac{\omega_d h_n^d}{m} = L_x h_n^d,$$

donde, $x \in S$ y $L_x = \frac{f(x)\omega_d}{2m} > 0$. Volviendo ahora a 1.53

$$P\big(\exists y \in B(x,r_n) : B(y,r_n)\cap \aleph_n = \emptyset\big) \leq m(1 - L_x h_n^d)^n \leq m e^{-nL_x h_n^d}.$$

Obsérvese que podemos garantizar que $L_x h_n^d \leq 1$ ya que $L_x h_n^d \leq P_X(U^u_{x,r_n})$. Entonces

$$\lim_{n\to\infty} P\big(\exists y \in B(x,r_n) : B(y,r_n)\cap \aleph_n = \emptyset\big) \leq \lim_{n\to\infty} m e^{-nL_x h_n^d}.$$

Por lo tanto 1.51 es consecuencia de la forma en que fue tomado h_n y de que $\lim_{n\to\infty} n r_n^d = \infty$.

Veamos ahora la prueba del recíproco. Supongamos que $\lim_{n\to\infty} E\big(d_\mu(S, S_n)\big) = 0$. Observemos que

$$\begin{aligned}
P\big(\exists y \in B(x,r_n) : B(y,r_n)\cap \aleph_n = \emptyset\big) &\geq P\big(B(x,r_n)\cap \aleph_n = \emptyset\big) \\
&= \Big(1 - P_X\big(B(x,r_n)\big)\Big)^n. \qquad (1.55)
\end{aligned}$$

Si la sucesión $\{nr_n^d\}$ no converge a infinito cuando $n \to \infty$, entonces podemos encontrar una subsucesión acotada $\{n_k r_{n_k}^d\}$. Es decir existe $M > 0$ tal que $n_k r_{n_k}^d \leq M$ para todo k. Si aplicamos el Lema 1.51, tenemos que para todo x y k suficientemente grande

$$P_X\big(B(x, r_{n_k})\big) = \int_{B(x,r_{n_k})} f(y)dy \leq 2f(x)\mu\big(B(0, r_{n_k})\big) = 2f(x)\omega_d r_{n_k}^d = L_x r_{n_k}^d, \tag{1.56}$$

siendo $L_x = 2f(x)\omega_d$. Para simplificar usemos la notación:

$$\psi_n(x) = P\big(\exists y \in B(x, r_n) : B(y, r_n) \cap \aleph_n = \emptyset\big),$$

y consideremos la subsucesión $\{\psi_{n_k}(x)\}$. Por lo tanto de 1.55 y 1.56 obtenemos que

$$\begin{aligned}
\liminf_{k\to\infty} \psi_{n_k}(x) &\geq \liminf_{k\to\infty} \Big(1 - P_X\big(B(x, r_{n_k})\big)\Big)^{n_k} \\
&\geq \liminf_{k\to\infty} (1 - L_x r_{n_k}^d)^{n_k} \\
&\geq \liminf_{k\to\infty} \exp\left(\frac{-n_k L_x r_{n_k}^d}{1 - L_x r_{n_k}^d}\right) \\
&\geq e^{-L_x M}, \qquad (1.57)
\end{aligned}$$

donde hemos usado que $(1-z)^n \geq \exp(-nz/(1-z))$ $z \in [0,1)$, y que $\lim_{k\to\infty} r_{n_k}^d = 0$. Aplicando entonces el Lema de Fatou

$$\begin{aligned}
\lim_{k\to\infty} E\big(d_\mu(S, S_{n_k})\big) &= \lim_{k\to\infty} \int_S \psi_{n_k}(x)\mu(dx) \\
&= \liminf_{k\to\infty} \int_S \psi_{n_k}(x)\mu(dx) \geq \int_S \liminf_{k\to\infty} \psi_{n_k}(x)\mu(dx) > 0. \qquad (1.58)
\end{aligned}$$

Lo cual contradice que $E\big(d_\mu(S, S_n)\big) \to 0$. Por lo tanto $\{nr_n^d\}$ converge a 0. □

En [16] se puede encontrar el cálculo del orden exacto de $E\big(d_\mu(S, S_n)\big)$, cálculo hecho a partir de algunos supuestos adicionales sobre S.

1.8. Estimación del Número de Clusters

El número, q, de c-clusters de una densidad f en $\mathbb{R}^d$ fue definido por Hartigan en 1975 como en número de componentes conexas de $\{f > c\}$ donde c es un número real dado. Este capítulo pretende ser una pequeña introducción al tema de estimar q, y se basa en el trabajo de Cuevas, Febrero y Fraiman, ver [8]. Los métodos propuestos están basados en la estimación de $\{f > c\}$ por $\{\hat{f}_n > c\}$ donde $\hat{f}_n$ es un estimador basado en núcleos. Dado que estimar las componentes conexas de $\{\hat{f}_n > c\}$ puede ser complicado, lo que se hace es tomar un estimador de Devroye y Wise para los $X_i \in \{\hat{f}_n > c\}$. Vamos a definir el estimador con mas detalle.

1.8.1. El estimador

Consideremos $X_1, X_2, \dots, X_n$ observaciones de una densidad f en $\mathbb{R}^d$ y $\hat{f}_n$ un estimador no paramétrico de f. Usaremos la notación $T(S)$ para el número q mencionado anteriormente, de componentes conexas de S siendo $S = S(f;c) = \{f > c\}$. Dado que calcular $T\big\{S(\hat{f}_n;c)\big\}$ presenta problemas prácticos obvios, reemplazamos $\{\hat{f}_n > c\}$ por el estimador, más simple

$$\hat{S}_n = \bigcup_{i=1}^{k_n} B(X_i, \varepsilon_n), \tag{1.59}$$

donde $X_i, i = 1, \dots, k_n$ denotan los $X_i \in \{\hat{f}_n > c\}$, (por lo tanto k_n es aleatorio), donde ε_n es un parámetro que se elige dependiendo del problema en cuestion. Por lo tanto el estimador es

$$T_n = T(\hat{S}_n) \tag{1.60}$$

Es suficiente tomar $\varepsilon_n = \varepsilon$ siendo ε menor que la mínima distancia entre las componentes de S, aunque es claro que elegir ε_n muy pequeño da como resultado una sobre estimacion de $T(S)$.
Dado que, si k_n es pequeño, obtendremos una mala estimación de $T\big(\{f_n > c\}\big)$, podemos considerar, si N es un número natural dado, un remuestreo $Z_1, \dots, Z_N$ de observaciones de $\hat{f}_n$, condicionadas a $\{\hat{f}_n > c\}$. Esto se hace tomando observaciones de $\hat{f}_n$ y rechazando aquellas que no verifican $\hat{f}_n(Z_i) > c$ hasta obtener N. Observese que, dada $\hat{f}_n$, puede aproximarse $\{\hat{f}_n > c\}$ tanto como se quiera, es decir N puede tomarse arbitrariamente grande. Nótese además que dada la simplicidad de la forma de $\hat{f}_n$, es fácil obtener observaciones de ella. A partir de $Z_1, \dots, Z_N$ construimos $T^0_{n,N} = T(\hat{S}^0_{n,N})$, donde $\hat{S}^0_{n,N}$ se define reemplazando los $X_1, \dots, X_{k_n}$ por $Z_1, \dots, Z_N$.
Otra forma, que da además una idea de la varianza de T_n, es considerar B remuestreos de tamaño n de $\hat{f}_n$ siendo B un número natural dado, y tomar para cada remuestreo, los $X^0_i \in \{\hat{f}_n > 0\}$. Luego el estimador es

$$T^0_n = T(\hat{S^0_n}), \tag{1.61}$$

donde $\hat{S}^0_n$ es unión de bolas de radio ε centradas en los X^0_i.
Es decir lo que se obtienen son B replicas independientes, de T_n. En el trabajo de Cuevas, Febrero y Fraiman se propone un algoritmo para calcular T_n que aquí omitiremos, y puede verse en [8]. Veamos ahora los resultados asintóticos de ambos estimadores, y para eso las siguientes hipótesis serán necesarias.

(S1) Existe un entorno U de c tal que, para todo $a \in U, T\big(\{f > a\}\big)$ es finito y constante ($T\big(\{f > a\}\big) = q$), y las componentes conexas de $\{f > a\}$ están separadas, es decir $D := \text{ínf}\{D_a : a \in U\} > 0$, donde

$$D_a = \text{mín}\left\{ \inf_{x\in A_1, y\in A_2} \|x - y\| : A_1 \text{ y } A_2 \text{ componentes conexas de } \{f > a\} \quad \text{con } A_1 \neq A_2 \right\} \tag{1.62}$$

(S2) $\{f > a\}$ es acotado, y $P(f(X) > a) \in (0,1)$ para todo a en algun entorno U de c. Supondremos además que existe $p > 0$ tal que para todo ρ tal que $c \pm \rho \in U$,

$$\{f > c - \rho\} \subset \{f > c + \rho\}^{p\rho}, \tag{1.63}$$

donde, dado $A \subset \mathbb{R}^d$ y $t > 0$, A^t denota el conjunto $A \oplus tB$.

(S3) Para todo a en un entorno U de c, el conjunto $\{f > a\}$ es estándar.

(S4) Se verifica que

(i) $\mu_L\big(\{f > c\}^t\big) - \mu_L\big(\{f > c\}\big) = \mathcal{O}(t)$, cuando $t \to 0$.

(ii) La función $h(s) = \mu_L\big(\{f > s\}\big)$ es continua en c.

(E1) $\hat{f}_n$ es un estimador de densidad de f tal que para alguna sucesión $\beta_n \to 0$,

$$\sup_{x \in C} |\hat{f}_n(x) - f(x)| = \mathcal{O}(\beta_n) \quad c.s.$$

donde C es un conjunto compacto tal que $\{f > a\} \subset C$ para todo $a \in U$, donde U es un entorno de c tal que *(S1)*,*(S2)* y *(S3)* se verifican.

(E2) La sucesión de parámetros ε_n en el estimador auxiliar $\hat{S}_n$ definido en 1.59 satisface una de las siguientes condiciones

(i) $\varepsilon_n \to 0$, y $n\varepsilon_n^d \to \infty$, además

$$\beta_n / \varepsilon_n \to 0, \tag{1.64}$$

(ii) la sucesión ε_n es constante, $\varepsilon_n = \varepsilon$, con $0 < \varepsilon < D/2$, siendo D como en *(S1)*

Teorema 1.52. *Sean $X_1, X_2, \ldots, X_n$ observaciones de una densidad f en $\mathbb{R}^d$, desconocida. Definamos $S = \{f > c\}$, donde $c > 0$ es un número real dado, denotemos como $T(S)$ el número de c-clusters (esto es, el número de componentes conexas de S). Si $\hat{f}_n$ es un estimador de densidad, no paramétrico, de f, sea T_n el estimador de $T(S)$ definido en 1.60. Supongamos (S1),(S2),(S3), (E1) y (E2). Entonces*

a) $T_n \to T(S)$, *c.s.*

b) Denotemos T_n^0 el estimador definido en 1.61 basado en un remuestreo $X_1^0, X_2^0, \ldots, X_n^0$ de $\hat{f}_n$. Denotemos como $\hat{P}_n$ la medida de probabilidad correspondiente a $\hat{f}_n$, entonces

$$\sum_{n=1}^{\infty} \hat{P}_n\{T_n^0 \neq T_n\} < \infty, \quad c.s. \tag{1.65}$$

en particular $T_n^0 = T_n = T(S)$ a partir de cierto n c.s., c.s.
Los mismos resultados valen cuando se remplaza T_n^0 por $T_{n,N}^0$ para $N \geq n$.

c) Bajo (S1),(S2),(S3),(E1),(E2)-(i) y (S4) se cumple que

$$\mu_L\big(\hat{S}_n \triangle \{f > c\}\big) \to 0, \quad c.s.,$$

donde $C \triangle D$ denota la diferencia simétrica de los conjuntos C y D.

Demostración. Tomemos U un entorno de c de modo que se cumplan las hipótesis *(S1)-(S3)*. Supongamos primero que se verifica *(E2)-(i)*. La hipótesis *(E1)* implica que para n suficientemente grande

$$\{f > c + \beta_n'\} \subset \{\hat{f}_n > c\} \subset \{f > c - \beta_n'\}, \quad c.s., \tag{1.66}$$

donde β_n' es una sucesión de orden $\mathcal{O}(\beta_n)$. A lo largo de la demostración supondremos que n es tal que 1.66 se verifica. Tomemos $\varepsilon_n < D/2,\ p\beta_n' < \varepsilon_n/4,\ c \pm \beta_n' \in U$. De 1.66 obtenemos que $\{\hat{f}_n \geq c\}$ tiene q componentes conexas, una por cada componente de $\{f > c \pm \beta_n'\}$, cuya distancia entre si es mayor o igual que D. Obsérvese que, siempre que $\varepsilon_n < D/2$ tenemos que $T_n \geq q$, $\quad c.s.$ para n suficientemente grande.

Consideremos un cubrimiento de $\{\hat{f}_n > c\}$, $\big\{B(z_j, \varepsilon_n/2) : j = 1, \dots, R(\varepsilon_n)\big\}$, por bolas de radio $\varepsilon_n/2$ centradas en puntos $z_j \in \{\hat{f}_n > c\}$. Obsérvese que podemos tomar un cubrimiento de modo que $R(\varepsilon_n) \leq M\varepsilon_n^{-d}$ para algún $M > 0$.

Por construcción de $\hat{S}_n$ y la inclusión en 1.66, se cumple que: si para cada centro z_j existe un punto de la muestra $X_i \in \{\hat{f}_n > c\}$ tal que $\|X_i - z_j\| < \varepsilon_n/2$, entonces $\hat{S}_n$ tiene necesariamente q componentes conexas. Entonces

$$\begin{aligned}
P(T_n \neq q) &\leq P\big(\exists j, 1 \leq j \leq R(\varepsilon_n) : \forall i, 1 \leq i \leq n, X_i \notin B(z_j, \varepsilon_n/2) \cap \{\hat{f}_n > c\}\big) \\
&\leq \sum_{j=1}^{R(\varepsilon_n)} \prod_{i=1}^{n} P\big(X_i \notin B(z_j, \varepsilon_n/2) \cap \{\hat{f}_n/2 > c\}\big) \\
&\leq \sum_{j=1}^{R(\varepsilon_n)} \prod_{i=1}^{n} P\big(X_i \notin B(z_j, \varepsilon_n/2) \cap \{f > c + \beta_n'\}\big) \qquad (1.67)
\end{aligned}$$

Observemos que, dado que $z_j \in \{\hat{f}_n > c\}$ y $\{\hat{f}_n > c\} \subset \{f > c + \beta_n'\}^{p\beta_n'}$ con $p\beta_n' < \varepsilon_n/4$, existe $w_j \in \{f > c + \beta_n'\}$ tal que $B(w_j, \varepsilon_n/4) \subset B(z_j, \varepsilon_n/2)$.
Por lo tanto, por *(S3)*

$$P\big(X_i \in B(z_j, \varepsilon_n/2) \cap \{f > c + \beta_n'\}\big) \geq c\delta\mu_L\big(B(0,1)\big)4^{-d}\varepsilon_n^d = \gamma\varepsilon_n^d.$$

Por lo tanto, usando ésto en 1.67

$$\sum_{n=1}^{\infty} P(T_n \neq q) \leq \sum_{n=1}^{\infty} R(\varepsilon_n)(1 - \gamma\varepsilon_n^d)^n \leq \sum_{n=1}^{\infty} M\varepsilon_n^{-d} e^{-n\gamma\varepsilon_n^d} < \infty.$$

Luego, una condición suficiente para que $\sum_{n=1}^{\infty} P(T_n \neq q)$,

$$\sum_{n=1}^{\infty} \exp\Big\{ - \gamma\big(n\varepsilon_n^d - \gamma^{-1}\log(\varepsilon_n^{-d})\big)\Big\} < \infty,$$

y ésta serie converge si

$$\frac{\log n}{n\varepsilon_n^d - \gamma^{-1}\log(\varepsilon_n^{-d})} \to 0.$$

Esto es equivalente a que

$$\frac{\log n}{n\varepsilon_n^d - \log(\varepsilon_n^d)} \to 0,$$

lo cual se verifica bajo *(E2)-(i)*.

Cuando en lugar de *(E2)-(ii)* suponemos *(E2)-(i)*, la demostración de la parte *(a)* es más simple. Las ideas principales de la prueba anterior pueden ser fácilmente adaptadas a este caso, sin necesidad de hacer hipótesis sobre el orden β_n de consistencia de $\hat{f}_n$.

En líneas generales, la demostración de la parte *(b)* se hace de forma similar a la anterior, reemplazando P por $\hat{P}$ en 1.67 para el caso de T_n^0 o considerando la distribución condicional correspondiente, en el caso de $T_{n,N}^0$.

Para demostrar *(c)*, consideremos $a_1, a_2 \in U$ tal que $a_1 < c < a_2$. Denotemos por

$$S_n^{(1)} = \bigcup_{\{X_i : f(X_i) > a_1\}} B(X_i, \varepsilon_n) \quad y \quad S_n^{(2)} = \bigcup_{\{X_i : f(X_i) > a_2\}} B(X_i, \varepsilon_n).$$

Dado *(E1)*, tenemos que

$$S_n^{(2)} \subset \hat{S}_n \subset S_n^{(1)}, \text{ a partir de cierto } n \text{ } c.s.$$

Con un argumento similar al que se usa en el Teorema 1.34 se prueba que

$$\mu_L\big(S_n^{(j)} \triangle \{f > a_1\}\big) \to 0, \text{ cuando } n \to \infty, \quad c.s., \text{ para } j = 1, 2.$$

Finalmente, el resultado se sigue de *(S4)* ya que

$$\mu_L\big(\hat{S}_n \triangle \{f > c\}\big) \leq \mu_L\big(S_n^{(1)} \triangle \{f > a_1\}\big) + \\ \mu_L\big(S_n^{(2)} \triangle \{f > a_1\}\big) + 2\mu_L\big(\{f > a_1\} \triangle \{f > a_2\}\big),$$

y el resultado se sigue de aplicar *(S4)*. □

Capítulo 2

Estimación de longitudes, áreas y volúmenes

2.1. Introducción

En el capítulo anterior vimos diferentes formas de estimar conjuntos, y sus bordes, a partir de una cantidad finita de puntos en ellos, así como la estimación de conjuntos de nivel de densidades. Y observamos que si bien estos pueden estar próximos en la distancias consideradas, eso no implicaba necesariamente que su forma fuese similar. En este capítulo trataremos el tema de cómo estimar por ejemplo la longitud de su borde (para conjuntos en el plano), o en general, su superficie lateral. Estos datos si bien van en la dirección de aportar mas información a la comprensión de la forma del conjunto, son de interés en si mismo. Es intuitivo que la proximidad de un conjunto a otro, e incluso la de sus bordes, no implica necesariamente la proximidad de la longitud (para el caso de conjuntos en el plano) de sus bordes, y por lo tanto no necesariamente es consistente el tomar como estimador de la longitud del conjunto, a la longitud de estimadores del conjunto. El estimador de Devroye y Wise es un buen ejemplo de eso. Veamos primero, a partir del concepto de *contenido de Minkowski,* la forma en que mediremos longitudes, áreas, etc. Asumiremos, a lo largo de este capítulo, salvo que se especifique lo contrario, que los conjuntos serán compactos, iguales a la clausura de su interior.

Definición 2.1 (Conjunto Rectificable)**.** Sea $E \subset X$ donde X es un espacio métrico, decimos que E es *m-rectificable* si existe una función de Lipschitz de un subconjunto acotado de $\mathbb{R}^m$ sobre E.

Definición 2.2 (Medida de Hausdorff de un conjunto)**.** Sean X un espacio métrico, $p \geq 0$, $\delta > 0$ números reales, y $A \subset X$ definimos:

$$H_{p,\delta}(A) = \text{ínf} \left\{ \sum_{j=1}^{\infty} \big(diam(B_j)\big)^p : A \subset \bigcup_j B_j, diam(B_j) \leq \delta \right\},$$

donde convenimos en que $\text{ínf}\, \emptyset = \infty$. Es claro que $H_{p,\delta}(A)$ decrece cuando δ tiende a 0.

El número

$$H_p(A) = \lim_{\delta \longrightarrow 0} H_{p,\delta}(A),$$

se denomina *medida de Hausdorff p-dimensional de A*

Definición 2.3 (Contenido de Minskowski)**.** Sean $S \subset \mathbb{R}^d$ no vacío y acotado, y $\varepsilon > 0$ un número real. Definimos el contenido inferior de Minkowski, de dimensión m $(m \leq d)$, de S como:

$$\mathscr{M}_{*m}(S)) = \liminf_{\varepsilon \to 0} \frac{\mu_L\{x : d(x, S) \leq \varepsilon\}}{\varepsilon^{d-m}\alpha(d-m)},$$

y el superior como:

$$\mathscr{M}^{*m}(S)) = \limsup_{\varepsilon \to 0} \frac{\mu_L\{x : d(x, S) \leq \varepsilon\}}{\varepsilon^{d-m}\alpha(d-m)},$$

donde μ_L es la medida de Lebesgue en $\mathbb{R}^d$ y $\alpha(d-m)$ es la medida de Lebesgue de la bola de centro 0 y radio 1 en $\mathbb{R}^{d-m}$. Si ambos valores coinciden, dicho valor se denomina contenido de Minkowski de dimensión m de S y se anota $\mathscr{M}^m(S)$. Si G es un subconjunto compacto de $\mathbb{R}^d$ e igual a la clausura de su interior, denotaremos $L_0(G)$ al contenido de Minkowski de dimensión $d-1$ de ∂G en caso de que éste exista. Es decir:

$$L_0(G) = \lim_{\varepsilon \to 0} \frac{\mu_L\{x : d(x, \partial G) \leq \varepsilon\}}{2\varepsilon}. \tag{2.1}$$

En [14] se encuentra una prueba del siguiente teorema:

Teorema 2.4. *Si E es un conjunto cerrado de $\mathbb{R}^d$ m-rectificable entonces:*

$$\mathscr{M}^m(E) = H_m(E).$$

Definición 2.5 (Alcance de un conjunto)**.** Se define el *alcance* de un conjunto $G \subset \mathbb{R}^d$ como el supremo de los $\varepsilon > 0$ (posiblemente infinito) tal que si $x \in \mathbb{R}^d$ y la distancia de x a G (definida como $\inf_{a \in G} d(x, a)$) es menor que ε existe un único $a \in G$ que está a esa distancia.

En [16] se demuestra el siguiente lema, que relaciona los conceptos de alcance de un conjunto y de rodamiento lbire de una bola.

Lema 2.6. *Sea $A \subset \mathbb{R}^d$ un conjunto compacto no vacío. Supongamos que una bola de radio $\alpha > 0$ rueda libremente en A y $\overline{A^c}$. Entonces A y $\overline{A^c}$ son de alcance positivo, siendo ambos alcances mayores o iguales que α.*

En [13] se demuestra que si G es un conjunto con alcance r_0 positivo la medida de Lebesgue de $G \oplus \varepsilon B$ coincide localmente ($\varepsilon \in (0, r_0)$) con un polinomio de grado a lo sumo d, cuyo término independiente es la medida de Lebesgue de G.

Así como sucedía en la estimación de conjuntos, tener algo de informacion respecto del conjunto, por ejemplo saber que es α-convexo, permite usar estimadores de la longitud, área, etc, que usen dicha información y permitan obtener velocidades de convergencia mejores que para los casos en que no se dispone de tal información. Veremos primero el trabajo de Cuevas, Fraiman, y Rodriguez-Casal, ver [5], para un enfoque general del problema, y luego el trabajo de Pateiro, ver [16], para el caso en que el conjunto es α-convexo.

2.2. Sin restricciones de forma

2.2.1. El Estimador

Vamos a considerar de aquí en adelante un subconjunto G de $[0,1]^d$ tal que su contenido de Minkowski esta bien definido, es estrictamente positivo y finito, y observaciones independientes identicamente distribuidas $(Z_1,\delta_1),\ldots,(Z_n,\delta_n)$ de una variable aleatoria (Z,δ) con distribución uniforme en $[0,1]^d$, donde $\delta=1$ si $Z\in G$, $\delta=0$ si $Z\in[0,1]^d\backslash G$. En adelante denotaremos por $R=[0,1]^d\backslash G$, P_X y P_Y la distribución condicional de G y R respectivamente, es decir, la distribución de $Z_{|\{\delta=1\}}$ y $Z_{|\{\delta=0\}}$. Observemos que tanto P_X como P_Y son uniformes en G y R respectivamente. Dado un $z\in[0,1]^d$ y $\varepsilon>0$ denotaremos por $G_z(\varepsilon)$ el número de observaciones que pertenecen a la bola $B(z,\varepsilon)$

$$G_z(\varepsilon)\equiv G_{n,z}(\varepsilon)=\sum_{i=1}^{n}\mathbb{I}_{\{\delta_i=1,\|Z_i-z\|\le\varepsilon\}},$$

y analogamente $R_z(\varepsilon)$

$$R_z(\varepsilon)\equiv R_{n,z}(\varepsilon)=\sum_{i=1}^{n}\mathbb{I}_{\{\delta_i=0,\|Z_i-z\|\le\varepsilon\}}.$$

Observación 2.7. *$G_z(\varepsilon)$ tiene distribución binomial con parámetros n y $p_X(z,\varepsilon)=P\big(\|Z-z\|\le\varepsilon,\delta=1\big)=\mu_L^d(G)P_X\big(B(z,\varepsilon)\big)$. Asimismo $R_z(\varepsilon)$ tiene distribución binomial con parámetros n y $p_Y(z,\varepsilon)=\big(1-\mu_L^d(G)\big)P_Y\big(B(z,\varepsilon)\big)$.*

Definición 2.8. Tomemos $\{\varepsilon_n\}$ una sucesión de números reales positivos, que converja a 0 cuando n tiende a infinito. Sea $T=\partial G$. El estimador que usaremos para $T\oplus\varepsilon_n B$ será:

$$T_n=\big\{z\in[0,1]^d : R_z(\varepsilon_n)\ge 1 \quad \text{y} \quad G_z(\varepsilon_n)\ge 1\big\}.$$

Como se dijo anteriormente, para el estimador propuesto nos interesan aquellos puntos tales que un entorno de radio ε_n interceptan a G y a R. Y para estimar $L_0=L_0(G)$:

$$L_n=\frac{\mu_L(T_n)}{2\varepsilon_n}.$$

Es claro que la convergencia de L_n a L_0 dependerá de una “buena” elección de ε_n así como también, de imponerle a G cierta regularidad mínima.

2.2.2. Consistencia del estimador

Teorema 2.9. *Supongamos que:*

a) G y R son estandar en T, es decir, existe una constante $C > 0$ tales que, para ε suficientemente pequeño:

$$P_X\big(B(t,\varepsilon)\big) \geq C\mu_L\big(B(t,\varepsilon)\big) \quad y \quad P_Y\big(B(t,\varepsilon)\big) \geq C\mu_L\big(B(t,\varepsilon)\big) \quad \textit{para todo} \quad t \in T.$$

b) La sucesión $\{\varepsilon_n\}$ cumple que:

$$\varepsilon_n \longrightarrow 0 \quad y \quad \frac{n\varepsilon_n^d}{\log(n)} \longrightarrow \infty,$$

entonces

$$L_n = \frac{\mu_L(T_n)}{2\varepsilon_n} \longrightarrow L_0, \quad c.s..$$

Demostración. La demostración es consecuencia de las siguientes afirmaciones:

Afirmación 1. $T_n \subset T \oplus \varepsilon_n B$ c.s..

Afirmación 2. Para todo $0 < \alpha < 1$ tenemos que c.s., a partir de un cierto n

$$T \oplus \varepsilon_n' B \subset T_n,$$

donde $\varepsilon_n' = \alpha\varepsilon_n, \quad 0 < \alpha < 1.$

Demostración de la afirmación 1. Si $z \in T_n$ entonces $B(z,\varepsilon_n)$ interseca a G y a su complemento R, por lo tanto $B(z,\varepsilon_n)$ interseca el borde de G, T, por lo tanto $z \in T \oplus \varepsilon_n B$, lo cual concluye la prueba de la afirmación 1.

Demostración de la afirmación 2. Por el Lema de Borel-Cantelli es suficiente demostrar que:

$$\sum_{n=1}^{\infty} P(T \oplus \varepsilon_n' B \not\subseteq T_n) < \infty.$$

Como

$$\begin{aligned} P(T \oplus \varepsilon_n' B \not\subseteq T_n) \;&\leq\; P\big(\exists z \in T \oplus \varepsilon_n' B : G_z(\varepsilon_n) = 0\big) \\ &\quad + P\big(\exists z \in T \oplus \varepsilon_n' B : R_z(\varepsilon_n) = 0\big), \end{aligned} \tag{2.2}$$

intentaremos encontrar cotas para las probabilidades del lado derecho de la última desigualdad. Si $z \in T \oplus \varepsilon_n'$ existe un $t \in T$ para el cual $B(t,\beta_n) \subset B(z,\varepsilon_n)$ donde $\beta_n = (1-\alpha)\varepsilon_n$. Por lo tanto:

$$P\big(\exists z \in T \oplus \varepsilon_n' B : G_z(\varepsilon_n) = 0\big) \leq P\big(\exists t \in T : G_t(\beta_n) = 0\big).$$

Denotemos $T(\beta_n)$ el conjunto (con cardinal $D(\beta_n)$) de los centros correspondientes a un cubrimiento minimal de T por bolas de radio $\beta_n/2$. Es decir, consideraremos el conjunto $\big\{B(s,\beta_n/2) : s \in T(\beta_n) \subset T\big\}$ tal que:

$$T \subset \bigcup_{s\in T(\beta_n)} B\left(s, \frac{\beta_n}{2}\right).$$

Como $\big\{\exists t \in T : G_t(\beta_n) = 0\big\} \subset \big\{\exists s \in T(\beta_n) : G_s(\beta_n/2) = 0\big\}$ tenemos:

$$\begin{aligned}
P(\exists t \in T : G_t(\beta_n) = 0) &\leq P\left(\exists s \in T(\beta_n) : G_s\left(\frac{\beta_n}{2}\right) = 0\right) \\
&\leq \sum_{s\in T(\beta_n)} P\left(G_s\left(\frac{\beta_n}{2}\right) = 0\right) \\
&= \sum_{s\in T(\beta_n)} \left(1 - p_X\left(s, \frac{\beta_n}{2}\right)\right)^n \\
&\leq \sum_{s\in T(\beta_n)} \exp\left(-np_X\left(s, \frac{\beta_n}{2}\right)\right).
\end{aligned}$$

En la última desigualdad hemos usado que $(1-x) \leq e^{-x}$ para $0 \leq x \leq 1$. El lado derecho de la desigualdad anterior puede ser fácilmente acotado usando la hipótesis de que G y R son estándar, para n suficientemente grande tenemos que:

$$p_X\left(s, \frac{\beta_n}{2}\right) \geq C\alpha(d)\mu_L(G)\frac{\beta_n^d}{2^d} = K_1\varepsilon_n^d,$$

donde $\alpha(d) = \mu_L(B(0,1))$ y K_1 es una constante que depende de d, α, $\mu_L(G)$ y C. Por lo tanto:

$$P\big(\exists z \in T \oplus \varepsilon'_n B : G_z(\varepsilon_n) = 0\big) \leq D(\beta_n)\exp(-nK_1\varepsilon_n^d).$$

Para acotar $D(\varepsilon)$ recordemos que éste representa el cardinal de un cubrimiento mínimo $\mathscr{C}(\varepsilon/2)$ de T por bolas de radio $\varepsilon/2$. Observar que las bolas de radio $\varepsilon/4$ con los mismos centros que en $\mathscr{C}(\varepsilon/2)$ son disjuntas, (de lo contrario $\mathscr{C}(\varepsilon/2)$ no sería minimal). Entonces su medida de Lebesgue debe ser menor o igual que $\mu_L\big(T \oplus (\varepsilon/4)B\big)$, luego:

$$D(\varepsilon)(\varepsilon/4)^d\alpha(d) \leq \mu_L\big(T \oplus (\varepsilon/4)B\big).$$

Como estamos suponiendo que existe el contenido de Minkowski de ∂G y es finito, tenemos que $D(\varepsilon) \leq A\varepsilon^{1-d}$ para alguna constante A y por lo tanto:

$$P\big(\exists z \in T \oplus \varepsilon'_n B : G_z(\varepsilon_n) = 0\big) \leq K_2\varepsilon_n^{1-d}\exp(-K_1 n\varepsilon_n^d),$$

donde $K_2 = (1-\alpha)^{1-d}A$. La condición $n\varepsilon_n^d/log(n) \to \infty$ nos asegura la convergencia de la serie $\sum_{n=1}^{\infty} \varepsilon_n^{1-d}\exp(-K_1 n\varepsilon_n^d)$. La otra probabilidad en (3.1) se acota de forma análoga, lo que concluye la prueba de la afirmación 2. El teorema 1, es consecuencia directa de las afirmaciones 1 y 2, tenemos que, con probabilidad 1,

$$\alpha L_0 = \lim_n \frac{\mu_L(T \oplus \varepsilon_n' B)}{2\varepsilon_n} \leq \liminf_n L_n \leq \limsup_n L_n \leq \lim_n \frac{\mu_L(T \oplus \varepsilon_n B)}{2\varepsilon_n} = L_0.$$

Como se cumple para todo $\alpha \in (0,1)$ se sigue la conclusión del teorema. □

2.2.3. Una cota para $E(L_n)$

Consideremos la función $L(\varepsilon) = \mu_L(T \oplus \varepsilon B)/(2\varepsilon)$, *la cual es una aproximación de* L_0. *Mostraremos que bajo ciertas condiciones de regularidad en* G *y* R *se cumple que* $|L(\varepsilon) - L_0| = \mathcal{O}(\varepsilon)$. *Observemos que:*

$$T \oplus \varepsilon B = (G \oplus \varepsilon B) \cap (R \oplus \varepsilon B),$$

de donde:

$$\mu_L(T \oplus \varepsilon B) = \mu_L(G \oplus \varepsilon B) + \mu_L(R \oplus \varepsilon B) - \mu_L([0,1]^d \oplus \varepsilon B).$$

De esta última identidad, y por lo expuesto en la sección anterior, sabemos que, si G *tiene alcance positivo, la medida de Lebesgue de* $T \oplus \varepsilon B$ *coincide localmente (* $\varepsilon \in (0, r_0)$ *) con* $P(\varepsilon)$ *un polinomio de grado a lo sumo* d *con término independiente igual a 0. Observemos que (en vistas de que asumimos que el* L_0 *es finito) el coeficiente en* ε *de* $P(\varepsilon)$ *coincide con* $2L_0$, *por lo tanto* $L(\varepsilon) - L_0$ *es un polinomio en* ε *con término independiente nulo. En particular obtenemos que* $L(\varepsilon)$ *es diferenciable en* $\varepsilon = 0$.
Como se ve en la prueba del teorema 3.1, tenemos, con probabilidad 1, que $T_n \subset T \oplus \varepsilon B$ *a partir de un cierto* n *y por lo tanto:*

$$L_n \leq L(\varepsilon_n) \quad c.s.. \tag{2.3}$$

Lo cual significa que L_n *aproxima a* L_0 *inferiormente, cuando el valor de* $L(\varepsilon)$ *es próximo a* $L(0) = L_0$.

Teorema 2.10. *Supongamos que se cumplen la condición a) del teorema anterior, que la función* $F(\varepsilon) := \mu_L(T \oplus \varepsilon B)$ *es diferenciable en un entorno de 0 y su derivada es continua en 0. Entonces*

$$E(L_n) \geq L(\varepsilon_n) - I_n, \tag{2.4}$$

donde

$$I_n = \frac{1}{\varepsilon_n} \int_{T \oplus \varepsilon_n B} \exp\Big(-Kn\big(\varepsilon_n - d(z,T)\big)^d \Big) dz,$$

siendo K *una constante positiva y* $d(z,T) = \inf\{\|z - t\| : t \in T\}$.
Además

$$I_n = \mathcal{O}\big((n\varepsilon_n^d)^{-1/d}\big).$$

Observación 2.11. *De acuerdo a lo comentado anteriormente es suficiente que* G *tenga alcance positivo para que se cumpla la condición sobre* $F(\varepsilon)$.

Demostración. La esperanza de L_n puede ser escrita como:

$$\begin{aligned} E(L_n) &= \frac{E\big(\mu_L(T_n)\big)}{2\varepsilon_n} = \frac{1}{2\varepsilon_n} E\left(\int \mathbb{I}_{\{z\in T_n\}}\mu_L(dz)\right) = \frac{1}{2\varepsilon_n}\int E(\mathbb{I}_{\{z\in T_n\}})\mu_L(dz) \\ &= \frac{1}{2\varepsilon_n}\int P(z\in T_n)\mu_L(dz) = \frac{1}{2\varepsilon_n}\int_{T\oplus\varepsilon_n B} P(z\in T_n)\mu_L(dz), \end{aligned}$$

donde hemos usado en la ultima igualdad que, con probabilidad 1, $T_n \subset T\oplus\varepsilon_n B$. Es claro que:

$$P(z\notin T_n) \leq P\big(G_z(\varepsilon_n)=0\big) + P\big(R_z(\varepsilon_n)=0\big). \tag{2.5}$$

Como $G_z(\varepsilon_n)$ tiene distribución binomial con parámetros n y $p_X(z,\varepsilon_n)$ tenemos que

$$P\big(G_z(\varepsilon_n)=0\big) = \big(1-p_X(z,\varepsilon_n)\big)^n \leq \exp\big(-np_X(z,\varepsilon_n)\big).$$

Tomemos $P_T z\in T$ la proyección de z sobre T. Entonces, para todo $z\in T\oplus\varepsilon_n B$

$$B\big(P_T z, \varepsilon_n - d(z,T)\big) \subset B(z,\varepsilon_n).$$

Usando la condición a) del teorema 3.1 tenemos que para ε_n suficientemente pequeño,

$$p_X\big(B(z,\varepsilon_n)\big) \geq C\alpha(d)\big(\varepsilon_n - d(z,T)\big)^d,$$

y por lo tanto

$$P(G_z=0) \leq \exp\Big(-K_1 n\big(\varepsilon_n - d(z,T)\big)^d\Big),$$

donde K_1 es una constante positiva que depende solamente de $\mu_L(G)$, C y d. Analogamente obtenemos que $P(R_z=0)\leq \exp(-K_2 n(\varepsilon_n - d(z,T))^d)$, para cierta constante positiva K_2 que solo depende de $\mu_L(R)$, C y d. Usando estas dos acotaciones y (3.4) obtenemos:

$$P(z\in T_n) \geq 1 - 2\exp\Big(-Kn\big(\varepsilon_n - d(z,T)\big)^d\Big),$$

donde $K = \text{mín}(K_1,K_2)$, finalmente

$$\begin{aligned} E(L_n) &= \frac{1}{2\varepsilon_n}\int_{T\oplus\varepsilon_n B} P(z\in T_n)dz \\ &\geq \frac{1}{2\varepsilon_n}\int_{T\oplus\varepsilon_n B} 1 - 2\exp\Big(-Kn\big(\varepsilon_n - d(z,T)\big)^d\Big)dz \\ &= L(\varepsilon_n) - \frac{1}{\varepsilon_n}\int_{T\oplus\varepsilon_n B} \exp\Big(-Kn\big(\varepsilon_n - d(z,T)\big)^d\Big)dz = L(\varepsilon_n) - I_n, \end{aligned}$$

si escribimos

$$I_n = \frac{1}{\varepsilon_n}\int_{T\oplus\varepsilon_n B} g_n\big(d(z,T)\big)dz,$$

siendo $g_n(w) = \exp\big(-Kn(\varepsilon_n - w)^d\big)$. Haciendo un cambio de variable obtenemos

$$I_n = \frac{1}{\varepsilon_n}\int_0^{\varepsilon_n} g_n(w)F(dw),$$

donde $F(w) = \mu_L\{z : d(z,T) \leq w\}$. Como supusimos que F es diferenciable en 0 y que L_0 existe y es finito, tenemos que $F'(0) = 2L_0$, además, como $F'(0)$ es continua, para w suficientemente pequeño se cumple que $F'(w) \leq 3L_0$. Finalmente para n suficientemente grande,

$$\begin{aligned} I_n &\leq \frac{3L_0}{\varepsilon_n}\int_0^{\varepsilon_n} \exp(-Knt^d)dt = \frac{3L_0}{\varepsilon_n}\int_0^{Kn\varepsilon_n^d} \exp(-u)\frac{1}{d(Kn)^{1/d}}u^{-(d-1)/d}du \\ &\leq \frac{3L_0}{dK^{1/d}(\varepsilon_n^d n)^{1/d}}\int_0^\infty \exp(-u)u^{-(d-1)/d}du = \frac{A}{(\varepsilon_n^d n)^{1/d}}, \end{aligned}$$

en la primera desigualdad hemos usado $F'(w) \leq 3L_0$ y el cambio de variable $t = \varepsilon_n - w$, en la primera igualdad el cambio de variable $u = Knt^d$, finalmente, la segunda desigualdad se debe a que el integrando es positivo. □

Corolario 2.12. *El siguiente corolario refiere a la convergencia en* $\mathtt{L}_1$ *de* L_n.

a) En las mismas hipótesis que en el teorema anterior, tenemos que:

$$E|L_n - L_0| \leq I_n + |L(\varepsilon_n) - L_0|, \tag{2.6}$$

es decir, si $\varepsilon_n \longrightarrow 0$ *y* $n\varepsilon_n^d \longrightarrow \infty$ *se cumple* $E|L_n - L_0| \longrightarrow 0$.

b) Si además asumimos que G tiene alcance positivo, tenemos que $|L(\varepsilon_n) - L_0| = \mathcal{O}(\varepsilon)$ *y por lo tanto el orden óptimo para (3.5) es* $\mathcal{O}(n^{-1/(2d)})$ *que se obtiene si* $\varepsilon_n = n^{-1/(2d)}$.

Demostración. a) Aplicando la desigualdad triangular tenemos que

$$E\big(|L_n - L_0|\big) \leq E\big(|L_n - L(\varepsilon_n)|\big) + |L(\varepsilon_n) - L_0| \leq I_n + |L(\varepsilon_n) - L_0|,$$

donde acotamos el primer sumando usando (3.2) y (3.3).

□

Como consecuencia del corolario anterior, tenemos que, en las hipótesis del Corolario parte b)

$$E|L_n - E(L_n)| \leq E\big(|L_n - L(\varepsilon_n)|\big) + |L(\varepsilon_n) - E(L_n)|$$

$$\leq 2I_n = \mathcal{O}\big((n\varepsilon_n^d)^{-1/d}\big),$$

$$L_0 - E(L_n) \leq \big(L(\varepsilon_n) - E(L_n)\big) + \big(L_0 - L(\varepsilon_n)\big)$$

$$\leq \mathcal{O}\big((n\varepsilon_n^d)^{-1/d}\big) + \mathcal{O}(\varepsilon_n).$$

2.3. Un estimador asintóticamente normal

En esta sección veremos resultados asintóticos para el estimador de L_0 ligeramente distinto al de la sección anterior. Tales resultados tienen, (además de ser de importancia teórica) interés desde el punto de vista práctico, ya que nos permiten construir intervalos de confianza, y realizar test de hipótesis. La demostración de estos resultados se basa en el supuesto de que las intensidades de (siguiendo la notación de la sección anterior) los puntos en R y en G son distintas, es decir, supondremos que tendremos mas puntos de un color que de otro. Esto será formalmente definido más adelante. Dichos resultados, y sus aplicaciones, pueden encontrarse también en [2].

El problema de estimar la longitud del borde de un conjunto aparece, en el área de oncología y cardiología, como un requisito para el cálculo de un índice (índice de contorno, que abreviaremos c.i por sus siglas en ingles) que permite medir la gravedad de un área dañada. Tal indice involucra tanto la longitud del borde del conjunto como su volumen, y se define como

$$C_0(G) = \frac{L_0(G)}{\mu(G)^{(d-1)/d}},$$

Más adelante veremos como estimarlo, así como también su consistencia, y normalidad asintótica. En esta sección necesitaremos definir una versión interior del contenido de Minkowski

$$L_0^-(G) = \lim_{\varepsilon \to 0} \frac{\mu\big(B(G^c, \varepsilon) \cap G\big)}{\varepsilon}. \tag{2.7}$$

En el trabajo de [1] se demuestra que si el borde de G es de Lipszchitz entonces $L_0(G) = L_0^-(G)$ y conincide con su medida de Hausdorff $(d-1)-$dimensional.

2.3.1. El estimador

Al igual que en la sección anterior, supondremos que el conjunto a estimar, G, es compacto, y su contenido de Minkowski $L_0 = L_0(G)$ esta bien definido y es finito. Asumiremos que G es el soporte de una medida de probabilidad absolutamente continua y que, sin perdida de generalidad, $G \subset (0,1)^d$. El objetivo es estimar $L_0(G)$ y $C_0(G)$ a partir de dos muestras $\{X_i\}$ e $\{Y_i\}$ donde

$$\{X_i\}_{1 \le i \le n} \text{ son i.i.d, con distribución uniforme en } G^c = [0,1] \setminus G,$$

y

$$\{Y_i\}_{1 \le i \le m} \text{ son i.i.d, con distribución uniforme en } G.$$

Este modelo, al que nos referiremos como *(M2)*, es similar al que empleamos en la sección anterior para estimar $L_0(G)$, que llamaremos *(M1)*. En la práctica podemos usar una versión *(M1')* de *(M1)* para obtener *(M2)* basta tomar (Z_i, δ_i) como en *(M1)*, definiendo $Z_i = X_i$ si $\delta_i = 0$ y $Z_i = Y_i$ en otro caso, hasta obtener m observaciones de G y n de G^c. Es necesario considerar un número aleatorio de observaciones $N = N(n,m) \geq n + m$. La diferencia entre *(M2)* y *(M1')* es más bien formal, bajo *M2)* consideramos dos muestras independientes, de tamaños m y n, mientras que bajo *(M1')* obtenemos $N - m - n$ observaciones extra en $[0,1]^d$. Desde un punto de vista práctico, el modelo *(M1')* es más adecuado ya que en general lo que se tienen son observaciones $Z_i \in [0,1]^d$ y por medio algún tipo de criterio se clasifican las que están en G o en G^c.

Veamos ahora algo de notación que será necesaria para definir los estimadores de $L_0(G)$ y $C_0(G)$. Para $A \subset [0,1]^d$, sea $N_m(A;G)$ la medida de conteo asociada a la muestra Y, es decir

$$N_m(A;G) = \#\{j \geq 1 : Y_j \in A\}.$$

Donde $\#S$ denota el cardinal del conjunto S. Denotaremos como

$$\mu_{m,G}(A) = \mu(G)\frac{N_m(A;G)}{m},$$

la renormalización de la medida empírica, es decir $\mu_{m,G}(A)$ es un estimador de la medida de Lebesgue de la intersección $A \cap G$. Consideremos también el estimador no paramétrico H_n de G^c, para eso tomemos ε_n una sucesión de numeros reales positivos que tienda a 0 y cuya velocidad de convergencia se especificará, y definimos

$$H_n = B\big(\{X_1, \ldots, X_n\}, \varepsilon_n\big).$$

Introduzcamos ahora un estimador de $L_0(G)$ para el caso hipotético de que conocemos $\mu(G)$.

$$A_n = \frac{\mu_{m,G}(H_n)}{\varepsilon_n},$$

el cual estima la ε_n-dilatación interna del borde de G.
Dado que en general $\mu(G)$ es desconocido, lo estimamos como

$$\hat{\mu}(G) = \frac{\#\{i : \delta_i = 1\}}{N},$$

que es consistente por la Ley de los Grandes Números.
Luego, el estimador propuesto para $L_0^-(G)$ es entonces

$$\hat{L}_n = \hat{\mu}(G)\frac{N_m(H_n;G)}{m\varepsilon_n}. \tag{2.8}$$

Y para $C_0(G)$ es

$$\hat{C}_n = \frac{\hat{L}_n}{\hat{\mu}(G)^{(d-1)/d}}. \tag{2.9}$$

2.3.2. Hipótesis

Veamos ahora todas las hipótesis que utilizaremos, más adelante especificaremos cuales son necesarias para la consistencia, y cuales para la normalidad asintótica.

(H0) El contenido de Minkowski L_0 satisface que $L_0(G) = L_0^-(G) < \infty$. Es claro que, cambiando G por G^c, con cambios obvios podríamos haber definido los estimadores para $L^+(G)$.

(H1) El conjunto G^c es estándar en $\Gamma = \partial G$, es decir existe una constante K tal que para todo η suficientemente pequeño, $P\big(X_1 \in B(x,\eta)\big) \geq K\mu\big(B(x,\eta)\big)$ para todo $x \in \Gamma$.

(H2) La sucesión $\{\varepsilon_n\}$ y $m = m(n)$ verifica

$$\varepsilon_n \to 0, \quad \frac{m\varepsilon_n}{\log n} \to \infty \text{ y } \frac{n\varepsilon_n^d}{\log n} \to \infty \quad \text{cuando } n \to \infty.$$

(H3) $|L(\varepsilon) - L_0| = \mathcal{O}(\varepsilon)$ cuando $\varepsilon \to 0$, donde $L(\varepsilon) = \mu\big(B(G^c,\varepsilon) \cap G\big)/\varepsilon$ Es claro que esta condición se cumple cuando $L(\varepsilon)$ es diferenciable en 0 y esto es cierto si por Γ rueda libremente una bola de radio α para algún $\alpha > 0$, como se mencionó en la introducción de éste capítulo.

(H4) La sucesión $m = m(n)$ y ε_n satisfacen que $m\varepsilon_n^3 \to 0$ cuando $n \to \infty$.

(H5) $\frac{1}{n}\left(\frac{m}{\varepsilon_n}\right)^{d/2} \log\left(\frac{m}{\varepsilon_n}\right) \to 0$ cuando $n \to \infty$.
Obsérvese que esta condición implica que el número n de observaciones en G^c es de orden mayor que el de observaciones en G.
Observemos además que las hipótesis requeridas sobre $m(n)$ y ε_n son compatibles, bastaría tomar $m = n^\alpha, \varepsilon_n = n^{-\beta}$ y tomar $\alpha > 0$ y $\beta > 0$ de modo que

$$\alpha + \beta < 2/d \quad y \quad \beta < \alpha < 3\beta. \tag{2.10}$$

Veamos ahora los resultados Teóricos. En lo que respecta a las demostraciones, denotaremos como P la medida en el espacio $\big([0,1]^d\big)^n \times \big([0,1]^d\big)^m$ determinada por la sucesión $(X_1,\ldots,X_n,Y_1,\ldots,Y_m)$, y como E a la esperanza respecto de P. Luego P es el producto de $n+m$ medidas de probabilidad, las primeras n distribuidas uniformemente en G^c y las restantes m con distribución uniforme en G. Las constantes que aparezcan se denotarán con la letra C.

Teorema 2.13. *Bajo (H0), (H1) y (H2) y el modelo (M2) tenemos que*

$$A_n \to L_0 \quad c.s., \text{ cuando } n \to \infty.$$

El mismo resultado vale para $\hat{L}_n$ bajo el modelo (M1').

Demostración. Usaremos *i.o.* como abreviatura de infinitamente seguido, por sus siglas en inglés. Con esta notación, es claro que si tomamos $\delta > 0$ basta demostrar que

$$P\left(\left|\frac{\mu_{m,G}(H_n)}{\varepsilon_n} - L_0\right| > \delta \ i.o.\right) = 0$$

Tenemos que

$$\begin{aligned}\left|\frac{\mu_{m,G}(H_n)}{\varepsilon_n} - L_0\right| \leq{}& \left|\frac{\mu_{m,G}(H_n)}{\varepsilon_n} - \frac{\mu(H_n \cap G)}{\varepsilon_n}\right| \\ &+ \left|\frac{\mu(H_n \cap G)}{\varepsilon_n} - \frac{\mu\big(B(G^c, \varepsilon_n) \cap G\big)}{\varepsilon_n}\right| \\ &+ \left|\frac{\mu\big(B(G^c, \varepsilon_n) \cap G\big)}{\varepsilon_n} - L_0\right| \\ ={}& T_1(n, m) + T_2(n) + T_3(n). \end{aligned} \tag{2.11}$$

Estudiaremos separadamente estos sumandos a partir de los siguientes lemas.

Lema 2.14. *Bajo* $(H0) - (H2)$.

$$P\big(T_1(n, m) > \delta/3 \ i.o.\big) = P\left(\left|\frac{\mu_{m,G}(H_n)}{\varepsilon_n} - \frac{\mu(H_n \cap G)}{\varepsilon_n}\right| > \frac{\delta}{3} \ i.o.\right) = 0 \quad n \to \infty.$$

Demostración. Observemos que la desigualdad

$$P\big(T_1(n, m) > \delta/3 \ i.o.\big) \leq \lim_{n_0 \to \infty} \sum_{n=n_0}^{\infty} P\left(T_1(n, m) \geq \frac{\delta}{3}\right), \tag{2.12}$$

implica que basta demostrar que la serie en 2.12 es convergente.

Consideremos la sucesión de variables aleatorias

$$\alpha_j = \mathbb{I}_{\{Y_j \in H_n\}} - \frac{\mu(H_n \cap G)}{\mu(G)}, \quad 1 \leq j \leq m,$$

y observemos que

$$E(\alpha_j | H_n) = 0, \quad E(\alpha_j^2 | H_j) = \frac{\mu(H_n \cap G)}{\mu(G)} - \left(\frac{\mu(H_n \cap G)}{\mu(G)}\right)^2.$$

Tenemos entonces que

$$P\big(T_1(n, m) > \delta/3 \, i.o.\big) = P\left(\left|\sum_{j=1}^{\infty} \alpha_j\right| \geq \frac{m\varepsilon_n}{3\mu(G)}\delta\right) = E\left(P\left[\left|\sum_{j=1}^{m} \alpha_j\right| \geq \frac{m\varepsilon_n}{3\mu(G)}\delta \,\middle|\, H_n\right]\right)$$

Si usamos ahora una desigualdad de tipo Bernstein, obtenemos que

$$P\left(\left|\sum_{j=1}^{m} \alpha_j\right| \geq \frac{m\varepsilon_n}{3\mu(G)}\delta \,\middle|\, H_n\right) \leq 2\exp\{-C(G, \delta)m\varepsilon_n\}.$$

En la última desigualdad hemos usado que $\mu(H_n \cap G) \leq \mu\big(B(G^c, \varepsilon_n) \cap G\big) = \mathcal{O}(\varepsilon_n)$, que se cumple por *(H0)*. La cota no depende de H_n por lo tanto puede aplicarse a la probabilidad, sin condicionar. Usando *(H2)* se concluye la tesis. $\square$

Lema 2.15. *Bajo* $(H0)-(H2)$.

$$P\left(T_2(n) > \frac{\delta}{3}\ i.o.\right) = P\left(\left|\frac{\mu(H_n\cap G)}{\varepsilon_n} - \frac{\mu\big(B(G^c,\varepsilon_n)\cap G\big)}{\varepsilon_n}\right| > \frac{\delta}{3}\ i.o.\right) = 0 \quad cuando\ n\to\infty.$$

Demostración. Al igual que en la demostración del Lema anterior, es suficiente demostrar que la serie con término general $P\big(T_2(n)\geq \delta/3\big)$ es sumable. Sea

$$R_n = \big(B(G^c,\varepsilon_n)\cap G\big)\setminus(H_n\cap G).$$

Veamos primero que si $T_2(n) > \delta/3$, o, equivalentemente

$$\mu(R_n)\geq \delta\varepsilon_n/3, \tag{2.13}$$

entonces existe $z\in\Gamma$ tal que

$$B(z,r\varepsilon_n)\cap\{X_1,\ldots,X_n\}=\emptyset, \tag{2.14}$$

para n suficientemente grande, y $r=\frac{\delta}{6L_0}$. Asumiremos que δ es tal que $r<1$.
En efecto, por *(H0)* tenemos que

$$\mu\big(B(G^c,\varepsilon)\cap G\big) = \mu\big(B(\Gamma,\varepsilon)\cap G\big) = \varepsilon L_0 + \rho(\varepsilon), \quad 0\leq\varepsilon\leq\varepsilon_n; \tag{2.15}$$

donde $\rho(\cdot)$ satisface $\rho(\varepsilon)=o(\varepsilon)$ cuando $\varepsilon\to 0$. Si 2.14 no se verificase tendríamos que $\Gamma\subset B(\{X_1,\ldots,X_n\},r\varepsilon_n)$, de donde $B\big(\Gamma,(1-r)\varepsilon_n\big)\cap G\subset H_n\cap G$. En particular

$$\mu(H_n\cap G)\geq(1-r)\varepsilon_n L_0+\rho\big((1-r)\varepsilon_n\big). \tag{2.16}$$

Veamos ahora que las ecuaciones 2.16 y 2.13 son incompatibles. Para ver esto observemos que los conjuntos $H_n\cap G$ y R_n son disjuntos por construcción, y que $(H_n\cap G)\cup R_n = B(G^c,\varepsilon_n)\cap G$, de donde

$$\begin{aligned}\mu\big(B(G^c,\varepsilon_n)\cap G\big) = \mu(H_n\cap G)+\mu(R_n) &\geq \frac{\delta\varepsilon_n}{3}+(1-r)\varepsilon_n L_0+\rho\big((1-r)\varepsilon_n\big)\\ &\geq \varepsilon_n L_0+\frac{\delta\varepsilon_n}{6}+\rho\big((1-r)\varepsilon_n\big),\end{aligned}$$

lo cual contradice 2.15, de donde se sigue 2.14.

Consideremos puntos $z_1,z_2,\ldots,z_M$ no aleatorios, con $z_i\in\Gamma$, tal que $\big\{B(z_i,r\varepsilon_n/2)\big\}$ forma un cubrimiento de Γ. Es claro que esto puede obtenerse con $M=\mathcal{O}(1/\varepsilon_n^d)$ ya que puede obtenerse un cubrimiento de todo $[0,1]^d$ con tal numero de bolas, de dicho radio.

Sea $z\in\Gamma$ como en 2.14 e i_0 tal que $1\leq i_0\leq M$, de modo que $z\in B(z_{i_0},r\varepsilon_n/2)$. Entonces $B(z_{i_0},r\varepsilon_n/2)\subset B(z,r\varepsilon_n)$, y, por lo tanto $B(z_{i_0},r\varepsilon_n/2)\cap\{X_1,\ldots,X_n\}=\emptyset$. Concluimos que $P\big(T_2(n)\geq\delta/3\big)$ puede ser acotado por la probabilidad de que ninguno de los n puntos, independientes, uniformemente distribuidos en G^c pertenezca a una de las bolas $B(z_j,r\varepsilon_n/2)$. Por *(H1)*, $P\big(X_1\in B(z_j,r\varepsilon_n/2)\big)\geq K\omega_d(r\varepsilon_n/2)^d$, ω_d es el volumen de la bola unitaria d-dimensional. Obtenemos que

$$\begin{aligned}P\left(T_2(n)\geq\frac{\delta}{3}\right) &\leq M\max_j P\big(B\left(z_j,r\varepsilon_n/2\right)\cap\{X_1,\ldots,X_n\}=\emptyset\big)\\ &= M\max_j P\left(X_1\notin B\left(z_j,\frac{r\varepsilon_n}{2}\right)\right)^n\\ &\leq M\big(1-C(d,G)(r\varepsilon_n)^d\big)^n\leq\exp\big\{-C(d,G,\delta)n\varepsilon_n^d\big\}.\end{aligned}$$

La serie con término general $\exp\big\{-C(d,G,\delta)n\varepsilon_n^d\big\}$ converge, por *(H2)*. $\square$

Sólo nos queda el último término en 2.11. Por *(H0)* tenemos, para n suficientemente grande

$$T_3(n) = \left|\frac{\mu\big(B(G^c,\varepsilon_n)\cap G\big)}{\varepsilon_n} - L_0\right| \leq \frac{\delta}{3}. \tag{2.17}$$

Por lo tanto, la demostración del teorema es consecuencia de aplicar los lemas anteriores y 2.17 a 2.11. Para demostrar que $\hat{L}_n \to L_0$ *c.s.* cuando $n\to\infty$ basta observar que en $\hat{L}_n$ el valor de $\mu(G)$ se remplaza por el del estimador consistente $\hat{\mu}(G)$. $\square$

Dado que el estimador $\hat{C}_n$ de C_0 se obtiene a partir de L_n simplemente dividiendo por $\hat{\mu}(G)^{(d-1)/d}$, obtenemos el siguiente corolario.

Corolario 2.16. *Bajo (H0), (H1) y (H2) y el modelo (M1') obtenemos que*

$$\hat{C}_n \to C_0 \quad c.s.\ \textit{cuando}\ n\to\infty.$$

Veamos ahora el resultado referido a la normalidad asintótica del estimador. Usaremos la notación $N(a,\sigma^2)$ para referirnos a la distribución normal, con esperanza a y varianaza σ^2. Denotaremos con el símbolo $\xrightarrow{\omega}$ a la convergencia débil.

Teorema 2.17. *Bajo las hipótesis (H0) a (H5) y el modelo (M2), tenemos que*

$$\sqrt{m\varepsilon_n}(A_n - L_0) \xrightarrow{\omega} N\big(0,\mu(G)L_0\big) \quad \textit{cuando}\ n\to\infty, \tag{2.18}$$

y, bajo (M1')

$$\sqrt{m\varepsilon_n}(\hat{L}_n - L_0) \xrightarrow{\omega} N\big(0,\mu(G)L_0\big) \quad \textit{cuando}\ n\to\infty. \tag{2.19}$$

Demostración. La demostración de 2.18 sigue el mismo esquema que el teorema anterior. Descomponemos $|\mu_{m,G}(H_n) - L_0|$ como en 2.11. Estudiaremos cada sumando de forma separada, los últimos dos terminos tenderán a 0, de modo que del primero obtendremos la convergencia débil del teorema. La demostración la haremos a partir de los siguientes lemas

Lema 2.18. *Bajo las hipótesis (H0)-(H2),*

$$\sqrt{m\varepsilon_n}\left(\frac{\mu_{m,G}(H_n)}{\varepsilon} - \frac{\mu(H_n\cap G)}{\varepsilon_n}\right) \xrightarrow{\omega} N\big(0,\mu(G)L_0\big)\ \textit{cuando}\ n\to\infty.$$

Demostración. Calculemos la transformada de Laplace de la expresión a la izquierda. Sea $\gamma>0$, tenemos que

$$\begin{aligned} E&\left(\exp\left\{-\gamma\sqrt{m\varepsilon_n}\left(\frac{\mu_{m,G}(H_n)}{\varepsilon_n} - \frac{\mu(H_n\cap G)}{\varepsilon_n}\right)\right\}\right) \\ &\qquad = E\left(E\left(\exp\left\{-\frac{\tilde{\gamma}}{\sqrt{m\varepsilon_n}}\left(\mathbb{I}_{\{Y_j\in H_n\}} - \beta_n\right)\right\}\middle| H_n\right)^m\right) \\ &\qquad\qquad = E\left(\left(1 + \frac{1}{2}\frac{\tilde{\gamma}^2}{m\varepsilon_n}(\beta_n - \beta_n^2) + \mathcal{O}\left[\frac{1}{(m\varepsilon_n)^{3/2}}\right]\right)^m\right), \end{aligned} \tag{2.20}$$

siendo $\tilde{\gamma} = \mu(G)\gamma$ y $\beta_n = \mu(H_n \cap G)/\mu(G) = E(\mathbb{I}_{\{Y_i \in H_n\}}|H_n)$. La última igualdad en 2.20 se sigue de hacer un desarrollo de Taylor de la función exponencial, hasta el segundo orden. Obsérvese que dado que $|\mathbb{I}_{\{Y_i \in H_n\}} - \beta_n| \leq 1$, el resto en el desarrollo esta uniformemente acotado. Por otro lado sabemos que $\beta_n/\varepsilon_n \xrightarrow{c.s.} L_0/\mu(G)$ por el Lema 2.15 y 2.17. Luego, el Lema se sigue de aplicar convergencia dominada. □

Lema 2.19. *Bajo (H0)-(H5)*

$$\sqrt{m\varepsilon_n}\left(\frac{\mu(H_n \cap G)}{\varepsilon_n} - \frac{\mu\big(B(G^c, \varepsilon_n) \cap G\big)}{\varepsilon_n}\right) \xrightarrow{\omega} 0 \text{ cuando } n \to \infty$$

Demostración. Simplemente tenemos que adaptar la prueba del Lema 2.15, usaremos la notación de dicho lema. El resultado queda demostrado si se demuestra que $\sqrt{m/\varepsilon_n}\mu(R_n) \to 0$ en probabilidad, cuando $n \to \infty$.
Sea $\delta > 0$ y supongamos que

$$\sqrt{\frac{m}{\varepsilon_n}}\mu(R_n) \geq \delta. \tag{2.21}$$

Definamos $\delta_n = \frac{\delta}{2\sqrt{m\varepsilon_n}}$. Las hipótesis *(H2)* y *(H3)* implican que

$$\mu\Big(B\big(G^c, \varepsilon_n(1-\delta_n)\big) \cap G\Big) = L_0\varepsilon_n(1-\delta_n) + \mathcal{O}(\varepsilon_n^2),$$

y luego, si estamos en el evento en el que 2.21 vale, existe $z \in \Gamma$ tal que

$$B(z, \varepsilon_n\delta_n) \cap \{X_1, \ldots, X_n\} = \emptyset.$$

Consideremos ahora un cubrimiento de Γ por M bolas $\{B_j\}_{1\leq j\leq M}$ con radio $\varepsilon_n\delta_n/2$, centradas en $\{z_j\}_{1\leq j\leq M}$ puntos de Γ. Es fácil ver que los centros pueden tomarse de modo que $M = \mathcal{O}((1/\varepsilon_n\delta_n)^d)$. Finalmente, si j_0 es tal que $\|z - z_{j_0}\| \leq \varepsilon_n\delta_n/2$, entonces $B_{j_0} \cap \{X_1, \ldots, X_n\} = \emptyset$. El problema se reduce entonces a calcular la probabilidad de tal evento.

$$\begin{aligned}
P\left(\sqrt{\frac{m}{\varepsilon_n}}\mu(R_n) \geq \delta\right) &\leq P\left(\bigcup_{j=1}^{M}\Big\{\{X_1, \ldots, X_n\} \cap B_j = \emptyset\Big\}\right) \\
&\leq M \max_j P\Big(\{X_1, \ldots, X_n\} \cap B_j = \emptyset\Big) \\
&\leq C(G)\left(\frac{m}{\varepsilon_n}\right)^{d/2}\left(1 - C(d,G)\left(\frac{\varepsilon_n}{m}\right)^{d/2}\right)^n \xrightarrow{n} 0,
\end{aligned}$$

donde el último límite se obtiene de reemplazar M y δ_n por sus valores y aplicar las hipótesis *(H1)* y *(H5)*. □

Para demostrar el teorema observemos que *(H4)* implica

$$\sqrt{m\varepsilon_n}T_3(n) = \sqrt{m\varepsilon_n}\left|\frac{\mu(B(G^c, \varepsilon_n) \cap G)}{\varepsilon_n} - L_0\right| \leq \delta,$$

si n es suficientemente grande. Lo cual, junto con los lemas 2.18 y 2.19 concluyen la prueba de 2.18. Para demostrar 2.19 observemos que la diferencia $\hat{\mu}(G) - \mu(G)$ es de orden $1/N \leq 1/\sqrt{n+m} \ll 1/\sqrt{m\varepsilon_n}$, por lo tanto podemos reemplazar A_n por $\hat{L}_n$ en 2.18, manteniendo los órdenes de convergencia y las distribuciones asintóticas. $\square$

2.4. Caso α-convexo

En esta sección veremos resultados referentes a la estimación de volúmenes para el caso en que el conjunto es un subconjunto de $\mathbb{R}^d$ α-convexo. Al igual que en la sección anterior, buscamos estimar el contenido de Minkowski del borde del conjunto. Supondremos que tenemos un conjunto $G \subset (0,1)^d$ y denotaremos $R = [0,1]^d \backslash int(G)$. Supongamos que tenemos observaciones $(Z_1, \delta_1), \ldots, (Z_n, \delta_n)$ de una variable aleatoria (Z, δ) con distribución uniforme en $[0,1]^d$, y $\delta = \mathbb{I}_{\{Z \in G\}}$. Denotemos P_X la distribución de Z dado que sabemos que $\delta = 1$ (es decir $Z \in G$) e Y la de Z dado que sabemos $\delta = 0$. Obsérvese que P_X es uniforme en G y P_Y en R.

2.4.1. El estimador

Considremos $\mathcal{X}_n = \{Z_i : \delta_i = 1\}$ y $\mathcal{Y}_n = \{Z_i : \delta_i = 0\}$,

$$
\begin{aligned}
G_n =& C_\alpha(\mathcal{X}_n) = (\mathcal{X}_n \oplus \alpha \overset{\circ}{B}) \ominus \alpha \overset{\circ}{B}, \\
R_n =& C_\alpha(\mathcal{Y}_n) = (\mathcal{Y}_n \oplus \alpha \overset{\circ}{B}) \ominus \alpha \overset{\circ}{B},
\end{aligned}
$$

y $\Gamma_n = B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n)$ siendo ε_n una sucesión de numeros positivos que converge a 0. Por lo tanto el estimador propuesto para el contenido de Minkowski $L_0(G)$ es

$$
L_n = \frac{\mu(\Gamma_n)}{2\varepsilon_n}.
$$

Veamos entonces uno de los resultados asintóticos.

Teorema 2.20. *Sea $G \subset (0,1)^d$ un conjunto compacto no vacío. Supongamos que una bola de radio $\alpha > 0$ rueda libremente por G y por $\overline{G^c}$, entonces, con probabilidad uno*

$$
\inf_{\varepsilon_n} |L_n - L_0| = \mathcal{O}\left(\frac{\log n}{n}\right)^{\frac{1}{d+1}}.
$$

El óptimo se obtiene para $\varepsilon_n = (\log n/n)^{1/(d+1)}$.

Demostración. El teorema será consecuencia de 3 proposiciones. En la primera se establece que si $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\rho_n) \cap B(Z_n^{\mathcal{Y}}, 2\rho_n)$ donde $Z_n^{\mathcal{X}} = \{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \rho_n^2\}$ y $Z_n^{\mathcal{Y}} = \{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \rho_n^2\}$ entonces $B(\Gamma, \varepsilon_n) \setminus \Gamma$ está contenida en $D_n = B(\Gamma, \varepsilon_n) \setminus B(\Gamma, \varepsilon_n - K\rho_n^2)$ para n suficientemente grande. En la segunda se encuentra, basándose en que $\mu(D_n) = \mathcal{O}(\rho_n^2)$, una cota para $|L_n - L_0|$ dependiente solo de ε_n y ρ_n. En la tercera se demuestra que efectivamente, con probabilidad 1, $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\rho_n) \cap B(Z_n^{\mathcal{Y}}, 2\rho_n)$.

Proposición 2.21. *Sea G un conjunto en las hipótesis del teorema 2.20, entonces*

(i) Con probabilidad 1, $\Gamma_n \subset B(\Gamma, \varepsilon_n)$.

(ii) Supongamos que $\rho_n \to 0$ y $\rho_n^2 \varepsilon_n^{-1} \to 0$ y que

$$P\big(\Gamma \subset B(Z_n^{\mathcal{X}}, 2\rho_n) \cap B(Z_n^{\mathcal{Y}}, 2\rho_n) \text{ a partir de cierto } n\big) = 1,$$

donde $Z_n^{\mathcal{X}} = \{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \rho_n^2\}$ y $Z_n^{\mathcal{Y}} = \{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \rho_n^2\}$. Entonces si $K \geq \text{máx}(2, 8/\alpha)$, tenemos que

$$P\big(B(\Gamma, \varepsilon_n - K\rho_n^2) \subset \Gamma_n \text{ a partir de un cierto } n\big) = 1.$$

Demostración. Veamos la demostración de la parte (i). Bajo las hipótesis de que G y $\overline{G^c}$ son α-convexos, se puede ver fácilmente que R también lo es. Dado que con probabilidad 1 tenemos que $\mathcal{X}_n \subset G$ y $\mathcal{Y}_n \subset R$,

$$G_n = C_\alpha(\mathcal{X}_n) \subset C_\alpha(G) = G \quad y \quad R_n = C_\alpha(\mathcal{Y}_n) \subset C_\alpha(R) = R.$$

Entonces, con probabilidad uno

$$\Gamma_n = B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n) \subset B(G, \varepsilon_n) \cap B(R, \varepsilon_n) = B(\Gamma, \varepsilon_n), \tag{2.22}$$

lo cual concluye la prueba de la parte (i). La parte (ii) quedará demostrada si se demuestra que

$$P\big(\Gamma \subset B(G_n, K\rho_n^2) \cap B(R_n, K\rho_n^2) \text{ a partir de un cierto } n\big) = 1, \tag{2.23}$$

dado que si $\Gamma \subset B(G_n, K\rho_n^2) \cap B(R_n, K\rho_n^2)$ y $\varepsilon_n > K\rho_n^2$ entonces

$$\begin{aligned} B(\Gamma, \varepsilon_n - K\rho_n^2) &\subset B\big(B(G_n, K\rho_n^2) \cap B(R_n, K\rho_n^2), \varepsilon_n - K\rho_n^2\big) \\ &\subset B(G_n, \varepsilon_n) \cap B(R_n, \varepsilon_n) = \Gamma_n. \end{aligned}$$

Para demostrar 2.23 es suficiente ver que con probabilidad 1, para n suficientemente grande,

$$x_G = x - K\rho_n^2 \eta(x) \in G_n \quad y \quad x_R = x + K\rho_n^2 \eta(x) \in R_n, \tag{2.24}$$

para todo $x \in \Gamma$, donde $\eta(x)$ es el vector normal saliente en el punto x. Para demostrar 2.24 necesitamos demostrar que x_G no puede estar contenido en una bola de radio α que no corte $\mathcal{X}_n$, análogamente para x_R. Para eso vamos a necesitar algunos lemas.

Lema 2.22. *Sea G un conjunto en las hipótesis del Teorema 2.20, supongamos que*

$$\Gamma \subset B(Z_n^{\mathcal{X}}, 2\rho_n) \cap B(Z_n^{\mathcal{Y}}, 2\rho_n),$$

donde $Z_n^{\mathcal{X}} = \{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \rho_n^2\}$ y $Z_n^{\mathcal{Y}} = \{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \rho_n^2\}$. Entonces para todo $y \in \mathbb{R}^d$, tal que $d(y, \Gamma) = \alpha - \kappa$ con $\text{máx}(2, 8/\alpha)\rho_n^2 < \kappa \leq \alpha$,

$$\mathring{B}(y, \alpha) \cap \mathcal{X}_n \neq \emptyset \quad y \quad \mathring{B}(y, \alpha) \cap \mathcal{Y}_n \neq \emptyset.$$

Demostración. Sea y tal que $d(y,\Gamma)=\alpha-\kappa$ con $\text{máx}(2,8/\alpha)\rho_n^2<\kappa\leq\alpha$. Denotemos como $P_\Gamma y$ la proyección de y sobre Γ. Dado que $\Gamma\subset B(Z_n^{\mathcal{X}},2\rho_n)$, existe $z_x\in Z_n^{\mathcal{X}}$ tal que $\|z_x-P_\Gamma y\|\leq 2\rho_n$, además $d(z_x,\Gamma)\leq\rho_n^2<\kappa/2$. Si $\|z_x-y\|\geq\alpha$ entonces, por el Lema 1.40

$$\|z_x-P_\Gamma y\|\geq\sqrt{\frac{\alpha\kappa}{2}}>2\rho_n,$$

lo cual es absurdo. La última desigualdad es consecuencia de que $\kappa>8\rho_n^2/\alpha$. Por lo tanto $\|z_x-y\|<\alpha$ y $\overset{\circ}{B}(y,\alpha)\cap\mathcal{X}_n\neq\emptyset$. Análogamente se demuestra que $\overset{\circ}{B}(y,\alpha)\cap\mathcal{Y}_n\neq\emptyset$. □

Antes de seguir con el siguiente lema vamos a introducir algo de notación, dado que estamos suponiendo que $G\subset(0,1)^d$ para todo $x\in G$

$$d\big(x,\mathbb{R}^d\setminus(0,1)^d\big)>0.$$

La función $d\big(\cdot,\mathbb{R}^d\setminus(0,1)^d\big)$ es continua, alcanza su mínimo en el conjunto compacto, G. Denotemos como e dicho mínimo.

$$e=\min_{x\in G}d\big(x,\mathbb{R}^d\setminus(0,1)^d\big)>0.$$

Obsérvese que $B(G,e)\subset[0,1]^d$.

Lema 2.23. *Sea $x\in\mathbb{R}^d$ un punto tal que $0\leq d(x,G)\leq e/2$, $y\notin[0,1]^d$ tal que $x\in\overset{\circ}{B}(y,\alpha)$. Entonces existe $z_0\in R$ para el cual $B(z_0,e/4)\subset\overset{\circ}{B}(y,\alpha)$.*

Demostración. la función

$$d(\lambda)=d\big(\lambda x+(1-\lambda)y,G\big),\quad 0\leq\lambda\leq 1,$$

es continua. Dado que $y\notin[0,1]^d$, tenemos que $d(0)=d(y,G)>e$. Por lo tanto

$$d(1)=d(x,G)\leq e/2.$$

Luego por el Teorema de Bolzano, existe z_0 en el segmento $[x,y]$ tal que $d(z_0,G)=3e/4$. Más aun, $z_0\in R$ ya que $z_0\in B(G,e)\subset[0,1]^d$ y $z_0\notin G$.
Veamos que $B(z_0,e/4)\subset\overset{\circ}{B}(y,\alpha)$. Sea $z\in B(z_0,e/4)$. Tenemos que

$$\|z-y\|\leq\|z-z_0\|+\|z_0-y\|\leq\frac{e}{4}+\|z_0-y\|.$$

Dado que z_0 está en el segmento $[x,y]$

$$\|z_0-y\|=\|x-y\|-\|x-z_0\|.$$

De $d(z_0,G)=3e/4$ y $d(x,G)\leq e/2$ se sigue que $\|x-z_0\|\geq e/4$ y, por lo tanto

$$\|z_0-y\|=\|x-y\|-\|x-z_0\|<\alpha-\frac{e}{4}.$$

Luego,

$$\|z-y\|<\frac{e}{4}+\alpha-\frac{e}{4}=\alpha.$$

□

Lema 2.24. *Supongamos que* $\Gamma \subset B(Z_n^{\mathcal{X}}, 2\rho_n) \cap B(Z_n^{\mathcal{Y}}, 2\rho_n)$ *y* $K\rho_n^2 < \text{mín}(e/2, \alpha)$, *donde* $K \geq \text{máx}(2, 8/\alpha)$. *Supongamos que* $d_H(\mathcal{X}_n, G) < \alpha$ *y* $d_H(\mathcal{Y}_n, R) < \text{mín}(e/4, \alpha)$. *Entonces, para todo* $x \in \Gamma$,

$$x - K\rho_n^2\eta(x) \in G_n \quad y \quad x + K\rho_n^2\eta(x) \in R_n,$$

donde $\eta(x)$ *es el vector normal saliente en el punto* x.

Demostración. Sea $x \in \Gamma$ y $x_G = x - K\rho_n^2\eta(x)$. El punto x_G pertenece a G_n si toda bola abierta de radio α que contenga a x_G corta a algún punto de la muestra $\mathcal{X}_n$. Sea $y \in \mathbb{R}^d$ un punto tal que $x_G \in \overset{\circ}{B}(y, \alpha)$. Queremos demostrar que $\overset{\circ}{B}(y, \alpha) \cap \mathcal{X}_n \neq \emptyset$. Esto es inmediato cuando $y \in G$ ya que por hipótesis $d_H(\mathcal{X}_n, G) < \alpha$. Supongamos entonces que $y \in G^c$. Ya que $x_G \in \overset{\circ}{B}(y, \alpha) \cap G$, entonces $d(y, \Gamma) = \alpha - \kappa$, donde $\kappa > K\rho_n^2 \geq \text{máx}(2, 8/\alpha)\rho_n^2$. Por el Lema 2.22 tenemos que $\overset{\circ}{B}(y, \alpha) \cap \mathcal{X}_n \neq \emptyset$.

Sea $x_R = x + K\rho_n^2\eta(x)$. Al igual que antes, para demostrar que x_R está en R_n basta demostrar que $\overset{\circ}{B}(y, \alpha) \cap \mathcal{Y}_n \neq \emptyset$. Para todo $y \in \mathbb{R}^d$ tal que $x_R \in \overset{\circ}{B}(y, \alpha)$. Nuevamente es inmediato cuando $y \in R$. Supongamos que $y \notin R$. Tenemos dos posibilidades: $y \in G$ o $y \notin [0,1]^d$. En el primer caso, como $x_R \in \overset{\circ}{B}(y, \alpha) \cap G^c$, tenemos que $d(y, \Gamma) = \alpha - \kappa$, con $\kappa > K\rho_n^2 \geq \text{máx}(2, 8/\alpha)\rho_n^2$. El Lema 2.22 implica entonces que $\overset{\circ}{B}(y, \alpha) \cap \mathcal{Y}_n \neq \emptyset$. Por último si $y \in [0,1]^d$, por la definición de x_R tenemos que $d(x_R, G) = K\rho_n^2 < e/2$, entonces, por el Lema 2.23 existe $z_0 \in R$ tal que $B(z_0, e/4) \subset \overset{\circ}{B}(y, \alpha)$. Dado que $d_H(\mathcal{Y}_n, R) < e/4$, tenemos que $B(z_0, e/4) \cap \mathcal{Y}_n \neq \emptyset$. Por lo tanto $\overset{\circ}{B}(y, \alpha) \cap \mathcal{Y}_n \neq \emptyset$. □

La Proposición 2.21 queda demostrada ya que las condiciones del Lema 2.24 se cumplen con probabilidad 1, para n suficientemente grande. □

Veamos ahora la segunda proposición, que establece una cota para $|L_n - L_0|$ que se obtiene a partir de separar la resta en sesgo $|L(\varepsilon_n) - L_0|$ y varianza $|L_n - L(\varepsilon_n)|$.

Proposición 2.25. *Bajo las hipótesis de la Proposición 2.21 tenemos, con probabilidad uno*

$$|L_n - L_0| \leq |L(\varepsilon_n) - L_0| + \mathcal{O}\left(\frac{\rho_n^2}{\varepsilon_n}\right) = \mathcal{O}(\varepsilon_n) + \mathcal{O}\left(\frac{\rho_n^2}{\varepsilon_n}\right),$$

y

$$\inf_{\varepsilon_n} |L_n - L_0| = \mathcal{O}(\rho_n).$$

Demostración. Sabemos que $|L(\varepsilon_n) - L_0| = \mathcal{O}(\varepsilon_n)$. Por otro lado, de la Proposición 2.21 tenemos que

$$P\big(B(\Gamma, \varepsilon_n - K\rho_n^2) \subset \Gamma_n \subset B(\Gamma, \varepsilon_n) \text{ a partir de cierto } n\big) = 1.$$

Entonces, con probabilidad 1, para n suficientemente grande

$$|L_n - L(\varepsilon_n)| = \frac{\mu\big(B(\Gamma, \varepsilon_n)\big)}{2\varepsilon_n} - \frac{\mu(\Gamma_n)}{2\varepsilon_n} \leq \frac{\mu\big(B(\Gamma, \varepsilon_n) - \mu\big(B(\Gamma, \varepsilon_n - K\rho_n^2)\big)}{2\varepsilon_n},$$

veamos ahora un lema que nos será necesario para concluir la demostración de la proposición anterior.

Lema 2.26. *Supongamos que $F(\varepsilon) = \mu\big(B(\Gamma, \varepsilon)\big)$ es diferenciable en un entorno de 0 y que la derivada F' es continua en 0, entonces*

$$\lim_{n\to\infty} \frac{\mu(D_n)}{2K\rho_n^2} = L_0,$$

donde $D_n = B(\Gamma, \varepsilon_n) \setminus B(\Gamma, \varepsilon_n - K\rho_n^2)$.

Demostración. Para n suficientemente grande, (dado que $\varepsilon_n, \rho_n \to \infty$ y $\rho_n^2 \varepsilon_n^{-1} \to 0$) tenemos que

$$\begin{aligned} \frac{\mu(D_n)}{2K\rho_n^2} &= \frac{\mu\big(B(\Gamma, \varepsilon_n\big) - \mu\big(B(\Gamma, \varepsilon_n - K\rho_n^2)\big)}{2K\rho_n^2} \\ &= \frac{F(\varepsilon_n) - F(\varepsilon_n - K\rho_n^2)}{2K\rho_n^2} = \frac{F'(\varsigma_n)K\rho_n^2}{2K\rho_n^2} = \frac{F'(\varsigma_n)}{2} \end{aligned}$$

donde hemos aplicado el Teorema del Valor Medio, siendo ς_n un punto en el intervalo $(\varepsilon_n - K\rho_n^2, \varepsilon_n)$. Dado que F' es continua en 0,

$$\lim_{n\to\infty} \frac{\mu(D_n)}{2K\rho_n^2} = \frac{F'(0)}{2} = L_0$$

□

Por el Lema 2.26, tenemos que, con probabilidad 1

$$|L_n - L(\varepsilon_n)| = \mathcal{O}\left(\frac{\rho_n^2}{\varepsilon_n}\right).$$

Por lo tanto, con probabilidad 1

$$|L_n - L_0| \leq |L(\varepsilon_n) - L_0| + \mathcal{O}\left(\frac{\rho_n^2}{\varepsilon_n}\right) = \mathcal{O}(\varepsilon_n) + \mathcal{O}\left(\frac{\rho_n^2}{\varepsilon_n}\right).$$

Por lo tanto si igualamos los órdenes a la derecha de la última expresión obtenemos que $\varepsilon_n = \rho_n$ y

$$\inf_{\varepsilon_n} |L_n - L_0| = \mathcal{O}\left(\frac{\log n}{n}\right)^{\frac{1}{d+1}}.$$

Esto concluye la demostración de la Proposición 2.25. □

Proposición 2.27. *Si $c > 0$ es suficientemente grande entonces*

$$P\big(\Gamma \subset B(Z_n^{\mathcal{X}}, 2\rho_n) \cap B(Z_n^{\mathcal{Y}}, 2\rho_n) \text{ a partir de un cierto } n\big) = 1,$$

donde

$$\rho_n = \left(\frac{c\log n}{n}\right)^{\frac{1}{d+1}},$$

$$Z_n^{\mathcal{X}} = \big\{Z_i \in \mathcal{X}_n : d(Z_i, \Gamma) \leq \rho_n^2\big\} \quad y \quad Z_n^{\mathcal{Y}} = \big\{Z_i \in \mathcal{Y}_n : d(Z_i, \Gamma) \leq \rho_n^2\big\}.$$

Demostración. El Teorema 1 de D*ü*mbgen y Walther (1996), ver [11], establece que, para $\rho_n > 0$,

$$P\big(\Gamma \nsubseteq B(Z_n^{\mathcal{U}}, 2\rho_n)\big) \leq \rho_n^{-d} \prod(G, Z_n^{\mathcal{U}}, \rho_n), \quad \mathcal{U} = \mathcal{X}, \mathcal{Y},$$

donde $\prod(G, Z_n^{\mathcal{U}}, \rho_n) = \sup_{x\in\Gamma} P\big(B(x,\rho_n) \cap Z_n^{\mathcal{U}} = \emptyset\big)$. Por el Lema de Borel-Cantelli es suficiente demostrar que

$$\sum_{n=1}^{\infty} \rho_n^{-d} \prod(G, Z_n^{\mathcal{U}}, \rho_n) < \infty, \quad \mathcal{U} = \mathcal{X}, \mathcal{Y}. \tag{2.25}$$

Dado $x \in \Gamma$. Como Z es uniformemente distribuido en $[0,1]^d$ tenemos que, para $\rho_n^2 < e$,

$$\begin{aligned} P\big(B(x,\rho_n) \cap Z_n^{\mathcal{X}} = \emptyset\big) &= P\big(Z_i \notin B(x,\rho_n) \cap B(\Gamma, \rho_n^2) \cap G,\ i = 1, \ldots, n\big) \\ &= \Big(1 - \mu\big(B(x,\rho_n) \cap B(\Gamma,\rho_n^2) \cap G\big)\Big)^n \\ &\leq \exp\Big(-n\mu\big(B(x,\rho) \cap B(\Gamma,\rho_n^2) \cap G\big)\Big). \end{aligned}$$

Análogamente

$$P\big(B(x,\rho_n) \cap Z_n^{\mathcal{Y}} = \emptyset\big) \leq \exp\Big(-n\mu\big(B(x,\rho_n) \cap B(\Gamma,\rho_n^2) \cap R\big)\Big).$$

Por 1.44, sabemos que

$$P\big(B(x,\rho_n) \cap Z_n^{\mathcal{X}} = \emptyset\big) \leq \exp\big(-n\gamma\rho_n^{d+1}\big), \quad \mathcal{U} = \mathcal{X}, \mathcal{Y}.$$

Esto concluye la demostración de la Proposición. □

El Teorema es una consecuencia inmediata de las Proposiciones 2.21, 2.25 y 2.27. □

Capítulo 3

Estimación de conjuntos: Datos insuficientes

3.1. Introducción

Hasta ahora hemos estudiado la estimación de ciertos subconjuntos de $\mathbb{R}^d$, y hemos propuesto estimadores para el borde de dichos conjuntos, bajo el supuesto de que dada una muestra $X_1, \ldots, X_n$ sabemos exactamente, o bien que todos los X_i pertenecen al conjunto, tal es el caso del estimador de Devroye y Wise, o que podemos dividir la muestra en puntos del conjunto y puntos del complemento, en el caso en que nos interesó estimar longitudes, áreas, etc. El presente capítulo pretende mostrar a partir de ejemplos que en ciertos casos de interés práctico no podemos hacer esa división, y lo que se tiene entonces son simplemente *mediciones* $X_1, \ldots, X_n$ realizadas en puntos $Z_1, \ldots, Z_n$ de $[0,1]^d$ y sabemos únicamente, que algunas de ellas tienen una distribución desconocida F, en un cierto conjunto desconocido S, y el resto otra distribución, también desconocida y distinta de F, que llamaremos G. Lo que se busca entonces es estimar el conjunto S. Si supiéramos cuáles X_i tienen distribución F y cuales G, sin saber necesariamente cuales son dichas distribuciones, estaríamos entonces en los casos estudiados en los capítulos anteriores.

Pensemos a modo de ejemplo muy simple, que estamos en el intervalo $[0,1]$ y nuestro conjunto S es de la forma $[0,\theta)$, que tenemos mediciones $X_1, \ldots, X_n$ en puntos $Z_1, \ldots, Z_n$ de $[0,1]$, los X_i para los cuales $Z_i \in S$, tienen cierta distribución F y los X_i para los cuales $Z_i \in S^c$ tienen distribución G, éste modelo tan simple ha sido ampliamente estudiado en la literatura, y se conoce con el nombre de *problema de punto de cambio*. Podemos pensar que lo que tenemos son mediciones a lo largo del tiempo de algún fenómeno físico y lo que buscamos es el punto de cambio en la distribución de dichas mediciones.

Para el caso $d = 2$ surge de manera natural imaginar que nuestro conjunto en cuestión es un área geográfica. Consideremos la siguiente aplicación a la forestación, pensemos que las observaciones $\{X_i : i \in I\}$ representan alturas de árboles, y que por ejemplo F es la distribución de árboles *sanos* y G la de árboles con alguna patología. El objetivo es entonces estimar el conjunto que delimita los árboles sanos del resto. Ejemplos análogos pueden pensarse en el área de oceanografía, siendo las

X_i mediciones de profundidades, etc.

3.2. Motivación intuitiva del estimador propuesto

Vamos a expresar primero con algo más de precisión lo dicho en la sección anterior. Supondremos que el conjunto que queremos estimar, $S \subset (0,1)^d$, es Borel medible, y que tenemos datos $(Z_i, X_i)_{i=1,\ldots,n}$ independientes donde Z_i tiene distribución uniforme en $[0,1]^d$ y X_i tiene distribución F si $Z_i \in S$ y $G \neq F$ si $Z_i \in S^c$. Las distribuciones F y G son totalmente desconocidas, la única suposición que se hará es que son distintas. El objetivo será elegir según algún criterio que se especificará, un conjunto que aproxime S en medida, que tomaremos de una familia que denotaremos Γ_n (cuyo cardinal anotamos $|\Gamma_n|$). Para definir la familia Γ_n consideremos una grilla U_d generada de la siguiente manera, dividimos el j-ésimo $(1 \leq j \leq d)$ eje en n_j divisiones igualmente espaciadas $(1/n_1, 2/n_2, \ldots, n_j/n_j)$ de modo que los nodos de la grilla son los puntos $l = (l_1/n_1, l_2/n_2, \ldots, l_d/n_d)$ con $l_j \in \{1, 2, \ldots, n_j\}$. Γ_n serán entonces conjuntos T generados como unión de rectángulos con vértices en la grilla.

Vamos a introducir primero algo de notación:

1) Si $A \subset (0,1)^d$ es medible, fijado n, $|A| = \#\{i : Z_i \in A\}$.

2) Para $\varepsilon_n > 0$ y $T \in \Gamma_n$ anotamos:

$$T \oplus \varepsilon_n B := \bigcup_{l \in T \cap U_d} B(l, \varepsilon_n).$$

Tomaremos ε_n de modo que $T \subset T \oplus \varepsilon_n B$

3) Por simplicidad si A es un conjunto medible, usaremos la notación $j \in A$ para $\{j : Z_j \in A\}$.

Dado un $T \in \Gamma_n$ consideremos las distribuciones empíricas

$$h_T(x) := \frac{1}{|T \oplus \varepsilon_n B|} \sum_{i \in T \oplus \varepsilon_n B} \mathbb{I}_{\{X_i \leq x\}}$$

y

$$h_{T^c}(x) := \frac{1}{|(T \oplus \varepsilon_n B)^c|} \sum_{i \in (T \oplus \varepsilon_n B)^c} \mathbb{I}_{\{X_i \leq x\}}.$$

Tenemos entonces que $h_T(x)$ es una estimación de:

$$\frac{|T \oplus \varepsilon_n B \cap S| F(x) + |T \oplus \varepsilon_n B \cap S^c| G(x)}{|T \oplus \varepsilon_n B|},$$

y $h_{T^c}(x)$ de:

$$\frac{|(T \oplus \varepsilon_n B)^c \cap S| F(x) + |(T \oplus \varepsilon_n B)^c \cap S^c| G(x)}{|(T \oplus \varepsilon_n B)^c|},$$

y por lo tanto

$$|h_T(x) - h_{T^c}(x)| \approx \left| \frac{|T \oplus \varepsilon_n B \cap S|}{|T \oplus \varepsilon_n B|} - \frac{|(T \oplus \varepsilon_n B)^c \cap S|}{|(T \oplus \varepsilon_n B)^c|} \right| |F(x) - G(x)|.$$

Observar que la expresión anterior está mayorada por $|F(x) - G(x)|$.

3.2.1. Definición del Estimador

Consideremos la diferencia $d_i^T := |h_T(X_i) - h_{T^c}(X_i)|$ para $i = 1, \ldots, n$, y una función $N(d_1^T, d_2^T, \ldots, d_n^T)$. La función N debe cumplir ciertas propiedades que se especificarán más adelante.Como ejemplos de funciones que cumplen esas propiedades tenemos la norma de *Kolmogorov-Smirnov* $N_{KS}(d_1, \ldots, d_n) := sup_{1\leq i\leq n} d_i$, la norma de *Cramér-von Mises* $N_{Cv}(d_1, \ldots, d_n) = (\sum_{i=1}^n d_i^2/n)^{1/2}$, y la media aritmética $N_{am}(d_1, \ldots, d_n) := \sum_{i=1}^n d_i/n$.

Finalmente, para definir el estimador, tenemos que tener en cuenta valores muy chicos de $|T \oplus \varepsilon_n B|$ (relativos a n) ya que estos producen valores de d_i muy grandes. Esto se logra multiplicando por $|T \oplus \varepsilon_n B|/n$. Por las mismas razones, multiplicamos por $|(T \oplus \varepsilon_n B)^c|/n$.

El candidato a estimador $\hat{S}_n$ se define como el conjunto T en Γ_n que maximiza la función

$$D(T) := \frac{|T \oplus \varepsilon_n B|}{n} \frac{|(T \oplus \varepsilon_n B)^c|}{n} N(d_1^T, \ldots, d_n^T),$$

es decir $\hat{S}_n := \text{argmax}_{T\in\Gamma_n} D(T)$.

3.2.2. Sobre la distancia entre conjuntos

Para medir la proximidad entre conjuntos vamos a tener en cuenta que tan próximo está $T \oplus \varepsilon B$ de S y que tan próximo está $(T \oplus \varepsilon_n B)^c$ de S^c. Para eso, si $A \triangle B$ denota la diferencia simétrica entre A y B, definimos:

$$\partial(A, B) := \text{mín}\left\{\mu(A\triangle B), \mu(A^c \triangle B)\right\}.$$

Observación 3.1. *$\partial(\ .\ ,\ .\ )$ tiene las siguientes propiedades:*

- *(1.a)* $\partial(A, B) \geq 0$
- *(1.b)* $\partial(A, A) = 0$
- *(1.c)* $\partial(A, B) = \partial(B, A)$
- *(1.d)* $\partial(A, C) \leq \partial(A, B) + \partial(B, C)$

Demostración. Las propiedades 1.a, 1.b y 1.c son obvias a partir de la definición, veamos la prueba de *(1.d)*. Si escribimos $\partial(A, B)$ en términos de A, B y C obtenemos:

$$\begin{aligned}
\partial(A, B) &= \text{mín}\Big\{\mu(A \cap B^c \cap C) + \mu(A \cap B \cap C^c) + \mu(A^c \cap B \cap C) + \mu(A^c \cap B \cap C^c), \\
&\qquad \mu(A^c \cap B^c \cap C) + \mu(A^c \cap B^c \cap C^c) + \mu(A \cap B \cap C) + \mu(A \cap B \cap C^c)\Big\}. \\
&=: \text{mín}\{U, V\} \qquad .
\end{aligned}$$

Si hacemos un desarrollo análogo para $\partial(A, C)$ y $\partial(C, B)$ en términos de A, B y C. Uno de los cuatro posibles valores de $\partial(A, C) + \partial(C, B)$, es:

$$\begin{aligned}
W := \mu(C \cap A^c \cap B) + \mu(C \cap A^c \cap B^c) + \mu(C^c \cap A \cap B) + \mu(C^c \cap A \cap B^c) + \mu(B \cap C^c \cap A) \\
+ \mu(B \cap C^c \cap A^c) + \mu(B^c \cap C \cap A) + \mu(B^c \cap C \cap A^c)
\end{aligned}$$

Como los sumandos de U son un subconjunto de los sumandos de W obtenemos que $\partial(A, B) = \text{mín}\{U, V\} \leq U \leq W$. Desigualdades análogas se obtienen para las otras tres posibles expresiones de $\partial(A, C) + \partial(C, B)$.

3.2.3. N(.)

Definición 3.2. Una función $N(.)$: $\mathbb{R}_+^n \to \mathbb{R}_+^1$ es una norma de *promedios-dominante* si se cumple:

D.a) **Simétria** $N(.)$ es simétrica en sus n variables.

D.b) **Homogeneidad** $N(\alpha d_1, \ldots, \alpha d_n) = \alpha N(d_1, \ldots, d_n)$ para todo $\alpha \geq 0$.

D.c) **Desigualdad Triangular** $N(d_1+c_1, \ldots, d_n+c_n) \leq N(d_1, \ldots, d_n)+N(c_1, \ldots, c_n)$.

D.d) **Identidad** $N(1, \ldots, 1) = 1$.

D.e) **Monotonía** $N(d_1, \ldots, d_n) \leq N(c_1, \ldots, c_n)$ siempre que $d_i \leq c_i$ para todo i.

D.f) **Promedio-Dominancia** $N(d_1, \ldots, d_n) \geq \sum_{i=1}^n d_i/n$.

Es fácil ver que las normas N_{KS}, N_{Cv}, y N_{am} son normas *promedio-dominante.*

3.2.4. Hipótesis sobre Γ_n

Supondremos que:

T1) $\forall T \in \Gamma_n$ $|T|/n > 0$ y $0 < \mu(S) < 1$.

T2) $\forall n$ existe $T_n \in \Gamma_n$ y ε_n tal que la sucesión $\{T_n\}$ cumple $n^\delta \partial(S, T \oplus \varepsilon_n B) \to 0$ cuando $n \to \infty$, para $\delta < 1/2$.

Es claro que si no tenemos tales candidatos en Γ_n no podremos elegir ninguno que aproxime S en medida.

T3) $\forall \gamma > 0$ $|\Gamma_n| \exp(-\gamma n^{1-2\delta}) < n^{-\alpha}$ con $\alpha > 1$ siendo δ como en **T2**.

El modelo que aquí se propone tiene como motivación el propuesto en [4]. La diferencia escencial respecto del trabajo de Carlstein y Krishnamoorthy es que en dicho trabajo se toman los Z_i en una grilla que depende de n y no del azar, por no tomar aquí tal grilla es que en ciertos cálculos vamos a tener que tomar esperanza condicionada. Otra diferencia a destacar es que en éste trabajo no se pide la condición **R4)** de regularidad del borde.

Capítulo 4

Convergencia en medida del estimador

4.1. Introducción

El propósito de este capítulo es presentar los resultados teóricos referidos a la consistencia del estimador propuesto en el capítulo anterior.

Teorema 4.1. ***Consistencia:*** *bajo las hipótesis* ***T1)*** *a* ***T3)***

$$n^{\delta}\partial(S,\hat{S}_n) \to 0 \text{ cuando } n \text{ tiende a infinito}, \quad c.s.$$

Teorema 4.2. ***Error en Probabilidad:*** *bajo las hipótesis* ***T1)*** *a* ***T3)*** *se cumple que*

$$P\big(\partial(S,\hat{S}_n) > \varepsilon\big) \leq c_1|\Gamma_n| \exp\big\{ - c_2\varepsilon^2 n\big\},$$

para n suficientemente grande, siendo c_1 y c_2 constantes positivas

Demostración. Vamos a demostrar el teorema 4.1 y 4.2 a partir de:

$$P\left(n^{\delta}\partial(S,\hat{S}_n) > \varepsilon\right) \leq c_1|\Gamma_n| \exp\big\{ - c_2\varepsilon^2 n^{1-2\delta}\big\},$$

para n suficientemente grande.
De la desigualdad anterior, tomando $\delta = 0$ se concluye el teorema 4.2. El teorema 4.1 es una aplicación directa del Lema de Borel-Cantelli y de la hipótesis **T3)**.

Vamos a introducir primero algo de notación que será utilizada más adelante.

Definición 4.3. Vamos a definir para $T \in \Gamma_n$

$$\begin{aligned}
\eta_T(x) &:= \big[\mu(T \oplus \varepsilon_n B \cap S)F(x) + \mu(T \oplus \varepsilon_n B \cap S^c)G(x)\big]/\mu(T \oplus \varepsilon_n B).\\
\eta_{T^c}(x) &:= \big[\mu\big((T \oplus \varepsilon_n B)^c \cap S\big)F(x) + \mu\big((T \oplus \varepsilon_n B)^c \cap S^c\big)G(x)\big]/\mu((T \oplus \varepsilon_n B)^c).\\
\delta_{ni}^T &:= \big|\eta_T(X_i^n) - \eta_{T^c}(X_i^n)\big|.\\
\Delta(T) &:= \mu\big(T \oplus \varepsilon_n B\big)\mu\big((T \oplus \varepsilon_n B)^c\big)S\big(\delta_{ni}^T : i = 1, \dots, n\big).
\end{aligned}$$

Veamos ahora dos lemas que nos permitirán trabajar con $\Delta(T)$ en lugar de con $D(T)$, esto presenta la ventaja de que en la definición de $\Delta(T)$ tenemos F y G en lugar de las distribuciones empíricas, y las medidas exactas de los conjuntos en lugar de las aproximaciones empíricas. Necesitaremos por lo tanto cotas exponenciales para la diferencia entre la distribución empírica de F, F_n, y F (de igual forma entre la distribución empírica de G, G_n, y G) y resultados de consistencia de la medida empírica de los conjuntos.

Lema 4.4. *Bajo las hipótesis **T1)** y **T3)**, dado* $\varepsilon_n > 0$

$$n^\delta \sup_{T\in\Gamma_n} \left| \mu(T \oplus \varepsilon_n B) - \frac{|T \oplus \varepsilon_n B|}{n} \right| \to 0 \quad c.s.$$

Demostración.

$$P\left(n^\delta \sup_{T\in\Gamma_n} \left| \mu(T \oplus \varepsilon_n B) - \frac{|T \oplus \varepsilon_n B|}{n} \right| > \varepsilon\right) \leq \sum_{T\in\Gamma_n} P\left(\left| \mu(T \oplus \varepsilon_n B) - \frac{|T \oplus \varepsilon_n B|}{n} \right| > \varepsilon n^{-\delta}\right).$$

Vamos a usar la desigualdad de Hoeffding que dice que si $X_1, \dots, X_n$ son independientes, $P(X_i \in [a_i, b_i]) = 1$ $S = \sum_{i=1}^n X_i$ entonces

$$P\big(S - E(S) \geq n\varepsilon\big) \leq \exp\left(-\frac{2n^2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

Consideremos Y_i $i = 1, \dots, n$ i.i.d con distribución Bernoulli$(\mu(T \oplus \varepsilon_n B))$ entonces

$$\begin{aligned} P\left(\frac{|T \oplus \varepsilon_n B|}{n} - \mu(T \oplus \varepsilon_n B) > \varepsilon n^{-\delta}\right) =& P\left(\frac{\sum_{i=1}^n Y_i}{n} - \mu(T \oplus \varepsilon_n B) > \varepsilon n^{-\delta}\right) \\ =& P\left(\sum_{i=1}^n Y_i - n\mu(T \oplus \varepsilon_n B) > \varepsilon n^{1-\delta}\right) \\ \leq& \exp\big\{ - 2n^{1-2\delta}\varepsilon^2\big\} \end{aligned}$$

Usando **T3)** se concluye la tesis. □

Con argumentos análogos se puede ver que

$$n^\delta \sup_{T\in\Gamma_n} \left| \mu(T \oplus \varepsilon_n B \cap S) - \frac{|T \oplus \varepsilon_n B \cap S|}{n} \right| \to 0 \quad c.s.$$

y lo mismo para $(T \oplus \varepsilon_n B)^c$.

Lema 4.5. *Para n suficientemente grande*

$$P\left(n^\delta \sup_{T\in\Gamma_n} |D(T) - \Delta(T)| > \varepsilon\right) \leq K|\Gamma_n| \exp\big\{ - k\varepsilon^2 n^{1-2\delta}\big\}.$$

Demostración. Definamos $\tilde{D}(T) := \mu\big(T \oplus \varepsilon_n B\big)\mu\big((T \oplus \varepsilon_n B)^c\big)S\big(d_{ni}^T : i = 1, \dots, n\big)$ y observemos que

$$\left|\tilde{D}(T) - D(T)\right| = \left|\left(\mu(T \oplus \varepsilon_n B) - \frac{|T \oplus \varepsilon_n B|}{n}\right)\mu\big((T \oplus \varepsilon_n B)^c\big) \right.$$
$$\left. + \left(\mu\big((T \oplus \varepsilon_n B)^c\big) - \frac{|(T \oplus \varepsilon_n B)^c|}{n}\right)\frac{|T \oplus \varepsilon_n B|}{n}\right| S\big(d_{ni}^T : i = 1, \dots, n\big).$$

Tenemos que por las propiedades **D.e)** y **D.c)** $S(d_{ni}^T : i = 1, \dots, n) \leq 1$ y, por el Lema 4.4 basta probar la desigualdad del Lema 4.5 para $\tilde{D}$. Anotemos $H_T^{ni} := \big|h_T(X_i) - \eta_T(X_i)\big|$ y $H_{T^c}^{ni} := \big|h_{T^c}(X_i) - \eta_{T^c}(X_i)\big|$ y $e_{ni}^T := H_T^{ni} + H_{T^c}^{ni}$ entonces $d_{ni}^T \leq e_{ni}^T + \delta_{ni}^T$ y $\delta_{ni}^T \leq e_{ni}^T + d_{ni}^T$, entonces usando las propiedades **D.e)** y **D.c)** de S tenemos que $\tilde{D}(T) - \Delta(T) \leq \mu\big(T \oplus \varepsilon_n B\big)\mu\big((T \oplus \varepsilon_n B)^c\big)S\big(e_{ni}^T : i = 1, \dots, n\big)$. Las mismas cotas se pueden obtener para $\Delta(T) - \tilde{D}(T)$ de donde usando **D.c)** se obtiene $|\tilde{D}(T) - \Delta(T)| \leq \mu(T \oplus \varepsilon_n B)S\big(H_T^{ni} : i = 1, \dots, n\big) + \mu\big((T \oplus \varepsilon_n B)^c\big)S\big(H_{T^c}^{ni} : i = 1, \dots, n\big)$.

$$\begin{aligned} H_T^{ni} \leq & \left|\frac{\sum_{j \in T \oplus \varepsilon_n B \cap S} \mathbb{I}_{\{X_j \leq X_i\}}}{|T \oplus \varepsilon_n B \cap S|} - F(X_i)\right| \frac{|T \oplus \varepsilon_n B \cap S|}{|T \oplus \varepsilon_n B|} \\ & + F(X_i)\left|\frac{|T \oplus \varepsilon_n B \cap S|}{|T \oplus \varepsilon_n B|} - \frac{\mu(T \oplus \varepsilon_n B \cap S)}{\mu(T \oplus \varepsilon_n B)}\right| \\ & + \left|\frac{\sum_{j \in T \oplus \varepsilon_n B \cap S^c} \mathbb{I}_{\{X_j \leq X_i\}}}{|T \oplus \varepsilon_n B \cap S^c|} - G(X_i)\right| \frac{|T \oplus \varepsilon_n B \cap S|}{|T \oplus \varepsilon_n B|} \\ & + G(X_i)\left|\frac{|T \oplus \varepsilon_n B \cap S^c|}{|T \oplus \varepsilon_n B|} - \frac{\mu(T \oplus \varepsilon_n B \cap S^c)}{\mu(T \oplus \varepsilon_n B)}\right|. \end{aligned} \tag{4.1}$$

El primer valor absoluto se acota por:

$$F_T := \sup_{x \in \mathbb{R}} \left|\frac{\sum_{j \in T \oplus \varepsilon_n B \cap S} \mathbb{I}_{\{X_j \leq x\}}}{|T \oplus \varepsilon_n B \cap S|} - F(x)\right|,$$

en el segundo sumando acotamos $F(X_i)$ por 1. De forma análoga se acota en tercer y cuarto sumando. Obtenemos $H_T^{ni} \leq H_T^n$ donde H_T^n no depende de i. Por argumentos análogos se puede acotar $H_{T^c}^{ni} \leq H_{T^c}^n$. Por lo tanto aplicando propiedades **D.e)**,**D.b)** y **D.d)** obtenemos que

$$\left|\tilde{D}(T) - \Delta(T)\right| \leq \mu\big(T \oplus \varepsilon_n B\big)H_T^n + \mu\big((T \oplus \varepsilon_n B)^c\big)H_{T^c}^n.$$

Si usamos la desigualdad 4.1, el lado derecho de esta desigualdad da lugar a 4 sumandos análogos. Uno de ellos es:

$$\left|\left(\mu(T \oplus \varepsilon_n B) - \frac{|T \oplus \varepsilon_n B|}{n}\right)\frac{|T \oplus \varepsilon_n B \cap S|}{n} + \frac{|T \oplus \varepsilon_n B \cap S|}{n}\right| F_T$$
$$+\left|\left(\mu(T \oplus \varepsilon_n B) - \frac{|T \oplus \varepsilon_n B|}{n}\right)\frac{|T \oplus \varepsilon_n B \cap S|}{n} + \left(\frac{|T \oplus \varepsilon_n B \cap S|}{n} - \mu(T \oplus \varepsilon_n B \cap S)\right)\right|.$$

Observemos que si aplicamos el Lema 4.4 es suficiente considerar

$$P\left(\sup_{T\in\Gamma_n}\frac{|T\oplus\varepsilon_n B\cap S|}{n}F_T>\varepsilon n^{-\delta}\right)\leq\sum_{T\in\Gamma_n}P\left(F_T>\varepsilon\frac{n^{1-\delta}}{|T\oplus\varepsilon_n B\cap S|}\right). \qquad (4.2)$$

Para acotar cada sumando de la ultima desigualdad usaremos la desigualdad de Dvoretzky, Kiefer, y Wolfowitz que dice que si $X_1, X_2, \dots, X_n$ son variables aleatorias i.i.d con distribución F y si

$$F_n(x)=\frac{1}{n}\sum_{i=1}^{n}\mathbb{I}_{[X_i<\infty)}(x) \quad x\in\mathbb{R},$$

entonces

$$P\left(\sup_{x\in\mathbb{R}}|F_n(x)-F(x)|>\varepsilon\right)\leq 2\exp\{-2n\varepsilon^2\}.$$

Por lo tanto si usamos esta desigualdad, condicionando a $|T\oplus\varepsilon_n B\cap S|$, tenemos que cada sumando en 4.2 se puede acotar por $K_1\exp\{-k_1\varepsilon^2 n^{1-2\delta}\}$. Esto concluye la demostración del Lema 4.5.

Los siguientes tres lemas tienen como objetivo relacionar la proximidad de T a θ, según la distancia $\partial(T,\theta)$, con la de $\Delta(T)$ a $\Delta(\theta)$.

Lema 4.6. *Para $T\in\Gamma_n$ si definimos $\delta_{ni}^S := |F(X_i)-G(X_i)|$ entonces*

$$\Delta(T)=\rho(T)S\big(\delta_{ni}^S : i=1,\dots,n\big),$$

siendo

$$\rho(T):=\big|\mu\big(T\oplus\varepsilon_n B\cap S\big)\mu\big((T\oplus\varepsilon_n B)^c\big)-\mu\big((T\oplus\varepsilon_n B)^c\cap S\big)\mu\big(T\oplus\varepsilon_n B\big)\big|.$$

Demostración. Observemos que

$$\delta_{ni}^T=\frac{\rho(T)\delta_{ni}^S}{\mu\big(T\oplus\varepsilon_n B\big)\mu\big((T\oplus\varepsilon_n B)^c\big)},$$

y por lo tanto el lema es consecuencia de la propiedad **D.b).**

Lema 4.7. *Para cada $T\in\Gamma_n$ tenemos que, si definimos*

$$\Delta(S)=\rho(S)S(\delta_{ni}^S : i=1,\dots,n)$$

y $\rho(S):=\mu(S)\mu(S^c)$ entonces

$$\Delta(T)\leq\Delta(S).$$

Demostración. Observemos que por el Lema 4.6 es suficiente probar que $\rho(T)\leq\mu(S)\mu(S^c)$. Consideremos la expresión dentro del módulo, en la definición de $\rho(T)$. Si esta expresión es positiva entonces $\rho(T)$ es igual a

$$\mu\big(T\oplus\varepsilon_n B\cap S\big)\mu\big((T\oplus\varepsilon_n B)^c\cap S\big)-\mu\big((T\oplus\varepsilon_n B)^c\cap S\big)\mu\big(T\oplus\varepsilon_n B\cap S\big)\leq\mu(S)\mu(S^c).$$

El caso negativo es análogo.

Lema 4.8. *Para $\gamma > 0$, si $\partial(S,T) < \gamma$ entonces $\rho(S)-\rho(T) \leq \gamma$ y si $\partial(S,T\oplus\varepsilon_n B) > \gamma$ entonces $\rho(S) - \rho(T) > k'\gamma > 0$.*

Demostración. Observemos que

$$\rho(S) - \rho(T) = \text{mín}\Big\{\mu\big(S^c \cap T \oplus \varepsilon_n B\big)\mu\big(S\big) + \mu\big(S \cap (T \oplus \varepsilon_n B)^c\big)\mu\big(S^c\big), \\ \mu\big(S \cap T \oplus \varepsilon_n B\big)\mu\big(S^c\big) + \mu\big(S^c \cap (T \oplus \varepsilon_n B)^c\big)\mu\big(S\big)\Big\}. \quad (4.3)$$

Si comparamos esta expresión con la definición de $\partial(S, T \oplus \varepsilon_n B)$, la primer implicancia es inmediata. Para la segunda, tenemos por hipótesis que

$$\text{máx}\Big\{\mu\big(S^c \cap T \oplus \varepsilon_n B\big), \mu\big(S \cap (T \oplus \varepsilon_n B)^c\big)\Big\} > \frac{\gamma}{2}$$

y que

$$\text{máx}\Big\{\mu\big(S \cap T \oplus \varepsilon_n B\big), \mu\big(S^c \cap (T \oplus \varepsilon_n B)^c\big)\Big\} > \frac{\gamma}{2}.$$

Por lo tanto usando la ecuación 4.3 tenemos que

$$\rho(S) - \rho(T) > \text{mín}\big\{\mu(S), \mu(S^c)\big\}\frac{\gamma}{2}.$$

El último lema, de caracter algo técnico que utilizaremos para demostrar el Teorema es el siguiente:

Lema 4.9. *Anotemos $\varphi_F := \int_{-\infty}^{\infty} |F(x)-G(x)|dF(x)$ y $\varphi_G := \int_{-\infty}^{\infty} |F(x)-G(x)|dG(x)$ y $\varphi := \varphi_F\mu(S) + \varphi_G\mu(S^c)$. Entonces $\varphi > 0$.*

Demostración. Por hipótesis tenemos que $\Phi := \big\{x \in \mathbb{R} : |F(x) - G(x)| > 0\big\} \neq \emptyset$. Es suficiente ver que $\int_\Phi dF(x) > 0$ o $\int_\phi dG(x) > 0$. El caso en que Φ contiene un punto de discontinuidad de F o G es trivial por lo que supondremos que F y G son continuas en cada $x \in \Phi$.

Tomemos $x_0 \in \Phi$ y supongamos $F(x_0) > G(x_0)$. Entonces $\sigma := \big\{y \in (-\infty, x_0] : F(x) > G(x)\forall x \in (y, x_0]\big\}$ es no vacío por continuidad. Anotemos $y_0 := \text{ínf}\{y \in \sigma\}$. Si $y_0 = -\infty$, entonces $(-\infty, x_0] \subset \Phi$ y $\int_\Phi dF(x) \geq F(x_0) > G(x_0) \geq 0$. Si $y_0 > -\infty$, entonces $F(x_0) \leq G(y_0)$ y también $(y_0, x_0] \subset \Phi$, de donde $\int_\Phi dF(x) \geq F(x_0)-F(y_0) \geq F(x_0) - G(x_0) > G(x_0) - G(y_0) \geq 0$.

Demostración del Teorema 2

Aplicando los Lemas 4.6, 4.7, 4.8 y la propiedad **D.f)** de S tenemos que:

$$\partial(S,T) > \gamma \Rightarrow |\Delta(S) - \Delta(T)| = \big(\rho(S) - \rho(T)\big)S\big(\delta_{ni}^S : i = 1, \ldots, n\big) > k'\gamma\delta_n^S.$$

donde $\delta_n^S = \sum_{i=1}^n \delta_{ni}^S/n$. Por lo tanto si $\omega := \varphi/2 > 0$:

$$P\big(\partial(S, \hat{S}_n) > \varepsilon n^{-\delta}\big) \leq P\big(|\Delta(S) - \Delta(\hat{S}_n)| > k'\varepsilon n^{-\delta}\delta_n^S\big) \leq \\ P\big(|\Delta(S) - \Delta(\hat{S}_n)| > k'\varepsilon n^{-\delta}\omega\big) + P\big(\delta_n^S < \omega\big). \quad (4.4)$$

Vamos a acotar el primer sumando, para eso

Lema 4.10. *Para n suficientemente grande:*

$$P\big(n^{\delta}\left|\Delta(\hat{S}_n)-\Delta(S)\right|>\varepsilon\big)\leq \tilde{K}|\Gamma_n|\exp\big\{-\tilde{k}\varepsilon^2 n^{1-2\delta}\big\}.$$

Demostración. Sea $T_n^0\in\Gamma_n$ el subconjunto que maximiza $\Delta(.)$ sobre Γ_n. Tenemos que, por el Lema 4.7 $\Delta(S)\geq\Delta(T_n^0)\geq\Delta(T)\quad\forall T\in\Gamma_n$, y por definición tenemos que $D(\hat{S}_n)\geq D(T)\quad\forall T\in\Gamma_n$.

$$|\Delta(\hat{S}_n)-\Delta(S)|\leq|\Delta(\hat{S}_n)-D(\hat{S})|+|D(\hat{S}_n)-\Delta(T_n^0)|+|\Delta(T_n^0)-\Delta(S)|. \quad (4.5)$$

$|D(\hat{S}_n)-\Delta(T_n^0)|$ se puede acotar por $\sup_{T\in\Gamma_n}|D(T)-\Delta(T)|$ ya que $D(\hat{S}_n)\geq\Delta(T_n^0)\geq\Delta(\hat{S}_n)$ o $\Delta(T_n^0)\geq D(\hat{S}_n)\geq D(T_n^0)$. La misma cota se puede aplicar en $|\Delta(\hat{S}_n)-D(\hat{S}_n)|$.

Usando los Lemas 4.6 y 4.7 y las propiedades **D.e)** y **D.c)** de S obtenemos que

$$|\Delta(T_n^0)-\Delta(S)|\leq\Delta(S)-\Delta(T)=\big(\rho(S)-\rho(T)\big)S\big(\delta_{ni}^S: i=1,\dots,n\big)\leq\rho(S)-\rho(T).$$

Por la hipótesis **T2)** sabemos que para n suficientemente grande $\partial(S,T_n)<\varepsilon n^{-\delta}$ para algún $T_n\in\Gamma_n$. Por lo tanto si usamos el Lema 4.8 hemos acotado $|\Delta(T_n^0)-\Delta(S)|$ por ε. El Lema 4.10 es entonces una aplicación del Lema 4.5. □

Para el segundo observemos que $P(\delta_n^S<\omega)\leq P(|\delta_n^S-\varphi|>\omega)$. Anotemos $\overline{\delta}_n^S:=\sum_{i\in S}\delta_{ni}^S/|S|$ y $\underline{\delta}_n^S:=\sum_{i\in S^c}\delta_{ni}^S/|S^c|$ por lo tanto tenemos $\delta_n^S=\overline{\delta}_n^S(|S|/n)+\underline{\delta}_n^S(|S|/n)$; escribimos

$$|\delta_n^S-\varphi|\leq\overline{\delta}_n^S\left|\frac{|S|}{n}-\mu(S)\right|+\mu(S)|\overline{\delta}_n^S-\varphi_F|+\underline{\delta}_n^S\left|\frac{|S^c|}{n}-\mu(S^c)\right|+\mu(S^c)|\underline{\delta}_n^S-\varphi_G|.$$

El primer y tercer sumando se acotan usando el Lema 4.4. Consideremos el segundo sumando, un argumento análogo se aplica al cuarto. Por la ecuación (2.3) de [12] tenemos que

$$P\Big(|\overline{\delta}_n^S-\varphi_F|>\omega/4\Big)=E\Big(P\big(|\overline{\delta}_n^S-\varphi_F|>\omega/4\,|\,|S|\big)\Big)\leq E\Big(\exp\big\{-\tilde{c}|S|\big\}\Big)=$$
$$\sum_{j=0}^{n}\left(\frac{\mu(S)}{e^{\tilde{c}}}\right)^j(1-\mu(S))^{n-j}C_j^n=\left(1-\mu(S)\left(1-\frac{1}{e^{\tilde{c}}}\right)\right)^n=\exp\{-nk\}$$

con k constante positiva.Al juntar la cota de Höeffding con la que obtuvimos de aplicar el Lema 4.9 se concluye la demostración del Teorema 2.

4.2. Una generalización del caso anterior

En esta sección veremos que es posible generalizar los resultados de la sección anterior para el caso en que (siguiendo la notación del capítulo anterior) las variables X_i toman valores en $\mathbb{R}^d$. Vamos a suponer que F o G son absolutamente continuas respecto de la medida de Lebesgue en $\mathbb{R}^d$. Tenemos que definir entonces quien será el estimador a maximizar, y mas precisamente, quienes son los d_i. De forma análoga a lo hecho en el capítulo anterior, definimos la familia Γ_n, y usaremos la misma notación que antes. Definimos además para $T \in \Gamma_n$ y $x \in \mathbb{R}^d$ las distribuciones empíricas:

$$h_T(x) := \frac{1}{|T \oplus \varepsilon_n B|} \sum_{i \in T \oplus \varepsilon_n B} \mathbb{I}_{\{X_i \leq x\}}$$

y

$$h_{T^c}(x) := \frac{1}{|(T \oplus \varepsilon_n B)^c|} \sum_{i \in (T \oplus \varepsilon_n B)^c} \mathbb{I}_{\{X_i \leq x\}},$$

donde entendemos por $\{X_i \leq x\}$ que si $X_i = (X_{i1}, \ldots, X_{id})$ y $x = (x_1, \ldots, x_d)$ entonces $X_{i1} \leq x_1, \ldots, X_{id} \leq x_d$. Observemos que con esta notación es posible definir los d_i y por lo tanto D de forma totalmente análoga a lo hecho anteriormente.

4.2.1. Hipótesis sobre Γ_n

Vamos a pedir que se cumplan las condiciones **T1)** y **T2)** y además una condición algo mas fuerte que **T3)**:

T'3) $\forall \gamma > 0$ $|\Gamma_n|(n+1)^d \exp(-\gamma n^{1-2\delta}) < n^{-\alpha}$ con $\alpha > 1$ siendo δ como en **T2**.

Lo que demostraremos será entonces,

Teorema 4.11. ***Consistencia:*** *bajo las hipótesis **T1)**, **T2)** y **T'3))***

$$n^\delta \partial(S, \hat{S}_n) \to 0 \textit{ cuando } n \textit{ tiende a infinito}, \quad c.s.$$

Teorema 4.12. ***Error en Probabilidad:*** *bajo las hipótesis **T1)**, **T2)** y **T'3)** se cumple que*

$$P\big(\partial(S, \hat{S}_n) > \varepsilon\big) \leq c_1 |\Gamma_n| (n+1)^d \exp\big\{ - c_2 \varepsilon^2 n \big\},$$

para n suficientemente grande, siendo c_1 y c_2 constantes positivas

Demostración. De forma análoga a lo hecho anteriormente, vamos a demostrar el teorema 4.11 y 4.12 a partir de:

$$P\left(n^\delta \partial(S, \hat{S}_n) > \varepsilon\right) \leq c_1 |\Gamma_n| (n+1)^d \exp\big\{ - c_2 \varepsilon^2 n^{1-2\delta} \big\}$$

para n suficientemente grande, (las constantes que aparecen no son necesariamente las mismas que las del teorema anterior)
De la desigualdad anterior, tomando $\delta = 0$ se concluye el teorema 4.12. El teorema 4.11 es una aplicación directa del Lema de Borel-Cantelli y de la hipótesis **T'3)**.
Observemos que el Lema 4.4 sigue siendo válido ya que no depende de las X_i y que la definición 4.3 se generaliza de forma trivial al caso que estamos considerando.

En lugar del Lema 4.5 veremos que siguiendo las mismas ideas de éste, podemos demostrar fácilmente

Lema 4.13. *Para n suficientemente grande*

$$P\left(n^{\delta}\sup_{T\in\Gamma_n}|D(T)-\Delta(T)|>\varepsilon\right)\leq K|\Gamma_n|(n+1)^d\exp\big\{-k\varepsilon^2 n^{1-2\delta}\big\}.$$

Demostración. Observemos que la demostración no cambia si en lugar de considerar X a valores reales la consideramos a valores en $\mathbb{R}^d$, salvo que en el momento de acotar

$$P\left(F_T>\varepsilon\frac{n^{1-\delta}}{|T\oplus\varepsilon_n B\cap S|}\right),$$

no podemos usar la desigualdad de Dvoretzky, Kiefer, y Wolfowitz, ya que ésta es para variables a valores reales. En lugar de eso vamos a usar los resultados del apéndice referentes a la teoría de Vapnik-Chervonenkis, en particular lo anterior se acota usando el teorema A.4 y el A.5 y es de allí de donde surge el factor $(n+1)^d$ que aparece en la tésis del Lema 4.13. □

Es claro que los Lemas 4.6, 4.7 y 4.8 también son válidos en este caso, y en lugar del Lema 4.10, es posible,demostrar

Lema 4.14. *Para n suficientemente grande:*

$$P\big(n^{\delta}\left|\Delta(\hat{S}_n)-\Delta(S)\right|>\varepsilon\big)\leq\tilde{K}|\Gamma_n|(n+1)^d\exp\big\{-\tilde{k}\varepsilon^2 n^{1-2\delta}\big\}.$$

Demostración. La demostración se sigue de manera análoga a la del Lema 4.10, aplicando en 4.5 el Lema 4.13 a los 2 últimos sumandos. □

Es posible también, demostrar el Lema 4.9 para el caso en que F y G son distribuciones en $\mathbb{R}^d$ distintas y al menos una de ellas es absolutamente continua respecto de la medida de Lebesgue. La demostración de los teoremas se concluyen entonces a partir de considerar éstas generalizaciones.

Apéndice A

Apéndice

A.1. Teoría de Vapnik-Chervonenkis

La idea de éste capítulo es presentar brevemente la Tería de Vapnik-Chervonenkis, que nos permite generalizar el conocido Teorema fundamental de la estadística de Glivenco-Cantelli, y encontrar cotas exponenciales de la diferencia

$$\sup_{z\in\mathbb{R}^d} |F_n(z) - F(z)|.$$

El contenido de éstas secciones pueden encontrarse en [10]. Vamos a presentar primero una demostración del Teorema de Glivenko-Cantelli que contiene ideas que luego se usaran en el caso general.

Teorema A.1. *Sean $Z_1, Z_2, \dots, Z_n$ variables aleatorias independientes idénticamente distribuidas con función de distribución $F(z) = P(\{Z_1 \leq z\})$. Denotemos la distribución empírica como:*

$$F_n(z) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{Z_i \leq z\}},$$

entonces

$$P\left\{\sup_{z\in\mathbb{R}} |F(z) - F_n(z)| > \varepsilon\right\} \leq 8(n+1)e^{-n\varepsilon^2/32},$$

en particular, por el lema de Borel-Cantelli

$$\lim_{n\to+\infty}\sup |F(z) - F_n(z)| = 0 \quad \text{con probabilidad 1.}$$

Demostración. Vamos a introducir algo de notación que usaremos a lo largo de esta sección. $\nu(A) = P\{Z_1 \in A\}$ y $\nu_n(A) = (1/n)\sum_{j=1}^{n} \mathbb{I}_{\{Z_j\in A\}}$ para todo conjunto medible $A \subset \mathbb{R}$. Denotemos como $\mathcal{A}$ la clase de los conjuntos de la forma $(-\infty, z]$ con $z \in \mathbb{R}$. Con esta notación

$$\sup_{z\in\mathbb{R}} |F(z) - F_n(z)| = \sup_{A\in\mathcal{A}} |\nu_n(A) - \nu(A)|.$$

Vamos a demostrar el teorema en varias etapas, siguiendo las ideas de simetrización de Dudley (1978) y Pollard (1984). Asumiremos que $n\varepsilon^2 \geq 2$ en caso contrario la cota es trivial.

PASO 1. Simetrización respecto de un remuestreo de Z. Definimos las variables $Z'_1, \ldots, Z'_n \in \mathbb{R}$ tal que $Z_1, \ldots, Z_n, Z'_1, \ldots, Z'_n$ son independientes e idénticamente distribuidas. Denotemos como ν'_n la medida empírica correspondiente a la nueva muestra:

$$\nu'_n(A) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}_{\{Z'_i \in A\}}.$$

Entonces, para $n\varepsilon^2 \geq 2$ tenemos que

$$P\left\{\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu(A)| > \varepsilon\right\} \leq 2P\left\{\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu'_n(A)| > \frac{\varepsilon}{2}\right\}.$$

Para ver esto, sea $A^* \in \mathcal{A}$ para el cual $|\nu_n(A^*) - \nu(A^*)| > \varepsilon$ si tal conjunto existe, y A^* cualquiera, fijo en $\mathcal{A}$, si no. Entonces

$$\begin{aligned}
P\left\{\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu'_n(A)| > \varepsilon/2\right\} &\geq P\big\{|\nu_n(A^*) - \nu'_n(A^*)| > \varepsilon/2\big\} \\
&\geq P\left\{|\nu_n(A^*) - \nu(A^*)| > \varepsilon, |\nu'_n(A^*) - \nu(A^*)| < \frac{\varepsilon}{2}\right\} \\
&= E\left\{\mathbb{I}_{\left\{|\nu_n(A^*)-\nu(A^*)|>\varepsilon\right\}} P\left\{|\nu'_n(A^*) - \nu(A^*)| < \frac{\varepsilon}{2}\Big| Z_1, \ldots, Z_n\right\}\right\}.
\end{aligned}$$

La probabilidad dentro de la esperanza puede ser acotada usando la desigualdad de Chebyshev:

$$\begin{aligned}
P\left\{|\nu'_n(A^*) - \nu(A^*)| < \frac{\varepsilon}{2}\Big| Z_1, \ldots, Z_n\right\} &\geq 1 - \frac{\nu(A^*)(1-\nu(A^*))}{n\varepsilon^2/4} \\
&\geq 1 - \frac{1}{n\varepsilon^2} \geq \frac{1}{2}
\end{aligned}$$

Ya que hemos supuesto que $n\varepsilon^2 \geq 2$. Obtuvimos entonces

$$\begin{aligned}
P\left\{\sup_{A\in\mathcal{A}} |\nu(A) - \nu'(A)| > \varepsilon/2\right\} &\geq \frac{1}{2}P\big\{|\nu_n(A^*) - \nu(A^*)|\big\} \\
&\geq \frac{1}{2}P\left\{\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu(A)| > \varepsilon\right\}
\end{aligned}$$

PASO 2. Simetrización por signos aleatorios. Sea $\sigma_1, \ldots, \sigma_n$ variables aleatorias independientes idénticamente distribuidas, con $P(\sigma_1 = -1) = P(\sigma_1 = 1)1/2$, independientes de $Z_1, \ldots, Z_n, Z'_1, \ldots, Z'_n$. Como las $Z_1, Z'_1, \ldots, Z_n, Z'_n$ son independientes identicamente distribuidas, la distribución de

$$\sup_{A\in A}\left|\sum_{i=1}^{n}\big(\mathbb{I}_A(Z_i) - \mathbb{I}_A(Z'_i)\big)\right|,$$

es la misma que la de

$$\sup_{A\in A}\left|\sum_{i=1}^{n}\sigma_i\big(\mathbb{I}_A(Z_i)-\mathbb{I}_A(Z_i')\big)\right|,$$

usando entonces el paso 1,

$$\begin{aligned}
P\left\{\sup_{A\in\mathcal{A}}|\nu_n(A)-\nu(A)|>\varepsilon\right\} &\leq 2P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\big(\mathbb{I}_A(Z_i)-\mathbb{I}_A(Z_i')\big)\right|>\frac{\varepsilon}{2}\right\}\\
&= 2P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\big(\mathbb{I}_A(Z_i)-\mathbb{I}_A(Z_i')\big)\right|>\frac{\varepsilon}{2}\right\}\\
&\leq 2P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i)\right|>\frac{\varepsilon}{4}\right\}+2P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i')\right|>\frac{\varepsilon}{4}\right\}\\
&= 4P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i)\right|>\frac{\varepsilon}{4}\right\}.
\end{aligned}$$

PASO 3. Condicionar a $Z_1,\ldots,Z_n$. Para acotar la probabilidad anterior vamos a condicionar a $Z_1,\ldots,Z_n$. Fijemos $z_1,\ldots,z_n\in\mathbb{R}$ y observemos que al variar z en $\mathbb{R}$ el número de vectores diferentes $\big(\mathbb{I}_{\{Z_1\leq z\}},\ldots,\mathbb{I}_{\{Z_n\leq z\}}\big)$ es a lo sumo $n+1$, luego condicionando a $Z_1,\ldots,Z_n$ podemos escribir

$$P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i)\right|>\frac{\varepsilon}{4}\Big|Z_1,\ldots,Z_n\right\}\leq(n{+}1)\sup_{A\in\mathcal{A}}P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i)\right|>\frac{\varepsilon}{4}\Big|Z_1,\ldots,Z_n\right\}.$$

PASO 4. Desigualdad de Hoeffding. Con $z_1,\ldots,z_n$ fijos, $\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(z_i)$ es la suma de n variables aleatorias independientes con media 0, a valores entre -1 y 1, por lo tanto si aplicamos la desigualdad de Hoeffding obtenemos

$$P\left\{\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i)\right|>\frac{\varepsilon}{4}\Big|Z_1,\ldots,Z_n\right\}\leq 2e^{-n\varepsilon^2/32}.$$

Entonces

$$P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i)\right|>\frac{\varepsilon}{4}\Big|Z_1,\ldots,Z_n\right\}\leq 2(n+1)e^{-n\varepsilon^2/32}.$$

Si tomamos valor esperado de ambos lados

$$P\left\{\sup_{A\in\mathcal{A}}\frac{1}{n}\left|\sum_{i=1}^{n}\sigma_i\mathbb{I}_A(Z_i)\right|>\frac{\varepsilon}{4}\right\}\leq 2(n+1)e^{-n\varepsilon^2/32}.$$

□

Veamos ahora como generalizar el resultado anterior para el caso en que las variables Z_i toman valores en $\mathbb{R}^d$, para eso veamos primero algunas definiciones.

Definición A.2. Sea $\mathcal{A}$ una familia de subconjuntos medibles. Para $(z_1, \ldots, z_n) \in \{\mathbb{R}^d\}^n$, sea $N_{\mathcal{A}}(z_1, \ldots, z_n)$ el número de conjuntos distintos, de la forma

$$\big\{\{z_1, \ldots, z_n\} \cap A : A \in \mathcal{A}\big\},$$

definimos el número

$$s(\mathcal{A}, n) := \max_{(z_1,\ldots,z_n)\in\{\mathbb{R}^d\}^n} N_{\mathcal{A}}(z_1, \ldots, z_n).$$

Observemos que claramente $s(\mathcal{A}, n) \leq 2^n$ y si $s(\mathcal{A}, k) < 2^k$ para algún entero k entonces $s(\mathcal{A}, n) < 2^n$ para todo $n > k$.

Definición A.3. Sea $\mathcal{A}$ una familia de conjuntos tal que $|\mathcal{A}| \geq 2$. El mayor entero $k \geq 1$ para el cual $s(\mathcal{A}, k) = 2^k$ se denotará $V_{\mathcal{A}}$ y se denomina dimensión de Vapnik-Chernonenkis de $\mathcal{A}$. Si $s(\mathcal{A}, n) = 2^n$ para todo n entonces $V_{\mathcal{A}} = \infty$. En el caso en que $S(\mathcal{A}, k) = 2^k$ decimos que $\mathcal{A}$ divide completamente a $\{z_1, \ldots, z_n\}$.

Es claro que si por ejemplo tomamos $\mathcal{A}$ como los subconjuntos de la forma $(-\infty, x]$ con $x \in \mathbb{R}$ entonces $s(\mathcal{A}, 2) = 3 < 2^2$ y $V_{\mathcal{A}} = 1$.

Teorema A.4. *(Vapnik-Chernonenkis (1971)). Sea ν una probabilidad, y una familia $\mathcal{A}$ de conjuntos, entonces, para todo n y para todo $\varepsilon > 0$ tenemos que*

$$P\left\{\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu(A)| > \varepsilon\right\} \leq 8s(\mathcal{A}, n)e^{-n\varepsilon^2/32}.$$

Demostración. La demostración sigue las ideas de la demostración del teorema anterior. De forma análoga asumimos que $n\varepsilon^2 \geq 2$, en los primeros 2 pasos demostramos que

$$P\left\{\sup_{A\in\mathcal{A}} |\nu_n(A) - \nu(A)| > \varepsilon\right\} \leq 4P\left\{\sup_{A\in\mathcal{A}} \frac{1}{n}\left|\sum_{i=1}^n \sigma_i \mathbb{I}_A(Z_i)\right| > \frac{\varepsilon}{4}\right\}.$$

Esto puede hacerse de manera totalmente análoga a lo hecho en el teorema anterior. Veamos el paso 3.
PASO 3. Condicionar. Para acotar la probabilidad

$$P\left\{\sup_{A\in\mathcal{A}} \frac{1}{n}\left|\sum_{i=1}^n \sigma_i \mathbb{I}_A(Z_i)\right| > \frac{\varepsilon}{4}\right\},$$

nuevamente condicionamos a $Z_1, \ldots, Z_n$. Fijemos $z_1, \ldots, z_n \in \mathbb{R}^d$ y observemos que al variar $A \in \mathcal{A}$ el número de vectores distintos $\big(\mathbb{I}(z_1), \ldots, \mathbb{I}(z_n)\big)$ es justamente el número de subconjuntos distintos, de $\{z_1, \ldots, z_n\}$, que se producen al intersectar con elementos de $\mathcal{A}$. Por definición este número no excede $s(\mathcal{A}, n)$.

$$P\left\{\sup_{A\in\mathcal{A}} \frac{1}{n}\left|\sum_{i=1}^n \sigma_i \mathbb{I}_A(Z_i)\right| > \frac{\varepsilon}{4}\middle| Z_1, \ldots, Z_n\right\} \leq s(\mathcal{A}, n) \sup_{A\in\mathcal{A}} P\left\{\frac{1}{n}\left|\sum_{i=1}^n \sigma_i \mathbb{I}_A(Z_i)\right| > \frac{\varepsilon}{4}\middle| Z_1, \ldots, Z_n\right\}.$$

Por lo tanto es suficiente acotar la probabilidad condicional

$$P\left\{\frac{1}{n}\left|\sum_{i=1}^n \sigma_i \mathbb{I}_A(Z_i)\right| > \frac{\varepsilon}{4}\middle| Z_1, \ldots, Z_n\right\}.$$

Finalmente si aplicamos la desigualdad de Hoeffding obtenemos

$$P\left\{\sup_{A\in\mathcal{A}}|\nu_n(A)-\nu(A)|>\varepsilon\right\}\leq 8s(\mathcal{A},n)e^{-n\varepsilon^2/32}.$$

□

Veamos ahora un teorema que nos permitirá acotar superiormente $s(\mathcal{A},n)$ en términos de la dimensión de Vapnik-Chervonenkis de la familia $\mathcal{A}$.

Teorema A.5. *Si $\mathcal{A}$ es una familia de conjuntos con dimensión de Vapnik-Chervonenkis $V_{\mathcal{A}}$ entonces para todo n*

$$s(\mathcal{A},n)\leq\sum_{i=1}^{V_{\mathcal{A}}}\binom{n}{i}$$

Demostración. Recordemos que $s(\mathcal{A},n)=\max\limits_{(x_1,\ldots,x_n)}N_{\mathcal{A}}(x_1,\ldots,x_n)$, donde

$$N_{\mathcal{A}}(x_1,\ldots,x_n)=\Big|\big\{\{x_1,\ldots,x_n\}\cap A:A\in\mathcal{A}\big\}\Big|.$$

Claramente es suficiente demostrar que para todo $x_1,\ldots,x_n$

$$N_{\mathcal{A}}(x_1,\ldots,x_n)\leq\sum_{i=1}^{V_A}\binom{n}{i}.$$

Observemos que por la definición de N_A, podemos asumir que $\mathcal{A}$ es una familia de subconjuntos de $\{x_1,\ldots,x_n\}$ con dimensión V_A. En éste caso $s(\mathcal{A},n)=|\mathcal{A}|$. Vamos a demostrar el teorema por inducción en n y $V_{\mathcal{A}}$. El resultado es obvio para $n=1$ y para cualquier clase con $V_{\mathcal{A}}\geq 1$. Es claro que también se cumple para $n\geq 1$ si $V_{\mathcal{A}}=0$, ya que en este caso $s(\mathcal{A},n)=1$ para todo n. Asumiremos entonces que $V_{\mathcal{A}}\geq 1$. Por hipótesis de inducción asumiremos que el resultado es cierto para todo $k<n$, para toda familia de subconjuntos de $\{x_1,\ldots,x_n\}$ de dimensión menor o igual que $V_{\mathcal{A}}$, y para n y toda familia de dimensión menor que $V_{\mathcal{A}}$. Definimos las siguientes clases de subconjuntos de $\{x_,\ldots,x_n\}$.

$$\mathcal{A}'=\{A-\{x_n\}:A\in\mathcal{A}\},$$

y

$$\hat{\mathcal{A}}=\{\hat{A}\in\mathcal{A}:x_n\notin\hat{A},\hat{A}\cup\{x_n\}\in\mathcal{A}\}.$$

Observemos que tanto $\mathcal{A}'$ como $\hat{\mathcal{A}}$ están formados por subconjuntos de $\{x_1,\ldots,x_{n-1}\}$. Veamos que $|\mathcal{A}|=|\mathcal{A}'|+|\hat{\mathcal{A}}|$, escribimos

$$\mathcal{A}'=\{A-\{x_n\}:x_n\in A,A\in\mathcal{A}\}\cup\{A-\{x_n\}:x_n\notin A,A\in\mathcal{A}\}=B_1\cup B_2.$$

Entonces

$$\begin{aligned}|\mathcal{A}'| &= |B_1|+|B_2|-|B_1\cap B_2|\\ &= \big|\{A-\{x_n\}:x_n\in A,A\in\mathcal{A}\}\big|+\big|\{A-\{x_n\}:x_n\notin\mathcal{A},A\in\mathcal{A}\}\big|-\big|\hat{\mathcal{A}}\big|\\ &= \big|\{A:x_n\in A,A\in\mathcal{A}\}\big|+\big|\{A:x_n\notin A,A\in\mathcal{A}\}\big|-\big|\hat{\mathcal{A}}\big|\\ &= \big|\mathcal{A}\big|-\big|\hat{\mathcal{A}}\big|.\end{aligned}$$

Como $|\mathcal{A}'| \leq |\mathcal{A}|$ y $\mathcal{A}'$ es una familia de subconjuntos de $\{x_1, \ldots, x_n\}$ podemos aplicar la hipótesis de inducción y obtenemos que

$$|\mathcal{A}'| = s(A', n-1) \leq \sum_{i=1}^{V_{\mathcal{A}}} \binom{n-1}{i}.$$

Veremos que $V_{\hat{\mathcal{A}}} \leq V_{\mathcal{A}} - 1$, lo cual implica que

$$|\hat{\mathcal{A}}| = s(\hat{\mathcal{A}}, n-1) \leq \sum_{i=1}^{V_{\mathcal{A}}-1} \binom{n-1}{i},$$

por hipótesis de inducción. Para ver ésto consideremos un conjunto $S \subset \{x_1, \ldots, x_{n-1}\}$ tal que $\hat{\mathcal{A}}$ lo divide completamente. Veremos que $S \cup \{x_n\}$ es divido completamente por $\mathcal{A}$. Para eso, tenemos que ver que para todo conjunto $S' \subset S$ y $S' \cup \{x_n\}$ es la intersección de $S \cup \{x_n\}$ y un elemento de $\mathcal{A}$. Como S es dividido completamente por $\hat{\mathcal{A}}$, si $S' \subset S$ entonces existe un $\hat{A} \in \hat{\mathcal{A}}$ tal que $S' = S \cap \hat{A}$. Pero como por definición $x_n \notin \hat{A}$, tenemos que

$$S' = (S \cup \{x_n\}) \cap \hat{A},$$

y

$$S' \cup \{x_n\} = (S \cup \{x_n\}) \cap (\hat{A} \cup \{x_n\}).$$

Por definición de $\hat{\mathcal{A}}$, $\hat{A}$ y $\hat{A} \cup \{x_n\}$ están en $\mathcal{A}$ y por lo tanto $S \cup \{x_n\}$ es dividido completamente por $\mathcal{A}$. Como el cardinal de cualquier conjunto dividido completamente por $\mathcal{A}$ no puede exceder $V_{\mathcal{A}}$ obtenemos que $|S| \leq V_{\mathcal{A}} - 1$. Pero como S era un elemento arbitrario, dividido completamente por $\hat{\mathcal{A}}$ obtenemos que $V_{\hat{\mathcal{A}}} \leq V_{\mathcal{A}} - 1$. Por lo tanto hemos demostrado que

$$s(\mathcal{A}, n) = |\mathcal{A}| = |\mathcal{A}'| + |\hat{\mathcal{A}}| \leq \sum_{i=1}^{V_{\mathcal{A}}} \binom{n-1}{i} + \sum_{i=1}^{V_{\mathcal{A}}-1} \binom{n-1}{i},$$

finalmente la conclusión se sigue de la identidad $\binom{n}{i} = \binom{n-1}{i} + \binom{n-1}{i-1}$. □

Observemos que del resultado anterior tenemos que $s(\mathcal{A}, n) \leq (n+1)^{V_{\mathcal{A}}}$, si $V_{\mathcal{A}} < \infty$. Veamos que si tomamos como $\mathcal{A}$ la familia de conjuntos de la forma $(-\infty, y_1] \times \cdots \times (-\infty, y_d]$ entonces $V_{\mathcal{A}} = d$. Para eso vamos a demostrar el siguiente teorema, que tiene interés por si mismo, y del cual se deduce lo que queremos.

Teorema A.6.

(i) Si $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$, entonces $s(\mathcal{A}, n) \leq s(\mathcal{A}_1, n) + s(\mathcal{A}_2, n)$.

(ii) Dada una clase $\mathcal{A}$ definimos $\mathcal{A}_c = \{A^c : A \in \mathcal{A}\}$. Entonces $s(A_c, n) \leq s(\mathcal{A}, n)$.

(iii) Si $\mathcal{A} = \left\{\hat{A} \cap \tilde{A} : \hat{A} \in \hat{\mathcal{A}}, \tilde{A} \in \tilde{\mathcal{A}}\right\}$. Entonces $s(\mathcal{A}, n) \leq s(\hat{\mathcal{A}}, n) s(\tilde{\mathcal{A}}, n)$.

(iv) Si $\mathcal{A} = \left\{\hat{A} \cup \tilde{A} : \hat{A} \in \hat{\mathcal{A}}, \tilde{A} \in \tilde{\mathcal{A}}\right\}$. Entonces $s(\mathcal{A}, n) \leq s(\hat{\mathcal{A}}, n) s(\tilde{\mathcal{A}}, n)$.

(v) Si $\mathcal{A} = \left\{ \hat{A} \times \tilde{A} : \hat{A} \in \hat{\mathcal{A}}, \tilde{A} \in \tilde{\mathcal{A}} \right\}$*. Entonces* $s(\mathcal{A}, n) \leq s(\hat{\mathcal{A}}, n) s(\tilde{\mathcal{A}}, n)$.

Demostración. (i), (ii) y (v) son triviales. Para demostrar (iii) fijemos n puntos $x_1, \ldots, x_n$ y supongamos que con $\hat{\mathcal{A}}$ tomamos para $N \leq s(\hat{\mathcal{A}}, n)$ los conjuntos $C_1, \ldots, C_N$. Luego $\tilde{\mathcal{A}}$ separa en C_i a lo sumo $s(\tilde{A}, |C_i|)$ subconjuntos. Por lo tanto los subconjuntos de la forma $\hat{A} \cap \tilde{A}$ separan a lo sumo

$$\sum_{i=1}^{N} s(\tilde{\mathcal{A}}, |C_i|) \leq s(\hat{\mathcal{A}}, n) s(\hat{\mathcal{A}}, n),$$

subconjuntos. Hemos usado que $s(\mathcal{A}, n) \leq s(\mathcal{A}, n+m)$.

(iv) es inmediato a partir de (ii) y (iii). □

Observación A.7. *Es claro que de la parte* (v) *del teorema anterior se sigue que si* $\mathcal{A}$ *es la familia de conjuntos de la forma* $(-\infty, y_1] \times \cdots \times (-\infty, y_d]$ *entonces* $V_{\mathcal{A}} = d$*. Otra demostración posible para lo mismo, que usa la forma particular de los subconjuntos de la familia que estamos considerando es observar que si tenemos* $x_1, \ldots, x_d, x_{d+1}$ *puntos en* $\mathbb{R}^d$ *y consideramos el conjunto construido de la siguiente manera, tomemos* $x_{(1)}$ *de entre* $x_1, \ldots, x_{d+1}$ *cuya primera coordenada sea mayor (en caso de haber 2 o más tomamos uno cualesquiera de ellos), luego* $x_{(2)}$ *de entre los restantes, como el que tenga mayor segunda coordenada, y así hasta* $x_{(d)}$*, tomemos el conjunto formado por estos puntos,* $x_{(1)}, \ldots, x_{(d)}$*, es claro que el punto que no tomamos no se puede separar de este subconjunto, con elementos de la forma* $(-\infty, y_1] \times \cdots \times (-\infty, y_d]$*, para ningún* $(y_1, \ldots, y_d)$*. Esto prueba que* $VA < d + 1$*, es claro que* $VA \geq d$*.*

Bibliografía

[1] AMBROSIO, L., COLESANTI, A. y VILLA, E. (2007). One-side Minkoski content for some classes of closed sets and applications to stochastic geometry. Manuscript.

[2] ARMENDÁRIZ, I., CUEVAS, A., y FRAIMAN, R. (2008). Nonparametric estimation of boundary measures and related functionals: asymptotic results. *Applied Probability Trust*

[3] BHATTACHARYA, R. N. y RANGA RAO, R. (1976). Normal aproximations and Asymptotic Expansions. Wiley, New York.

[4] CARLSTEIN, E. y KRISHNAMOORTHY, C. (1992). Boundary Estimation, *Journal of the American Statisttical Association*, Vol. 87, No. 418.

[5] CUEVAS, A., FRAIMAN, R. y RODRIGUEZ-CASAL, A. (2006). A nonparametric approach to the estimation of lengths and surface areas, *Ann. Statist.* 0, Vol. 0 No. 00, 1-23.

[6] CUEVAS, A. y RODRIGUEZ-CASAL, A. (2004). On boundary estimation. *Adv. in Appl. Probab.* **36** 340-354. MR2058139 (2005b:62068).

[7] CUEVAS, A. y FRAIMAN, R. (1997). A plug-in approach to support estimation. *Ann. Statist.* **25** 2300-2312. MR1604449(99m:62040).

[8] CUEVAS, A., FEBRERO, M. y FRAIMAN, R. (1991). Estimating the number of clusters. Can. J. Statist. 28, 367-382.

[9] DEVROYE, L. y WISE, G.L. (1980). Detection of abnormal behavior via nonparametric estimation of the support. SIAM J. *Appl. Math* **38**, 480-488.

[10] DEVROYE, L., GYÖRFI, L., y LUGOSI, G. A probabilistic Theory of Pattern Recognition. *Springer-Verkag* ISBN 0-387-94618-7.

[11] DÜMBGEN, L. y WALTHER, G. (1996). Rates of convergence for random approximations of convex sets, Adv. Appl. Probab. 28 384-393.

[12] HÖEFFDING, W. (1963). Probability Inequalities for Sums of Bounded Random Variables, Journal of the American Statistical Association, 58, 13-30.

[13] FEDERER, H. (1959). Curvature measures. *Trans. Amer. Math. Soc.* **93** 418-491. MR0110078.

[14] FEDERER, H. (1969). Geometric measure theory. *Springer Verlag* ISBN 3-540-60656-4.

[15] MATTILA, P. (1995). Geometry of sets and measures in Euclidean spaces. *Cambridge Studies in Advanced Mathematics*, Vol. **44**. Cambridge University Press, Cambridge. Fractals and rectifiability. MR1333890(96H:28006).

[16] PATEIRO, B. (2008). Estimación de conjuntos bajo restricciónes de forma de tipo convexo. Tésis de doctorado. Santiago de compostela.

[17] PRAKASA RAO, B. L. S. (1983) Nonparametric Functional Estimation. Academic Press, New York.

[18] RODRIGUEZ-CASAL, A. (2002). Estimación de conjuntos y sus fronteras, un enfoque geométrico. Tésis de doctorado. Santiago de Compostela.

[19] RODRIGUEZ-CASAL, A. (2006). Set estimation under convexity type assumptions. *Annales de l'I.H.P.- Probabilités & Statistiques,* vol. 43, pp. 763-774.

[20] WALTHER, G. (1999). On a generalization of Blaschke's rolling theorem and the smoothing of surfaces. *Math. Methods Appl. Sci.* **22** 301-316. MR1671447(99M:46174).

[21] WALTHER, G. (1997). Granulometric smoothing. *Ann. Statist.* **25** 2273-2299. MR1604445 (99B:6270).