

UNIVERSIDAD NACIONAL DEL LITORAL
FACULTAD DE INGENIERÍA QUÍMICA

TESIS A SER PRESENTADA COMO PARTE DE LOS REQUISITOS DE LA UNIVERSIDAD NACIONAL
DEL LITORAL PARA LA OBTENCIÓN DEL GRADO ACADÉMICO DE

Doctor en Matemática

EN EL CAMPO DE: **Estadística**

TÍTULO DE LA TESIS:

Densidades, regresión y clasificación para datos funcionales.

INSTITUCIÓN DONDE SE REALIZÓ:

Departamento de Matemática – FIQ – UNL
Instituto de Matemática Aplicada del Litoral (IMAL)

AUTOR:

Pamela Nerina Llop

DIRECTORES DE TESIS:

Dr. Ricardo Fraiman y Dra. Liliana Forzani

DEFENDIDA ANTE EL JURADO COMPUESTO POR:

Pablo Ferrari

Pedro Morín

Víctor Yohai

AÑO DE PRESENTACIÓN: 2011

A mis padres y hermanos.

AGRADECIMIENTOS

En este pequeño apartado quiero expresar mi más sincero agradecimiento a todas y cada una de las personas quienes con su apoyo y su ayuda desinteresada hicieron posible la realización de esta tesis.

A mis directores Ricardo y Lili, por su enseñanza y apoyo constante, por ser mi guía y permitirme aprender de sus conocimientos. Gracias por su tiempo compartido de manera generosa y por toda la confianza y paciencia que han tenido durante estos años.

A mis padres y hermanos porque a pesar de los duros momentos que hemos pasado durante los primeros años de la tesis siempre, absolutamente siempre me brindaron su apoyo y jamás permitieron que abandonara este camino. Gracias mamá, papá, Ale, Romi y Jose por toda toda la confianza que han depositado en mi.

A Romi, por enseñarme a no bajar los brazos y por darme la fuerza necesaria para seguir adelante.

A Diego, por infinitas cosas pero por sobre todo por haber sido mi cable a tierra durante estos cinco años, por permitirme tantas ausencias y por estar siempre ahí, esperándome. Gracias gordo por apoyarme en este camino que elegí y por sobre todas las cosas, gracias por brindarme tu amor verdadero.

A mis amigas María, Marilina y Olga y a las de toda la vida Eli y Luci por los buenos momentos.

Al CONICET por haber apoyado económicamente la realización de esta tesis y a Rubén, por su co-dirección en los primeros años. Gracias a todo el IMAL por tanta calidez no sólo profesional sino personal y a Marce, Coca, Ale y Betty, por su buena voluntad y predisposición ante cualquiera de mis pedidos.

A todo los profesores quienes durante estos diez años de carrera matemática me han enseñado todo lo que se, gracias por compartir toda su sabiduría conningo.

A mis compañeros del IMAL y en especial a aquellos con quienes durante estos cinco años he compartido la salita verde, Gise, Mary, Edu, Luis, Kari, Diego, Nacho, Marisa, Adriana, Anibal,

Mauri, Marce, Emi, Marilina gracias por crear siempre ese agradable ambiente de trabajo y estudio, por los buenos mates, las lindas charlas y los valiosos consejos compartidos.

A mis compañeras de la cátedra de Matemática Discreta y en especial a María, por cubrirme cada vez que necesitaba ausentarme de la cátedra para dedicarme a cualquier actividad relacionada con la tesis. A Malva y al Departamento de Sistemas de la UTN por permitirme esas ausencias.

A Cintia, por su amistad y por brindarme su casa cada vez que viajaba a Buenos Aires a trabajar con Ricardo. Gracias también a los compañeros de la Universidad San Andrés, por su cálido trato en cada una de mis visitas a la Universidad.

A los cinco miembros del jurado, por haber aceptado el compromiso de evaluar esta tesis y a aquellos que han tenido la ardua tarea de leerla, gracias por sus valiosos comentarios y correcciones.

Por último pero no menos importante, quiero agradecer a Marilina por su ayuda con el latex, a Pedro y Diego por su ayuda con el Matlab y a todas aquellas personas que me han facilitado desde una palabra justa hasta alguna referencia, archivo, base de datos, programa o cualquier tipo de ayuda necesaria para la realización de esta tesis.

ÍNDICE GENERAL

Agradecimientos	I
Resumen	VII
Introducción	IX
Capítulo 1. El mundo no paramétrico para datos finito dimensionales	1
1.1 Estimadores de la función de densidad	2
1.1.1 Histograma	3
1.1.2 Estimador de núcleo	4
1.1.3 Estimador de k -vecinos más cercanos (k -NN)	5
1.2 Estimadores de la función de regresión	6
1.2.1 Regresograma	8
1.2.2 Estimador de núcleo	9
1.2.3 Estimador de k -vecinos más cercanos (k -NN)	9
1.2.4 Un resultado de consistencia universal	12
1.3 Eligiendo el parámetro de suavización óptimo	12
1.4 Clasificación supervisada (Discriminación)	15
1.5 Datos funcionales	17
Capítulo 2. Estimación de densidades para datos funcionales	19
2.1 Estimando la densidad de un proceso estacionario	21
2.1.1 Preliminares y definición del estimador	21
2.1.2 Hipótesis generales	22
2.1.3 Resultados asintóticos	24
2.2 Estimando la densidad de un proceso no estacionario	44
2.2.1 Extensión del estimador	44
2.2.2 Resultados asintóticos	45
2.3 Una nueva regla de clasificación para datos funcionales	49

2.4	Estudios de simulación	50
2.5	Aplicación a datos reales	54
Capítulo 3.	Estimación de densidades para campos aleatorios	59
3.1	Nociones de dependencia	60
3.2	Estimando la función de densidad de un campo estacionario	63
3.2.1	Preliminares y definición del estimador	63
3.2.2	Hipótesis generales	64
3.2.3	Resultados asintóticos	65
3.3	Estimando de función de densidad de un campo no estacionario	69
3.3.1	Extensión del estimador	69
3.3.2	Resultados asintóticos	69
3.4	Estudios de simulación	70
3.5	Apéndice	72
Capítulo 4.	Regresión para datos funcionales	79
4.1	Preliminares	80
4.2	Extensión del Teorema de Stone	81
4.3	Consistencia del estimador de k -vecinos más cercanos (k -NN)	86
4.4	Consistencia del estimador de núcleos	88
4.5	Apéndice	92
Apéndice A.	Resultados y herramientas útiles	97
A.1	Espacios métricos	97
A.2	Teoría de la medida	98
A.3	Integración	100
A.4	Los espacios L^p	102
A.5	Probabilidad	103
A.6	Convergencias estocásticas	107
A.7	Teoremas límites	108
A.8	Desigualdades exponenciales	109
A.8.1	De Bernstein	109
A.8.2	Para variables aleatorias débilmente dependientes	109
A.8.3	Para variables aleatorias α -mixing geométricas	110
A.9	Procesos estocásticos y campos aleatorios	110

Conclusiones generales	115
Bibliografía.....	117
Índice alfabético	123

RESUMEN

El objetivo de esta tesis es obtener herramientas estadísticas para el análisis de datos infinito dimensionales. En particular, estamos interesados en estudiar resultados de consistencia para estimadores de la función de densidad marginal y de la función de regresión para datos que son curvas o superficies.

En el contexto de datos funcionales, más precisamente para un camino muestral observado continuamente sobre $[0, T_n]$, [Labrador, 2008] definió un estimador no paramétrico de la densidad marginal como una extensión del estimador de k -vecinos más cercanos al contexto continuo. En esta dirección y basados en el trabajo antes citado, el primer problema que abordamos en esta tesis es estimar de manera no paramétrica, la densidad marginal de un proceso estocástico estacionario a partir de una muestra i.i.d. observada sobre un intervalo fijo $[0, T]$ con valores en \mathbb{R} . En este caso, probamos la consistencia del estimador, encontramos velocidades de convergencia puntual de orden $n^{-\alpha}$ para $\alpha < 1/2$, velocidades de convergencia uniforme de orden $\frac{n^{-\alpha}}{\log n}$ para $\alpha < 1/2$ y probamos su distribución normal asintótica de orden $n^{-1/2}$. Luego extendemos la definición del estimador a una clase particular de procesos no estacionarios obteniendo en este caso velocidades de convergencia de orden $n^{-\alpha}$ para $\alpha < 1/4$. Finalmente, como una aplicación de este método de estimación presentamos una nueva regla de clasificación funcional.

El segundo problema que abordamos en esta tesis es la extensión del estimador obtenido en la primera parte a campos aleatorios, es decir, a procesos definidos sobre espacios de mayores dimensiones con valores en \mathbb{R} . Más precisamente, a partir de una colección dependiente de procesos estocásticos definidos en un subconjunto de \mathbb{R}^d con valores en \mathbb{R} , definimos un estimador de la densidad marginal y encontramos velocidades de convergencia de orden $n^{-\alpha}$ para $\alpha < 1/4$. Para campos aleatorios no estacionarios, extendemos la definición dada para el caso estacionario obteniendo nuevamente velocidades de convergencia de orden $n^{-\alpha}$, para $\alpha < 1/4$.

Para estimadores de regresión, [**Stone, 1977**] probó un importante resultado de consistencia universal para datos d -dimensionales y lo usó para probar la consistencia del estimador de k -vecinos más cercanos. Basados en este trabajo, probamos un resultado de consistencia para estimadores de regresión en el contexto infinito dimensional el cual se adapta a una amplia familia de estimadores. Utilizando este resultado, probamos la consistencia de estimadores de regresión ya conocidos como el estimador de k -vecinos más cercanos y el estimador de núcleo en un espacio métrico separable.

INTRODUCCIÓN

Si bien el estudio de herramientas estadísticas para analizar elementos aleatorios infinito dimensionales ha comenzado hace ya varias décadas, podría decirse que la mayoría de los resultados estadísticos existentes en el área son producto de las últimas dos décadas de investigación. Este actual estallido tiene que ver con el gran avance de la tecnología computacional y los nuevos sistemas informáticos que nos permiten recolectar y analizar datos que pueden ser curvas, superficies o imágenes (y por lo tanto viven en dimensión infinita).

Es cierto que en la práctica los datos no vienen dados por curvas infinito dimensionales sino por vectores que, aunque pueden llegar a tener dimensión de cientos de miles, todavía pueden ser tratados como tales. Sin embargo, pensarlos y analizarlos como curvas permite explotar la naturaleza infinito dimensional de dichas observaciones para inferir la estructura de los datos y así, obtener más información y como consecuencia, mejores resultados.

El conjunto de herramientas utilizados para analizar este tipo de datos es conocido como Análisis de Datos Funcionales (FDA, [Ramsay and Dalzell, 1991]). Si bien la expresión datos funcionales puede utilizarse para cualquier elemento infinito dimensional, en esta tesis la utilizaremos para referirnos a curvas mientras que, para referirnos a superficies, utilizaremos la expresión campos aleatorios. Dentro de este contexto, estaremos interesados en realizar inferencias no paramétricas sobre los datos, es decir, realizar inferencias sin hacer suposiciones sobre la distribución subyacente de los mismos ni sobre la estructura de un modelo a priori.

En una primera etapa de esta tesis y basados en el trabajo de [Labrador, 2008] quien define un estimador de la densidad marginal para un proceso estocástico en el contexto dependiente, estudiamos la extensión de este estimador al contexto independiente, es decir, cuando n trayectorias i.i.d. de un proceso estocástico son observadas sobre un intervalo de tiempo finito. Dentro de este marco, consideramos el caso en el cual el proceso es estacionario y el caso

en el que no lo es. Para el primer caso, definimos el estimador y presentamos resultados de consistencia, velocidades de convergencia puntual y uniforme y distribución asintótica. Para el caso no estacionario, definimos la familia de estimadores indexada en el tiempo y presentamos un resultado de consistencia y velocidades de convergencia. Estos resultados fueron publicados por el *Journal of Multivariate Analysis* ([Llop et al., 2011]) y pueden encontrarse online en: <http://dx.doi.org/10.1016/j.jmva.2010.08.002>.

En una segunda etapa de la tesis extendemos el estimador obtenido en la primera parte a campos aleatorios. Aquí un campo aleatorio será una hipersuperficie aleatoria y el estimador será calculado a partir de datos espacio-temporales, es decir, a partir de una hipersuperficie aleatoria que evoluciona en el tiempo. Teniendo en cuenta la evolución temporal de este tipo de datos, definimos una nueva noción de dependencia, llamada (α, ϕ) -débil, la cual probamos que es más débil que las condiciones de dependencia mixing (definida por [Rosenblatt, 1956]) y que es una clase particular de dependencia $(\alpha, \mathcal{G}, \psi)$ -débil (definida por [Doukhan and Louhichi, 1999]). Además mostramos que sucesiones de campos aleatorios (α, ϕ) -débilmente dependientes generan sucesiones de variables aleatorias $(\alpha, \mathcal{G}, \psi)$ -débilmente dependientes y es por ello que desigualdades de tipo Bernstein para este último tipo de dependencia serán utilizadas para obtener velocidades de convergencia tanto para campos estacionarios como para no estacionarios.

Finalmente, basados en el resultado de consistencia universal para estimadores de regresión en \mathbb{R}^d presentado por [Stone, 1977], probamos un resultado de consistencia similar en el contexto infinito dimensional y lo utilizamos para probar la consistencia de estimadores ya conocidos como son el estimador de k -vecinos más cercanos y el estimador de núcleo. Si bien en el contexto infinito dimensional existen resultados individuales de consistencia para estos estimadores, las pruebas presentadas aquí mediante la aplicación del resultado general son considerablemente más sencillas que las ya existentes.

La tesis esta organizada de la siguiente manera. El **Capítulo 1** está dedicado a introducir métodos de estimación de densidades, estimación de regresión y clasificación existentes en el contexto finito dimensional. Al final de este capítulo introducimos brevemente el contexto infinito dimensional, que es en el cual trabajaremos a lo largo de toda la tesis.

En el **Capítulo 2** definimos un estimador de la densidad marginal para una muestra i.i.d. de procesos estocásticos estacionarios y no estacionarios y presentamos resultados de consistencia para dicho estimador. Además, aplicamos nuestros resultados para obtener una nueva regla de clasificación funcional. Finalmente, presentamos ejemplos numéricos que muestran el desempeño del estimador y de la correspondiente regla de clasificación.

En el **Capítulo 3** extendemos los resultados obtenidos en el Capítulo 2 a campos aleatorios y damos resultados de consistencia y velocidades de convergencia para la extensión del estimador. Además, extendemos la definición y los resultados para campos no estacionarios y presentamos ejemplos numéricos que muestran el desempeño del estimador. Aquí las estimaciones son realizadas utilizando sucesiones de campos que satisfacen una nueva noción de dependencia.

El **Capítulo 4** está dedicado al problema de estimar el operador de regresión en modelos donde la covariable es funcional y la respuesta real. En este contexto, extendemos al contexto infinito dimensional un importante resultado general de consistencia para cierta clase de estimadores de regresión y lo utilizamos para probar la consistencia de estimadores ya conocidos.

Finalmente, en el **Apéndice A** introducimos conceptos básicos y resultados clásicos previos necesarios para la comprensión de la tesis.

CAPÍTULO 1

EL MUNDO NO PARAMÉTRICO PARA DATOS FINITO DIMENSIONALES

En contraste con el análisis paramétrico, en el cual se supone que la distribución subyacente sigue un cierto modelo y por lo tanto, el problema consiste en estimar un número finito de parámetros que lo describen, el análisis no paramétrico se basa en la idea de no hacer suposiciones sobre la distribución subyacente de los datos ni sobre la estructura de un modelo a priori. Esto hace que los métodos no paramétricos sean aplicables a un número mucho mayor de problemas que los paramétricos aunque, en general, los primeros producen resultados estadísticamente menos eficientes que los segundos ya que usan la misma información para estimar objetos en dimensión infinita.

Dentro del análisis no paramétrico, el problema de estimar funcionales de la distribución tales como la función de densidad f y la función de regresión es un tema que ha cobrado mucha importancia e interés por parte de los investigadores en los últimos años. Ambos problemas tienen un doble objetivo: por un lado, encontrar un método que estime la cantidad deseada lo mejor posible y por otro lado, obtener un estimador tan simple como sea posible, fácil de entender e implementar numéricamente.

En este capítulo damos una breve introducción a la estimación no paramétrica **en el contexto finito dimensional** (en el sentido que la estimación se realiza en base a una muestra de datos finito dimensionales) presentando conceptos y resultados que nos ayudarán a comprender los próximos capítulos. Además, damos una breve introducción al contexto infinito dimensional que es el contexto en el cual trabajaremos a lo largo de la tesis.

1.1. Estimadores de la función de densidad

Sea $\{X_1, \dots, X_n\}$ una muestra aleatoria distribuida como la variable aleatoria absolutamente continua $X \in \mathbb{R}$ con función de densidad f . Supongamos que f es desconocida y queremos estimarla de manera no paramétrica. Dado que X es absolutamente continua podemos escribir

$$F(x) = \int_{-\infty}^x f(u) du,$$

por lo que, para estimar f podríamos pensar en estimar F y luego derivar. Un estimador natural de la función de distribución $F(x)$ es la función de distribución muestral o empírica

$$F_n(x) = \frac{\#\{X_i : X_i \leq x\}}{n}$$

la cual, como es bien sabido, no es una función diferenciable en todo punto por lo que debemos pensar en otra alternativa para estimar la densidad f . Como una opción, en lugar de considerar directamente la derivada de F_n , es decir, en lugar de considerar el límite

$$\lim_{h \rightarrow 0} \frac{F_n(x+h) - F_n(x-h)}{2h}$$

que, como dijimos puede no existir, consideremos el cociente incremental

$$\frac{F_n(x+h) - F_n(x-h)}{2h}$$

como un posible estimador de f al que llamaremos \hat{f}_n con $h = h_n$ un número positivo convenientemente elegido al que llamaremos **ventana o banda**. Este parámetro juega un papel muy importante en la construcción del estimador y discutiremos su elección unas secciones más adelante. Por ahora, lo único que pedimos es que $\lim_{n \rightarrow \infty} h_n = 0$. Luego,

$$\begin{aligned} \hat{f}_n(x) &= \frac{\#\{X_i : X_i \leq x + h_n\} - \#\{X_i : X_i \leq x - h_n\}}{2nh_n} \\ &= \frac{\#\{X_i : x - h_n < X_i \leq x + h_n\}}{2nh_n} \\ &= \frac{\#\{X_i : X_i \in (x - h_n, x + h_n]\}}{2nh_n} \\ &= \frac{1}{2nh_n} \sum_{i=1}^n \mathbb{I}_{(x-h_n, x+h_n]}(X_i) \\ &= \frac{1}{2nh_n} \sum_{i=1}^n \mathbb{I}_{(-1,1]} \left(\frac{x - X_i}{h_n} \right) \\ &= \frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} K \left(\frac{x - X_i}{h_n} \right), \end{aligned} \tag{1.1}$$

con $K(x) \doteq \frac{1}{2}\mathbb{I}_{(-1,1]}(x)$. Por lo tanto, hemos escrito a $\widehat{f}_n(x)$ como un promedio pesado de una función K de los datos muestrales. Si en lugar de usar la función K como en (1.1) elegimos cualquier otra función medible de integral 1, tendremos diferentes estimadores de f ,

$$\widehat{f}_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right). \quad (1.2)$$

1.1.1. Histograma

El ejemplo más simple y antiguo de estimación no paramétrica de la densidad es el **histograma**. Para la representación y exploración de datos, el histograma es muy útil a la hora de obtener información acerca de la forma que tiene la distribución subyacente en los datos. Por ejemplo, es una manera rápida de saber si la distribución es sesgada o no, si es unimodal, multimodal, etc.

Para dar una definición formal del histograma supongamos que particionamos el subconjunto $[-M, M] \subset \mathbb{R}$ en N subintervalos B_j disjuntos de ancho fijo h_n . Denotamos por n_j a la cantidad de observaciones X_i que caen en B_j , $i = 1, \dots, n$. Luego,

$$n_j = \sum_{i=1}^n \mathbb{I}_{B_j}(X_i) \quad \text{y} \quad n = \sum_{j=1}^N n_j.$$

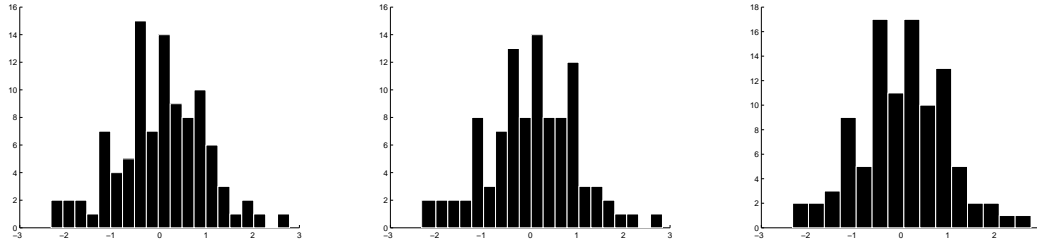
Para $x \in B_j$ definimos el histograma como

$$\widehat{f}_n(x) = \frac{n_j}{nh_n} = \frac{1}{nh_n} \sum_{i=1}^n \mathbb{I}_{B_j}(X_i),$$

la cual es una expresión similar a la dada en (1.2) para $K(x) = \mathbb{I}_{B_j}(x)$. Para $x \in \mathbb{R}$ el histograma se define como

$$\widehat{f}_n(x) = \frac{1}{nh_n} \sum_{j=1}^N \sum_{i=1}^n \mathbb{I}_{B_j}(X_i) \mathbb{I}_{B_j}(x). \quad (1.3)$$

Aunque el histograma es fácil de implementar, como estimador de densidad presenta muchas desventajas ya que no es continuo, su forma es extremadamente sensible a la elección del origen y al ancho de la banda, se necesitan muchos datos muestrales si se quiere identificar la forma de la distribución subyacente, etc. Es posible observar este comportamiento en la Figura 1 donde hemos graficado tres histogramas correspondientes a la misma muestra aleatoria $\{X_1, \dots, X_{100}\}$, con $X_i \sim \mathcal{N}(0, 1)$ para tres situaciones diferentes: la Figura 1 (a) corresponde al histograma con 20 bandas de ancho 0.2558, en la Figura 1 (b) hemos graficado el mismo histograma, con la misma cantidad de bandas pero con el punto inicial x_0 corrido en 0.05 con respecto al anterior y finalmente, en la Figura 1 (c) hemos graficado el histograma con 15 bandas de ancho 0.3411.



(a) 20 bandas y $x_0 = -2.3289$ (b) 20 bandas y $x_0 = -2.2789$ (c) 15 bandas y $x_0 = -2.3289$

FIGURA 1. Histogramas correspondientes a $\{X_1, \dots, X_{100}\}$, con $X_i \sim \mathcal{N}(0, 1)$.

1.1.2. Estimador de núcleo

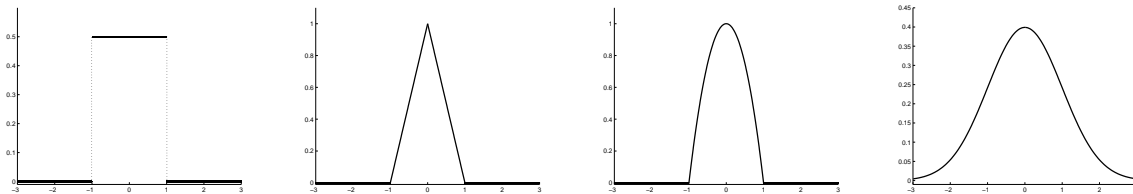
Los **estimadores de núcleo** o estimadores de **Akaike-Rosenblatt-Parzen** son tal vez los más utilizados en la estadística no paramétrica ya que evitan la mayoría de los problemas que presenta el histograma. Las ventajas principales de este estimador ante el histograma son su continuidad y suavidad.

Para definir un estimador de núcleo, en (1.2) tomamos como $K(x)$ a cualquier **función núcleo**, es decir, cualquier función $K : \mathbb{R} \rightarrow \mathbb{R}_0^+$ tal que $\int K(u) du = 1$ y $\int uK(u) du = 0$ y definimos

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

donde h_n es la banda (o ventana). En general, no existen restricciones sobre la elección de la función núcleo aunque los más utilizados son los simétricos, como por ejemplo:

- *Núcleo uniforme*: $K(x) = \frac{1}{2}\mathbb{I}_{[-1,1]}(x)$. (presentado en (1.1))
- *Núcleo triangular*: $K(x) = (x + 1)\mathbb{I}_{[-1,0]}(x) + (1 - x)\mathbb{I}_{[0,1]}(x)$.
- *Núcleo cuadrático* (de Epanechnikov): $K(x) = \frac{3}{4}(1 - x^2)\mathbb{I}_{[-1,1]}(x)$.
- *Núcleo gaussiano*: $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}x^2\right\}$.



(a) Núcleo uniforme (b) Núcleo triangular (c) Núcleo cuadrático (d) Núcleo gaussiano

FIGURA 2. Diferentes funciones núcleo.

La Figura 3 muestra cómo varía el estimador con la elección del núcleo.

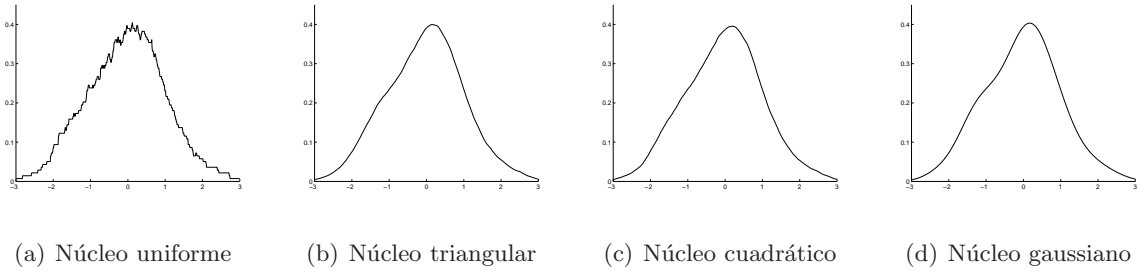


FIGURA 3. Estimadores de la densidad de $\{X_1, \dots, X_{100}\}$, con $X_i \sim \mathcal{N}(0, 1)$, para diferentes elecciones del núcleo.

Esta definición se puede extender a \mathbb{R}^d donde, para cada x el estimador queda definido como

$$\hat{f}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n \tilde{K}\left(\frac{x - X_i}{h_n}\right),$$

donde $\tilde{K} : \mathbb{R}^d \rightarrow \mathbb{R}_0^+$ es la función núcleo y h_n es el radio de la bola d -dimensional. Observemos que \tilde{K} puede ser definido por ejemplo, combinando una función núcleo en \mathbb{R} y una norma $\|\cdot\|$ en \mathbb{R}^d como sigue

$$\tilde{K}(x) = K(\|x\|), \text{ para todo } x \in \mathbb{R}^d.$$

(ver [Härdle and Müller, 1997] para una lectura más rigurosa sobre este tema).

Los primeros trabajos sobre estimación de densidades en \mathbb{R} con núcleos aparecieron recién en los años 50, con los trabajos de [Akaike, 1954] y [Rosenblatt, 1956], [Parzen, 1962] y [Cacoullos, 1966]. El caso multivariado fue estudiado originalmente por [Cacoullos, 1966] y [Epanechnikov, 1969]. Entre las referencias más recientes sobre estimación de densidades con núcleo podemos destacar a [Marron and Nolan, 1988], [Vieu, 1993], [Giné and Guillou, 2002], [Dabo-Niang, 2004], [Dony and Einmahl, 2006] y [Mynbaev and Martins-Filho, 2010].

1.1.3. Estimador de k-vecinos más cercanos (k -NN)

Recordemos que para construir el histograma y el estimador de núcleo en un punto x , fijábamos el ancho de banda y teníamos en cuenta las observaciones que caían dentro de esa banda elegida previamente. Pensemos en hacer este proceso de manera inversa, es decir, en vez de fijar el ancho de banda de antemano y tener en cuenta la cantidad de observaciones cercanas a x , fijemos la cantidad de observaciones que queremos tener en cuenta a la hora de calcular el estimador en x y luego calculemos el ancho de banda. Más precisamente, supongamos que para estimar la densidad en x queremos tomar las k observaciones más próximas a x . Dado que k

dependerá del tamaño de la muestra lo llamaremos k_n . El siguiente paso entonces es encontrar el ancho de banda h_n el cual no sólo dependerá de la muestra sino también del punto x por lo que lo denotaremos $h_n(x)$. Por lo tanto para cada x , la cantidad de observaciones que caen en $[x - h_n(x), x + h_n(x)]$ es k_n por lo que en (1.1) tendremos que

$$\hat{f}_n(x) = \frac{F_n(x + h_n(x)) - F_n(x - h_n(x))}{2h_n(x)} = \frac{k_n}{2nh_n(x)}.$$

Este estimador de la densidad se conoce como estimador de **k -vecinos más cercanos** o **k -NN**. Un aspecto muy importante a tener cuenta en este tipo de estimadores es que la banda ya no es fija sino aleatoria. Además, ya no es el ancho de banda quien juega el papel de parámetro de suavización sino la cantidad de vecinos que tomemos para realizar la estimación.

La extensión de este problema a más dimensiones es casi inmediata, en este caso k_n es la cantidad de elementos en \mathbb{R}^d que viven en la bola $\mathcal{B}(x, h_n(x))$ de centro x y radio $h_n(x)$ y

$$\hat{f}_n(x) = \frac{k_n}{n\lambda_d(\mathcal{B}(x, h_n(x)))},$$

donde $\lambda_d(\mathcal{B}(x, h_n(x)))$ denota la medida de Lebesgue en \mathbb{R}^d de la bola $\mathcal{B}(x, h_n(x))$.

El problema de estimar la densidad usando k -NN ha sido considerado por varios autores en diferentes contextos. En el caso independiente, [Moore and Henrichon, 1969] demostraron la convergencia uniforme en probabilidad en \mathbb{R} . Para vectores aleatorios p -dimensionales [Loftsgaarden and Quesenberry, 1965] fueron los primeros en probar su consistencia, [Wagner, 1973] probó la convergencia puntual completa, [Moore and Yackel, 1977] probaron la convergencia puntual casi segura de un estimador de k -vecinos más cercanos más general y [Boente and Fraiman, 1988] probaron la consistencia uniforme en el caso dependiente.

1.2. Estimadores de la función de regresión

Sean $\{(X_i, Y_i)\}_{i=1}^n$ elementos i.i.d. con la misma distribución que la del par (X, Y) . Supongamos que los datos verifican el modelo

$$Y = \eta(X) + e,$$

donde $\eta : \mathbb{R} \rightarrow \mathbb{R}$ es la función de regresión desconocida y e es el error con $\mathbb{E}(e|X) = 0$. De esta última igualdad se sigue que

$$\eta(X) = \mathbb{E}_Y(Y|X). \tag{1.4}$$

Supongamos que queremos estimar de manera no paramétrica la función de regresión η . Para ello, sean $f_{X,Y}(x,y)$ la densidad conjunta de (X,Y) y $f_{Y|X}(y|x)$ la densidad de $Y|X$ (desconocidas). Por definición de esperanza condicional resulta

$$\mathbb{E}_Y(Y|X=x) = \int y f_{Y|X}(y|x) dy = \int y \frac{f_{X,Y}(x,y)}{f_X(x)} dy,$$

por lo tanto, podríamos definir

$$\hat{\eta}(x) = \hat{\mathbb{E}}_Y(Y|X) = \int y \frac{\hat{f}_{X,Y}(x,y)}{\hat{f}_X(x)} dy. \quad (1.5)$$

Recordemos de (1.1) que un estimador de la densidad $f_X(x)$ es dado por

$$\hat{f}_X(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right),$$

con $K(x) \doteq \frac{1}{2}\mathbb{I}_{(-1,1]}(x)$. De la misma manera podemos deducir que un estimador para la densidad conjunta $f_{X,Y}(x,y)$ es dado por

$$\hat{f}_{X,Y}(x,y) = \frac{1}{nh_n^2} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) K\left(\frac{y-Y_i}{h_n}\right),$$

con K como antes. Reemplazando estos estimadores en (1.5) resulta

$$\begin{aligned} \hat{\eta}(x) &= \int y \frac{\frac{1}{nh_n^2} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) K\left(\frac{y-Y_i}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} dy \\ &= \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \frac{1}{h_n} \int y K\left(\frac{y-Y_i}{h_n}\right) dy}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}. \end{aligned} \quad (1.6)$$

Ahora, haciendo el cambio de variable $z = \frac{y-Y_i}{h_n}$ y teniendo en cuenta que

$$\int K(z) dz = \frac{1}{2} \int_{-1}^1 dz = 1 \quad \text{y} \quad \int zK(z) dz = \frac{1}{2} \int_{-1}^1 z dz = 0$$

tenemos que

$$\begin{aligned} \frac{1}{h_n} \int y K\left(\frac{y-Y_i}{h_n}\right) dy &= \int (h_n z + Y_i) K(z) dz \\ &= h_n \int zK(z) dz + Y_i \int K(z) dz \\ &= Y_i. \end{aligned}$$

Finalmente, reemplazando esta igualdad en (1.6) tenemos que

$$\hat{\eta}_n(x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) Y_i}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \doteq \sum_{i=1}^n W_{ni}(x) Y_i, \quad (1.7)$$

con $K(x) \doteq \frac{1}{2}\mathbb{I}_{(-1,1]}(x)$ y $W_{ni}(x) \doteq \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} = \frac{\frac{1}{2}\mathbb{I}_{(-1,1]}(x)}{\sum_{i=1}^n \frac{1}{2}\mathbb{I}_{(-1,1]}(x)}$ denominados **pesos**. Si en lugar de usar la función K como en (1.7) elegimos cualquier otra función medible de integral 1, tendremos diferentes estimadores de η ,

$$\widehat{\eta}_n(x) = \sum_{i=1}^n W_{ni}(x)Y_i, \quad (1.8)$$

con $W_{ni}(x) \doteq \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}$. Este tipo de estimadores generales es el que estudiaremos en lo que resta de esta sección.

1.2.1. Regresograma

El estimador de regresión no paramétrico más simple es el **regresograma**. Éste, es una función escalonada que asigna a cada intervalo el valor medio de los valores que caen en él y su nombre se debe a su analogía con el histograma.

El regresograma es muy útil a la hora de obtener información acerca de la forma que tiene la función de regresión y su definición es la siguiente. Supongamos que particionamos el subconjunto $[-M, M] \subset \mathbb{R}$ en N intervalos B_j de ancho fijo h_n . Denotamos por n_j a la cantidad de observaciones X_i que caen en B_j , $i = 1, \dots, n$ y por \bar{Y}_j al promedio de las respuestas correspondientes a dichas observaciones. Más precisamente,

$$n_j = \sum_{i=1}^n \mathbb{I}_{B_j}(X_i), \quad n = \sum_{j=1}^N n_j, \quad \bar{Y}_j = \frac{1}{n_j} \sum_{i=1}^N Y_i \mathbb{I}_{B_j}(X_i).$$

Para $x \in B_j$ definimos el regresograma como

$$\widehat{\eta}_n(x) = \bar{Y}_j = \frac{1}{n_j} \sum_{i=1}^n Y_i \mathbb{I}_{B_j}(X_i) = \frac{\sum_{i=1}^n Y_i \mathbb{I}_{B_j}(X_i)}{\sum_{i=1}^n \mathbb{I}_{B_j}(X_i)} \quad (1.9)$$

la cual es una expresión como la dada en (1.8) con $K(x) = \mathbb{I}_{B_j}(x)$ y $W_{ni}(x) = \frac{\mathbb{I}_{B_j}(X_i)}{\sum_{i=1}^n \mathbb{I}_{B_j}(X_i)}$. Para $x \in \mathbb{R}$ el regresograma se define como

$$\widehat{\eta}_n(x) = \sum_{j=1}^N \frac{1}{n_j} \left(\sum_{i=1}^n Y_i \mathbb{I}_{B_j}(X_i) \right) \mathbb{I}_{B_j}(x). \quad (1.10)$$

Este estimador presenta las mismas desventajas que el histograma, no es continuo, su forma es extremadamente sensible a la elección del ancho de banda, se necesitan muchos datos muestrales si se quiere identificar la función de regresión subyacente, etc. Es posible observar este comportamiento en la Figura 4 donde hemos graficado los regresogramas correspondientes al diámetro versus la altura de 115 árboles del bosque experimental de la Universidad de Idaho medidos a 137 centímetros del suelo (ver [Weisberg, 2005] pág 112).

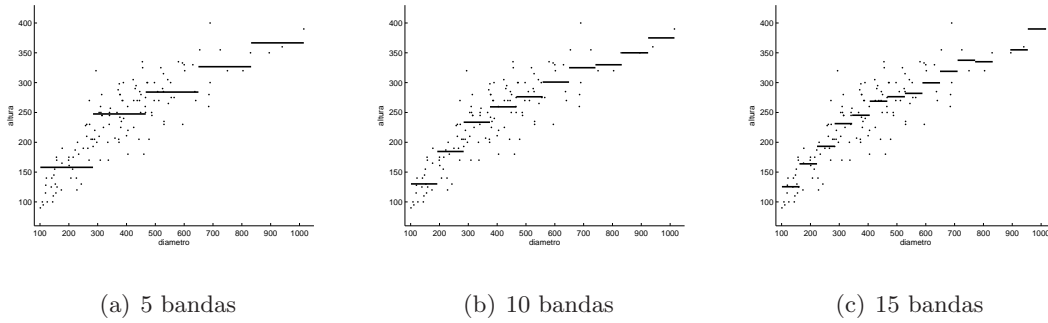


FIGURA 4. Regresogramas correspondientes a una muestra del diámetro de 115 árboles.

1.2.2. Estimador de núcleo

Este tipo de estimador es uno de los más utilizados en regresión y es conocido como el estimador de regresión de **Nadaraya-Watson** debido a sus creadores [**Nadaraya, 1964**, **Nadaraya, 1965**] y [**Watson, 1964**].

Para definir un estimador de núcleo en (1.8) tomamos como $K(x)$ a cualquier **función núcleo**, es decir, cualquier función $K : \mathbb{R} \rightarrow \mathbb{R}_0^+$ tal que $\int K(u) du = 1$ y $\int uK(u) du = 0$ y definimos

$$\hat{\eta}_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)},$$

donde h_n es la banda (o ventana). En general, no existen restricciones sobre la elección del núcleo aunque los más utilizados son los que fueron dados en la Sección 1.1.2.

La consistencia universal débil de este estimador, bajo ciertas condiciones sobre K y h_n , ha sido demostrada entre otros por [**Devroye, 1978**], [**Devroye and Wagner, 1980**] y [**Spiegelman and Sacks, 1980**].

Para extender este estimador a \mathbb{R}^d , es necesario considerar un núcleo multivariado $\tilde{K} : \mathbb{R}^d \rightarrow \mathbb{R}$, que puede ser definido por ejemplo (ver [**Härdle and Müller, 1997**]) combinando una función núcleo en \mathbb{R} y una norma $\|\cdot\|$ en \mathbb{R}^d como sigue

$$\tilde{K}(x) = K(\|x\|), \text{ para todo } x \in \mathbb{R}^d.$$

1.2.3. Estimador de k -vecinos más cercanos (k -NN)

La primera noción de vecinos más cercanos fue introducida por [**Fix and Hodges, 1951**] para el problema de clasificación (en la Sección 1.4 veremos este tema más en detalle). La regla de clasificación de k -vecinos más cercanos sirve para predecir la naturaleza de una observación contando la mayoría de votos entre las k observaciones que están más próximas al punto.

En este contexto, [Cover and Hart, 1967] probaron la consistencia para $k = 1$ en espacios métricos separables cuando X es absolutamente continua y η es continua en casi todo punto. [Devroye, 1981b] probó este resultado en \mathbb{R}^d , sin hacer suposiciones sobre la distribución de los datos. Luego este mismo autor ([Devroye, 1981a]) extendió el resultado anterior al contexto de regresión y probó la consistencia en media p del estimador sin hacer suposiciones sobre la distribución de X pero suponiendo acotación de la respuesta Y y convergencia a cero de la media del error absoluto condicionada a la muestra. En ese entonces, [Stone, 1977] ya había probado un importante resultado de consistencia universal para estimadores de regresión en \mathbb{R}^d (Teorema 1.4) y en particular, lo había usado para probar la consistencia del estimador k -NN (Teorema 1.2).

Para definir los vecinos más cercanos a x en \mathbb{R} con la métrica usual, seguimos los siguientes pasos:

- calculamos las distancias de cada punto observacional a x ,

$$d_i = |x - X_i|, \quad i = 1, \dots, n;$$

- calculamos los estadísticos de orden para esas distancias,

$$d_{(1)} \leq d_{(2)} \leq \dots \leq d_{(k)} \leq \dots \leq d_{(n)};$$

- las observaciones correspondientes a esos estadísticos de orden serán los **vecinos más cercanos** a x ,

$$X_{(1)}, X_{(2)}, \dots, X_{(k)}, \dots, X_{(n)}.$$

Así, $X_{(1)}$ es el 1-vecino más cercano de x , $X_{(2)}$ es el 2-vecino más cercano de x y $X_{(k)}$ es el k -vecino más cercano de x . Si hay empates, es decir si para $i \neq j$, $d_{(i)} = d_{(j)}$, elijiremos al azar entre $X_{(i)}$ y $X_{(j)}$ con una distribución uniforme discreta. Sin embargo, por simplicidad de notación, supondremos que no hay empates, es decir, supondremos que para $i \neq j$, $d_{(i)} \neq d_{(j)}$ con probabilidad uno.

Al igual que ocurría para el estimador de la función de densidad, k dependerá del tamaño de la muestra por lo que lo llamaremos k_n aunque seguiremos utilizando la notación k -NN para referirnos a este tipo de estimadores en lugar de k_n -NN. Definimos el estimador de **k -vecinos más cercanos** como

$$\hat{\eta}_n(x) = \sum_{i=1}^{k_n} W_{ni}(x) Y_i,$$

donde los pesos son de la forma

$$W_{ni}(x) = \begin{cases} \frac{1}{k_n} & \text{si } |x - X_i| \leq |x - X_{(k_n)}|, \\ 0 & \text{en otro caso.} \end{cases}$$

Observemos que si definimos $h_n(x) \doteq |x - X_{(k_n)}|$ podemos escribir

$$k_n = \sum_{i=1}^n \mathbb{I}_{[x-h_n(x), x+h_n(x)]}(X_i) = \sum_{i=1}^n \mathbb{I}_{[-1,1]} \left(\frac{x - X_i}{h_n(x)} \right)$$

y

$$W_{ni}(x) = \frac{1}{k_n} \mathbb{I}_{[x-h_n(x), x+h_n(x)]}(X_i) = \frac{\mathbb{I}_{[-1,1]} \left(\frac{x - X_i}{h_n(x)} \right)}{\sum_{i=1}^n \mathbb{I}_{[-1,1]} \left(\frac{x - X_i}{h_n(x)} \right)},$$

por lo que el estimador de k -vecinos más cercanos queda expresado como

$$\hat{\eta}_n(X) = \frac{\sum_{i=1}^n \mathbb{I}_{[-1,1]} \left(\frac{x - X_i}{h_n(x)} \right) Y_i}{\sum_{i=1}^n \mathbb{I}_{[-1,1]} \left(\frac{x - X_i}{h_n(x)} \right)}.$$

Notemos que este es un estimador de la forma (1.8) con la diferencia principal que aquí el ancho de banda no es fijo, sino que varía de acuerdo al punto donde se quiera realizar la estimación.

Observación 1.1. Es importante notar que si bien hemos definido los vecinos más cercanos y el estimador de este tipo en \mathbb{R} , la extensión a \mathbb{R}^d es inmediata reemplazando $|\cdot|$ por $\|\cdot\|$.

El siguiente resultado es uno de los resultados de consistencia más importantes para el estimador k -NN y fue probado por [Stone, 1977]. Para una demostración en un contexto más simple de este resultado nos referimos a [Devroye et al., 1996], Teorema 6.4, pág 101.

Teorema 1.2. Si $k_n \rightarrow \infty$ y $k_n/n \rightarrow 0$ entonces,

$$\lim_{n \rightarrow \infty} \mathbb{E} ((\eta(X) - \hat{\eta}_n(X))^2) = 0.$$

Un punto interesante de observar dentro de la demostración de este resultado es que el k -vecino más cercano de X converge a X con probabilidad uno. Más precisamente,

Lema 1.3. Si $x \in \text{sop}(\mu)$ y $k_n/n \rightarrow 0$, entonces $|x - X_{(k_n)}| \xrightarrow{a.s.} 0$. Si X es independiente de los datos y tiene medida de probabilidad μ , entonces $|X - X_{(k_n)}| \xrightarrow{a.s.} 0$ siempre que $k_n/n \rightarrow 0$.

Este resultado fue probado primeramente por [Cover and Hart, 1967] quienes probaron que, en espacios métricos separables, el primer vecino más cercano de X converge a él con probabilidad uno. Luego, [Stone, 1977] Corolario 3, pág 600 probó este resultado para k -vecinos más cercanos en \mathbb{R}^d . Nosotros nos referimos a [Devroye et al., 1996] Lema 5.1, pág

63 para la demostración de este lema que, si bien está dada en \mathbb{R}^d con su norma usual, puede ser extendida a cualquier espacio métrico separable.

1.2.4. Un resultado de consistencia universal

Para estimadores de la forma (1.8) [Stone, 1977] Corolario 1, pág 598 probó el siguiente resultado de consistencia universal.

Teorema 1.4. *Sea $\{(X_i, Y_i)\}_{i=1}^n$ una muestra aleatoria independiente e idénticamente distribuida como el par $(X, Y) \in \mathbb{R}^d \times \mathbb{R}$ que satisface el modelo (1.4) y supongamos que los pesos del estimador general (1.8) satisfacen las siguientes tres condiciones:*

(i) *existe $c > 0$ tal que, para toda f , $\mathbb{E}(f(X)) < \infty$,*

$$\mathbb{E} \left(\sum_{i=1}^n W_{ni}(X) f(X_i) \right) \leq c \mathbb{E}(f(X));$$

(ii) *para todo $a > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{i=1}^n W_{ni}(X) \mathbb{I}_{\{|x - X_i| > a\}} \right) = 0;$$

(iii)

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\max_{1 \leq i \leq n} W_{ni}(X) \right) = 0.$$

Entonces,

$$\lim_{n \rightarrow \infty} \mathbb{E} ((\eta(X) - \hat{\eta}(X))^2) = 0.$$

Para una demostración en un contexto más simple de este resultado nos referimos a [Devroye et al., 1996] Teorema 6.3, pág 98.

1.3. Eligiendo el parámetro de suavización óptimo

Ya hemos visto en las secciones anteriores que cuando deseamos estimar una función de densidad o de regresión de manera no paramétrica, ya sea utilizando un histograma/regresograma, un estimador de núcleo o un estimador de k -vecinos más cercanos, un problema básico es la elección de la cantidad de suavización que se requiere. En el caso de los histogramas/regresogramas y estimadores de núcleo, esto está controlado por el ancho de banda h_n y en el caso de estimadores de k -NN esto está controlado por la cantidad de vecinos más cercanos que se tomen para realizar la estimación. Estos parámetros son llamados **parámetro de suavización** y su elección es fundamental para obtener buenas estimaciones.

Para entender cuán importante es elegir un buen parámetro de suavización, veamos por ejemplo el problema de estimar la densidad mediante el histograma. Como lo hemos dicho anteriormente, el ancho de banda (parámetro de suavización) que se tome es muy importante ya que la forma del histograma es muy sensible a la cantidad de bandas consideradas (y por lo tanto a su ancho).

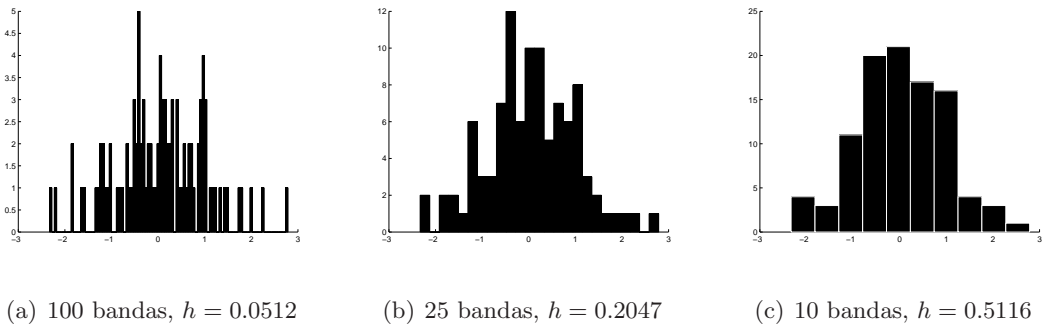


FIGURA 5. Histogramas correspondientes a $\{X_1, \dots, X_{100}\}$, con $X_i \sim \mathcal{N}(0, 1)$.

La Figura 5 muestra tres histogramas basados en el mismo conjunto de datos realizados con tres anchos de banda diferentes. Observemos como una banda ancha produce un estimador sobresuavizado y una banda estrecha produce un estimador poco suavizado. Lamentablemente este problema no se evita usando estimadores de núcleo como se muestra en la Figura 6.

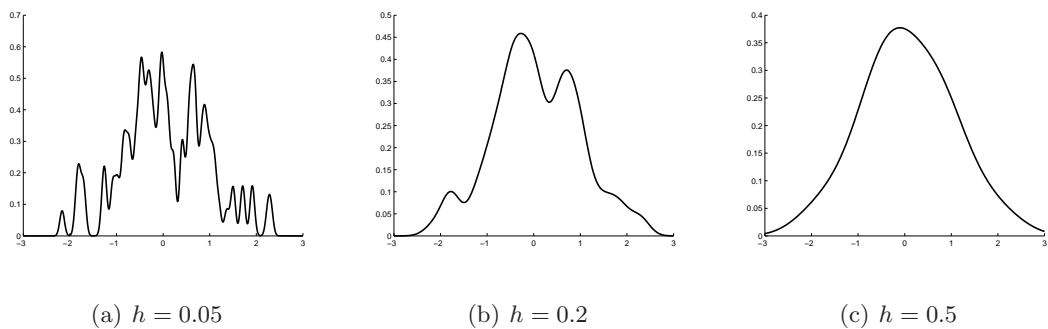


FIGURA 6. Estimadores de núcleo correspondientes a $\{X_1, \dots, X_{100}\}$, con $X_i \sim \mathcal{N}(0, 1)$.

Este fenómeno también puede observarse al tratar de estimar la función de regresión usando regresogramas (ver Figura 4) o estimadores de Nadaraya-Watson (ver Figura 7). En este caso, un ancho de banda pequeño produce esencialmente una interpolación de los datos mientras que un ancho de banda mayor produce un estimador sobresuavizado.

Para poder elegir el mejor parámetro de suavización existen métodos llamados **métodos automáticos** de selección de parámetros. Tal vez uno de los más utilizados es el método de **validación cruzada por mínimos cuadrados**, introducido por [Rudemo, 1982] y

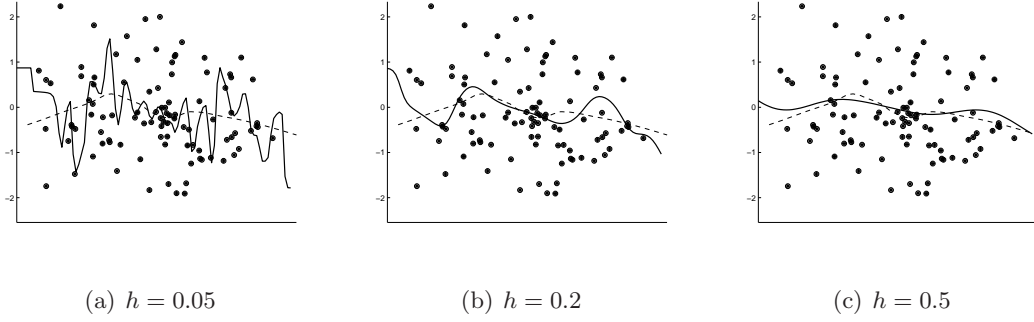


FIGURA 7. Estimadores de Nadaraya-Watson correspondientes a $\{X_1, \dots, X_{100}\}$, con $X_i \sim \mathcal{N}(0, 1)$.

[Bowman, 1984]. Este método se basa en encontrar el parámetro óptimo minimizando el **error cuadrático integrado** $ECI(h_n) = \int (\hat{g}_n(x) - g(x))^2 dx$. Más precisamente, supongamos que queremos estimar $g : \mathcal{F} \rightarrow \mathbb{R}$ con parámetro óptimo h_n^* entonces, éste será el que mininice el error cuadrático integrado

$$\begin{aligned} ECI(h_n) &= \int (\hat{g}_n(x) - g(x))^2 dx \\ &= \int \hat{g}_n^2(x) dx - 2 \int \hat{g}_n(x)g(x) dx + \int g^2(x) dx. \end{aligned} \quad (1.11)$$

Observemos que $\int g^2(x) dx$ no depende de h_n y puede ser ignorado en el problema de minimización. Luego, en lugar de minimizar $ECI(h_n)$ minimizaremos

$$LS(h_n) = \int \hat{g}_n^2(x) dx - 2 \int \hat{g}_n(x)g(x) dx. \quad (1.12)$$

Dado que el primer término puede ser calculado de los datos muestrales, sólo resta calcular el segundo término. Para ello observemos que

$$\int \hat{g}_n(x)g(x) dx = \mathbb{E}_g(\hat{g}_n),$$

pero como g es desconocida, no podemos calcular esta esperanza por lo que la estimamos mediante

$$\widehat{\mathbb{E}}_g(\hat{g}_n) = \frac{1}{n} \sum_{i=1}^n \hat{g}_{n,-i}(X_i),$$

donde $\hat{g}_{n,-i}(x)$ es el estimador denominado **leave-one-out** de g el cual es calculado con todos los elementos de la muestra excepto X_i . Luego reemplazamos (1.12) por

$$\widehat{LS}_n(h_n) = \int \hat{g}_n^2(x) dx - \frac{2}{n} \sum_{i=1}^n \hat{g}_{n,-i}(X_i),$$

y por lo tanto, el **ancho de banda óptimo** será aquel que minimice la expresión anterior, esto es,

$$h_n^* = \min_{h_n} \widehat{L\hat{S}}_n(h_n).$$

Otros métodos de selección de parámetros son aquellos que consisten en minimizar por ejemplo, el **error cuadrático medio** $ECM(x) = \mathbb{E} \left((\widehat{g}_n(x) - g(x))^2 \right)$ o el **error cuadrático medio integrado** $ECMI(h_n) = \int ECM(x) dx = \int \mathbb{E} \left((\widehat{g}_n(x) - g(x))^2 \right) dx$.

Supongamos por ejemplo que en lugar de minimizar el error cuadrático integrado (1.11) queremos minimizar el error cuadrático medio $ECM(x) = \mathbb{E} \left((\widehat{g}_n(x) - g(x))^2 \right)$. Para ello, si $g_n \doteq \mathbb{E}(\widehat{g}_n)$ podemos escribir

$$\begin{aligned} (\widehat{g}_n(x) - g(x))^2 &= ((\widehat{g}_n(x) - g_n(x)) + (g_n(x) - g(x)))^2 \\ &= (\widehat{g}_n(x) - g_n(x))^2 + 2(\widehat{g}_n(x) - g_n(x))(g_n(x) - g(x)) + (g_n(x) - g(x))^2. \end{aligned}$$

Por la definición de g_n , $\mathbb{E}(\widehat{g}_n(x) - g_n(x)) = 0$ por lo que,

$$\begin{aligned} ECM(x) &= \mathbb{E} \left((\widehat{g}_n(x) - g(x))^2 \right) \\ &= \mathbb{E} \left((\widehat{g}_n(x) - g_n(x))^2 \right) + \mathbb{E} \left((g_n(x) - g(x))^2 \right) \\ &= \text{var}(\widehat{g}_n(x)) + \text{sesgo}^2(\widehat{g}_n(x)). \end{aligned}$$

Luego, para que \widehat{g}_n sea **consistente en media cuadrática**, es decir, para que $ECM(x) \rightarrow 0$, es necesario encontrar un equilibrio entre $\text{var}(\widehat{g}_n(x))$ y $\text{sesgo}^2(\widehat{g}_n(x))$. Recordemos de la Sección 1.3 que parámetros muy grandes producían estimadores sobreesuavizados, con poca varianza pero mucho sesgo y que parámetros muy chicos producían estimadores poco suaves, con poco sesgo pero con mucha varianza. Este desafío de obtener un balance entre la varianza y el sesgo de un estimador se conoce como el problema de **compensación entre el sesgo y la varianza**, más conocido como **trade-off** entre el sesgo y la varianza.

1.4. Clasificación supervisada (Discriminación)

Clasificación supervisada o **discriminación** es una técnica utilizada para predecir o conjeturar la naturaleza no conocida Y de una observación X que toma valores en un espacio \mathcal{F} . Y es comunmente llamada **clase** y toma valores en un conjunto finito $\bar{G} = \{1, \dots, G\}$. Para ello, creamos una función $g : \mathcal{F} \rightarrow \bar{G}$ llamada **clasificador**, la cual representa la conjetura de Y dado X . Un clasificador se equivoca si $g(X) \neq Y$, por lo que la **probabilidad de error** para

un clasificador es

$$L(g) = P(g(X) \neq Y).$$

El **mejor clasificador posible** g^* será aquel que minimiza la probabilidad de error $P(g(X) \neq Y)$. Es decir,

$$g^* = \arg \min_{g: \mathcal{F} \rightarrow \bar{G}} P(g(X) \neq Y).$$

(ver [Devroye et al., 1996] Teorema 2.1 pág 10). El problema de encontrar g^* se conoce como **problema de Bayes** y g^* se denomina **clasificador o regla de Bayes**. La mínima probabilidad de error es llamada **error de Bayes** y se denota por $L^* = L(g^*)$.

Un problema equivalente a minimizar la probabilidad de error es el de maximizar la probabilidad a posteriori $P(g(X) = Y)$. Es decir, la **regla de Bayes** g^* también está dada por

$$g^* = \arg \max_{g: \mathcal{F} \rightarrow \bar{G}} P(g(X) = Y).$$

Observemos que g^* depende de la distribución de (X, Y) por lo que para poder calcularla dicha distribución debe ser conocida. Como este no es típicamente el caso, calculamos un clasificador en base a una muestra aleatoria i.i.d. $\{(X_i, Y_i)_{i=1}^n\}$ de (X, Y) ,

$$g_n(\cdot; X_1, Y_1, \dots, X_n, Y_n) : \mathcal{F} \times (\mathcal{F} \times \{1, \dots, m\})^n \rightarrow \{1, \dots, m\},$$

con probabilidad de error dada por la probabilidad condicional

$$L_n(g_n) = P(g_n(\cdot; X_1, Y_1, \dots, X_n, Y_n) \neq Y | X_1, Y_1, \dots, X_n, Y_n).$$

El proceso de construir g_n se llama **aprendizaje supervisado** y una sucesión de clasificadores $\{g_n; n \geq 1\}$ es llamada **regla de clasificación**.

Diremos que una regla es **(débilmente) consistente** si

$$\lim_{n \rightarrow \infty} \mathbb{E}(L_n) = L^*$$

y **(fuertemente) consistente** si

$$L_n \xrightarrow{a.s.} L^*.$$

Si una regla es consistente para todas las distribuciones de (X, Y) se dice que es **universalmente consistente**.

En el caso finito dimensional, existen varias reglas universalmente consistentes. En particular, la regla más popular es la de k -vecinos más cercanos, la cual asigna a un nuevo elemento la clase por mayoría de votos (es decir, la más frecuente) entre sus k -vecinos más cercanos. Esta regla es universalmente consistente siempre que $k_n \rightarrow \infty$ y $k_n/n \rightarrow 0$

(ver [Devroye et al., 1996] Teorema 6.4 pág 101). Los libros de [Devroye et al., 1996], [Duda and Stork, 2000] y [Hastie et al., 2001] ofrecen una amplia cobertura de estos temas para $\mathcal{F} = \mathbb{R}^d$. Sin embargo, las definiciones son esencialmente las mismas para cualquier espacio \mathcal{F} .

1.5. Datos funcionales

En los últimos años, el gran avance en la tecnología computacional (desempeño de los procesadores y la capacidad de almacenamiento) nos permite cada vez más recolectar y analizar datos continuamente en el tiempo. Esta clase de datos provienen de diversas áreas como la la salud (electrocardiogramas, imágenes de resonancias magnéticas), biomecánica (análisis de los movimientos del cuerpo humano), química (curvas de espectrometría), econometría (índices bursátiles), genética (micro-arreglos), geofísica (imágenes satelitales, series de tiempo espaciales), meteorología (mapa de temperaturas, precipitaciones, etc.), geología (mapas geotécnicos), entre otros.

Cuando los datos son grabados densamente sobre un período de tiempo/espacio, estadísticamente son pensados y analizados como curvas/superficies. En este contexto, cada curva/superficie observada corresponde a un individuo diferente en la muestra y cada punto de ella corresponde a una misma variable, medida a lo largo del tiempo.

En la práctica, en general, los datos no provienen de forma continua sino de manera discretizada, es decir, las mediciones están dadas por vectores/matrices finito dimensionales los cuales pueden llegar a tener dimensiones del orden de cientos de miles. Sin embargo, podemos trasladar este problema al contexto de datos funcionales suponiendo que existe una curva/superficie subyacente la cual es observada sólo en instantes discretos de tiempo. Así podemos explotar la naturaleza infinito dimensional de dichas observaciones para inferir la estructura de los datos y obtener más información y como consecuencia, mejores resultados.

El conjunto de herramientas usadas para analizar esta clase de datos es llamado **Análisis de Datos Funcionales** (FDA, [Ramsay and Dalzell, 1991]) y los autores que de alguna manera popularizaron este área dentro de la estadística son Ramsay, Silverman, Ferraty y Vieu cuyos libros [Ramsay and Silverman, 2002], [Ramsay and Silverman, 2005], [Ferraty and Vieu, 2006] y [Ferraty and Romain, 2010] presentan importantes resultados teóricos y prácticos, lo que los convierte en los libros más utilizados en este área (ver [González Manteiga and Vieu, 2007]).

Si bien la expresión datos funcionales se refiere tanto a curvas como a superficies, en el transcurso de esta tesis, utilizaremos la expresión **datos funcionales** para referirnos a curvas mientras que, para referirnos a superficies de mayores dimensiones, utilizaremos la expresión **campos aleatorios**.

En los próximos capítulos extenderemos los conceptos de las Secciones 1.1–1.4 al contexto infinito dimensional. Estudiaremos propiedades asintóticas de estimadores de la función de densidad para datos funcionales y para campos aleatorios y de la función de regresión para datos funcionales. Además, aplicaremos los resultados obtenidos al problema de clasificación funcional y presentaremos ejemplos que muestren el desempeño de los métodos presentados.

CAPÍTULO 2

ESTIMACIÓN DE DENSIDADES PARA DATOS FUNCIONALES

La estimación de la función de densidad marginal de un proceso estocástico a partir de una trayectoria muestral $\mathcal{X}(t)$ observada en el intervalo $[0, T_n]$, es un tema que ha sido estudiado en el marco de la estadística de procesos estocásticos desde finales de los años 70. Algunas referencias al respecto son los trabajos de [Banon, 1978], [Nguyen, 1979] y [Banon and Nguyen, 1981] para procesos de difusión (aquellos que son soluciones de ecuaciones diferenciales estocásticas) estacionarios, donde la teoría asintótica se estudia cuando la longitud T_n del intervalo de observación tiende a infinito.

Para procesos irregulares estacionarios, una referencia clave para el problema de estimar la densidad marginal es la de [Castellana and Leadbetter, 1986] quienes sostienen que *para procesos estacionarios Gaussianos, la irregularidad de las trayectorias corresponde a una menor correlación entre $\mathcal{X}(t)$ y $\mathcal{X}(t+s)$ y por lo tanto, las trayectorias contienen más información que permite obtener mejores velocidades de convergencia de la varianza a cero*. Estos autores encontraron, bajo ciertas condiciones llamadas CL, tasas de convergencia paramétricas $\sqrt{T_n}$ para estimadores de densidad de tipo δ (estimadores basados en núcleos). Este fenómeno motivó un gran número de trabajos, algunos de los cuales están resumidos en [Bosq, 1998] para estimadores de núcleo y en [Kutoyants, 2004] para el caso particular de procesos de difusión. Más precisamente, la convergencia casi segura del estimador de núcleo fue estudiado por [Bosq, 1997] bajo las condiciones CL para procesos a tiempo continuo y por [Kutoyants, 1997] para el caso particular de procesos de difusión ergódicos. Por otro lado, [Blanke and Bosq, 1997] especificaron condiciones suficientes y necesarias para obtener velocidades de convergencia paramétrica para estimadores de núcleo, [Blanke, 2004] obtuvo diferentes velocidades de convergencia (casi

segura, puntual y uniforme) según la regularidad del proceso y [Bosq and Davydov, 1999] estudiaron estimadores de densidad basados en tiempos locales.

El primer autor en definir y estudiar el estimador k -NN cuando un camino muestral simple es observado sobre $[0, T_n]$ vía la medida de ocupación fue [Labrador, 2006, Labrador, 2008], quién definió un estimador $\sqrt{\frac{T_n}{\log T_n}}$ -consistente como

$$\hat{f}_{T_n}(x) = \frac{k_{T_n}}{2T_n\delta_{T_n}} \quad (2.1)$$

donde $k_{T_n} \in (0, T_n)$ representa la cantidad de tiempo que el proceso $\mathcal{X}(t)$ pasa en el intervalo $I_{(x, \delta_{T_n})}$ de centro x y de radio aleatorio $\delta_{T_n, x} = \delta_{T_n}$.

El objetivo de este capítulo es, a partir de n trayectorias independientes de un proceso estocástico \mathcal{X} definido en un intervalo fijo $[0, T] \doteq T$ que verifica el modelo

$$\mathcal{X}(t) = \mu(t) + e(t), \quad t \in T, \quad (2.2)$$

estimar la densidad marginal de \mathcal{X} mediante la extensión del estimador definido por Labrador al contexto independiente. Aquí, $\mu(t)$ es la función media, $e(t)$ es un proceso estocástico estacionario de primer orden con media nula. Si la media $\mu(t)$ es constante, entonces estaremos en un contexto similar al de Labrador pero ahora con n trayectorias observadas en el intervalo fijo T y la teoría asintótica será para $n \rightarrow \infty$. Si $\mu(t)$ no es constante, entonces el proceso observado $\mathcal{X}(t)$ no será estacionario por lo que, para obtener resultados teóricos, será necesario utilizar herramientas diferentes a las utilizadas para el caso estacionario.

Es intuitivamente claro que la información provista por n muestras independientes de un proceso estacionario observado en el intervalo fijo T debe tener al menos la misma información (en general más aún) que una sola trayectoria observada en el intervalo creciente $[0, T_n]$ por lo que, para el caso estacionario, debería ser posible extender los resultados anteriores y obtener además de resultados de velocidades de convergencia, la distribución límite del estimador. Por otro lado, y a diferencia del caso en que observamos una sola trayectoria en un intervalo creciente, podremos abordar el caso no estacionario dado por el modelo (2.2) a partir de la estimación de la media $\mu(t)$.

Al igual que Castellana y Leadbetter, estamos interesados en procesos irregulares ya que ellos nos permitirán obtener mejores resultados. Por lo tanto, contrariamente a lo que ocurre en general en el análisis de datos funcionales, ningún *procedimiento de regularización* ([Hastie et al., 1995]) o *método de filtrado* será utilizado para suavizar los datos.

Este capítulo está organizado como sigue. En la Sección 2.1 definimos un estimador para la densidad marginal de un proceso estocástico estacionario el cual probamos que es \sqrt{n} -consistente ya que obtenemos normalidad asintótica de orden $n^{-1/2}$ y en la Sección 2.2 extendemos la definición anterior a procesos no estacionarios y probamos que este estimador es $n^{1/4}$ -consistente. Finalmente, en las Secciones 2.3-2.5 aplicamos nuestros resultados para definir una nueva regla de clasificación funcional y mediante estudios de simulación y ejemplos con datos reales, mostramos el desempeño de nuestros métodos.

2.1. Estimando la densidad de un proceso estacionario

En esta sección definimos el estimador de la densidad marginal desconocida de un proceso estocástico estacionario de primer orden. Para ello supondremos que en el modelo (2.2) la función media $\mu(t)$ es constante con respecto al tiempo, es decir,

$$\mathcal{X}(t) = \mu + e(t), \quad t \in T,$$

donde $e(t)$ es un proceso estocástico estacionario de primer orden con media nula. Para este estimador presentamos resultados de consistencia puntual, velocidades de convergencia puntual y uniforme y distribución asintótica.

2.1.1. Preliminaries y definición del estimador

Sea $T \subset \mathbb{R}$ un intervalo finito de longitud $|T|$ y $\{\mathcal{X}(t, \omega) : t \in T, \omega \in \Omega\}$ un proceso estocástico a valores reales, estacionario de primer orden definido en un espacio de probabilidad (Ω, \mathcal{A}, P) . Supongamos además que la densidad marginal no conocida de $\mathcal{X}(t, \omega)$ es $f_{\mathcal{X}}$ y que $\mathcal{X}(t, \omega)$ admite un tiempo local $l_T(\cdot, \mathcal{X})$. Esto es, si la medida de ocupación ν asociada con el proceso $\mathcal{X}(t, \omega)$

$$\nu(A, \mathcal{X}) \doteq \int_T \mathbb{I}_A(\mathcal{X}(t, \omega)) d\lambda(t), \quad A \in \mathfrak{B}(\mathbb{R}), \quad \omega \in \Omega$$

es absolutamente continua con respecto a la medida de Lebesgue λ , entonces el **tiempo local** es definido como una versión regular de la derivada de Radon–Nikodym $d\nu/d\lambda$ para casi todo ω . Más precisamente, el tiempo local es la única función $l_T(\cdot, \mathcal{X}) \in [0, \infty)$ (Teorema de Radon–Nikodym) tal que

$$\nu(A, \mathcal{X}) = \int_T \mathbb{I}_A(\mathcal{X}(t, \omega)) d\lambda(t) = \int_A l_T(u, \mathcal{X}) d\lambda(u).$$

Notación 2.1. En lo que resta del capítulo, omitiremos la variable ω en la expresión $\mathcal{X}(t, \omega)$ sin olvidar que implícitamente está en la misma.

Sea $\{\mathcal{X}_i(t)\}_{i=1}^n$ una muestra aleatoria de $\mathcal{X}(t)$ y sea $I_{(x,r)} = [x - r, x + r]$ el intervalo de centro x y radio r . Para $\{k_n\}$, $k_n/n < |T|$, una sucesión de números reales positivos que tiende a infinito, definimos la variable aleatoria $h_n^{\mathcal{X}} \doteq h_n^{\mathcal{X}}(x)$ de manera tal que $\{\mathcal{X}_i(t)\}_{i=1}^n$ pasen en el intervalo $I_{(x, h_n^{\mathcal{X}}(x))}$, k_n del tiempo. Esto es,

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{I_{(x, h_n^{\mathcal{X}}(x))}}(\mathcal{X}_i(t)) dt. \quad (2.3)$$

Definimos el estimador para la densidad $f_{\mathcal{X}}$ como

$$\widehat{f}_{\mathcal{X}}(x) \doteq \frac{k_n}{2n|T|h_n^{\mathcal{X}}(x)}. \quad (2.4)$$

2.1.2. Hipótesis generales

Consideraremos las siguientes hipótesis:

- H1 $\mathcal{X}(t)$ es un proceso estocástico estacionario de primer orden con función de densidad desconocida $f_{\mathcal{X}}$ el cual admite un tiempo local $l_T(\cdot, \mathcal{X})$;
- H2 la densidad $f_{\mathcal{X}}$ es una función Lipschitz con constante K ;
- H3 $\{\mathcal{X}_i(t), t \in T\}_{i=1}^n$ es una sucesión i.i.d. con la misma distribución que $\{\mathcal{X}(t), t \in T\}$;
- H4 $\{k_n\}$ es una sucesión de números reales positivos que tiende a infinito tal que $k_n/n = o(1)$ y $\sum_{n=1}^{\infty} \exp(-ak_n) < \infty$ para cada $a > 0$;
- H5 para cada $C > 0$,

$$\begin{aligned} & (Cc_n)^{-2} \int_T \int_T \int_{\{x: |u-x| \leq Cc_n\}} \int_{\{x: |v-x| \leq Cc_n\}} (f_{st}(u, v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) du dv ds dt \\ & \rightarrow \int_{T \times T} (f_{st}(x, x) - f_{\mathcal{X}}^2(x)) ds dt \doteq C_0^2(x) > 0, \end{aligned}$$

donde C puede depender de x , $c_n = \frac{k_n}{n}$ y f_{st} es la densidad conjunta de $(\mathcal{X}(s), \mathcal{X}(t))$.

Observación 2.2. Una condición suficiente para que H5 se cumpla es que exista una función integrable $\psi(\cdot)$ tal que

$$|f_{st}(u, v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)| \leq \psi(t - s). \quad (2.5)$$

Más precisamente, el Teorema de la convergencia dominada de Lebesgue junto con el Teorema de diferenciación de Lebesgue implican que de (2.5) se sigue H5.

Observación 2.3. (*Algunos comentarios sobre las hipótesis*)

- En la hipótesis H1, la existencia del tiempo local es necesaria para garantizar la existencia y unicidad de $h_n^{\mathcal{X}}$ y en este caso, $\hat{f}_{\mathcal{X}}$ está bien definido. Más precisamente, consideremos la función

$$G(r) \doteq \frac{1}{n} \sum_{i=1}^n \int_{I(x,r)} l_T(u, \mathcal{X}_i) du$$

para n y x fijos. Dado que l_T es una función positiva, tenemos que $G(r)$ es una función creciente de r con $G(0) = 0$. Más aún, dado que los procesos $\{\mathcal{X}_i(t)\}_{i=1}^n$ admiten un tiempo local entonces, $\lambda\{t : \mathcal{X}'_i(t) \text{ es finita}\} = 0$ para casi todo $t \in T$ para todo i (ver [Geman and Horowitz, 1980] Ejemplo 1, pág. 3) luego, G es una función estrictamente creciente de r . Es decir, las trayectorias $\mathcal{X}_i(t)$ son suficientemente irregulares como para asegurar el crecimiento estricto de G . Por otro lado, debido a la existencia del tiempo local podemos escribir

$$G(r) = \frac{1}{n} \sum_{i=1}^n \int_{I(x,r)} l_T(u, \mathcal{X}_i) du = \frac{1}{n} \sum_{i=1}^n \int_T \mathbb{I}_{I(x,r)}(\mathcal{X}_i(t, \omega)) dt$$

luego, $G(r) \rightarrow |T|$ cuando $r \rightarrow \infty$ y por lo tanto, dado cualquier $a \in \mathbb{R}$ con $a < |T|$ existe un único (la unicidad está asegurada ya que G es estrictamente creciente) $r \in \mathbb{R}$ tal que $G(r) = a$. En nuestro caso, para x fijo, dado k_n , $k_n/n < |T|$, existe $h_n^{\mathcal{X}}(x)$ tal que

$$\frac{k_n}{n} = G(h_n^{\mathcal{X}}(x)) = \frac{1}{n} \sum_{i=1}^n \int_T \mathbb{I}_{I(x, h_n^{\mathcal{X}}(x))}(\mathcal{X}_i(t, \omega)) dt$$

o equivalentemente,

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{I(x, h_n^{\mathcal{X}}(x))}(\mathcal{X}_i(t, \omega)) dt.$$

- Las condiciones en H4 son exactamente las mismas que debe cumplir el parámetro k_n el marco clásico finito dimensional para estimadores de densidad de k -vecinos más cercanos. H2 es también una condición usual para el caso finito dimensional y sirve para controlar el término del sesgo.
- Finalmente, la hipótesis H5 está relacionada a la irregularidad del proceso. Más precisamente, como se señala en [Castellana and Leadbetter, 1986], la condición (2.5) es una limitación fuerte de dependencia entre $\mathcal{X}(s)$ y $\mathcal{X}(t)$, cuando $t - s \rightarrow 0$, una característica que no tiene una analogía en tiempo discreto. En particular, ellos demostraron

que para procesos Gaussianos estacionarios con media cero y función de covarianza

$$\text{cov}(\mathcal{X}(s), \mathcal{X}(t)) = 1 - C|t - s|^\alpha + o(|t - s|^\alpha), \quad C > 0,$$

cuando $0 < \alpha < 2$ la condición (2.5) se verifica con $\psi(t - s) = 1 + K|t - s|^{-\alpha/2}$ en un entorno de $t - s = 0$ y $\psi(t - s) = K'|t - s|$ fuera del entorno de $t - s = 0$ si la covarianza es acotada lejos de 1 e integrable. Aquí, K y K' son constantes. Es bien sabido que procesos Gaussianos con $\alpha < 2$ tienen trayectorias irregulares en contraste con el caso más regular $\alpha = 2$. Más precisamente, si un proceso Gaussiano centrado tiene función de covarianza dada por

$$\text{cov}(\mathcal{X}(s), \mathcal{X}(t)) = 1 - \frac{1}{2}C|t - s|^2 + O\left(\frac{(t - s)^2}{|\log|t - s||^a}\right), \quad C > 0, a > 3,$$

existe una versión del proceso con trayectorias C^1 . Ver por ejemplo, el Ejemplo 1.2 y la Proposición 1.11 en [Azaïs and Wschebor, 2009] pág 28.

2.1.3. Resultados asintóticos

Teorema 2.4. Consistencia. *Supongamos que $H1$, $H3$ y $H4$ se cumplen. Entonces, para todo $x \in \mathbb{R}$ punto de Lebesgue de $f_{\mathcal{X}}$,*

$$\lim_{n \rightarrow \infty} \widehat{f}_{\mathcal{X}}(x) = f_{\mathcal{X}}(x) \quad a.co.$$

Observación 2.5. Si $f_{\mathcal{X}}$ es Lipschitz, todo punto en \mathbb{R} es punto de Lebesgue de $f_{\mathcal{X}}$ lo que implica que el Teorema anterior vale para todo $x \in \mathbb{R}$.

Antes de la demostración del teorema, probaremos un lema que será necesario no sólo en este capítulo sino también el próximo.

Lema 2.6. *Dado un proceso estocástico estacionario $\mathcal{X}(t)$, consideremos la variable aleatoria $Y^c \doteq \int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) dt$ y la variable aleatoria centrada $\overline{Y}^c \doteq Y^c - \mathbb{E}(Y^c)$ donde c puede depender de x . Para t fijo, sea p_c la medida de probabilidad de $\mathcal{X}(t)$, es decir,*

$$p_c \doteq P(\mathcal{X}(t) \in I(x,c)) = \int \mathbb{I}_{\{u: |u-x| \leq c\}} f_{\mathcal{X}}(u) du.$$

Entonces,

$$\mathbb{E}(Y^c) = |T|p_c, \quad |\overline{Y}^c| \leq 2|T| \quad y \quad \text{var}(\overline{Y}^c) = \text{var}(Y^c) \leq |T|^2 p_c.$$

Demostración.

$$\begin{aligned}
\mathbb{E}(Y^c) &= \mathbb{E}\left(\int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) dt\right) \\
&= \int_T \mathbb{E}\left(\mathbb{I}_{I(x,c)}(\mathcal{X}(t))\right) dt && \text{(Teorema de Tonelli)} \\
&= \int_T P(\mathcal{X}(t) \in I(x,c)) dt \\
&= |T|p_c. && \text{(e es estacionario)}
\end{aligned}$$

Usando la igualdad anterior, tenemos que

$$\begin{aligned}
|\bar{Y}^c| &\leq |Y^c| + |\mathbb{E}(Y^c)| \\
&= \left|\int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) dt\right| + |T|p_c \\
&\leq \int_T \left|\mathbb{I}_{I(x,c)}(\mathcal{X}(t))\right| dt + |T| && (|p_c| \leq 1) \\
&\leq 2|T|.
\end{aligned}$$

Además, dado que

$$\begin{aligned}
\text{var}(Y^c) &\leq \mathbb{E}((Y^c)^2) \\
&= \mathbb{E}\left(\left(\int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) dt\right)^2\right) \\
&= |T| \mathbb{E}\left(\int_T \left(\mathbb{I}_{I(x,c)}(\mathcal{X}(t))\right)^2 dt\right) && \text{(Cauchy-Schwartz)} \\
&= |T| \int_T \mathbb{E}\left(\mathbb{I}_{I(x,c)}(\mathcal{X}(t))\right) dt && \text{(Teorema de Tonelli)} \\
&= |T| \int_T P(\mathcal{X}(t) \in I(x,c)) dt \\
&= |T|^2 p_c && \text{(e es estacionario)}
\end{aligned}$$

entonces

$$\text{var}(\bar{Y}^c) = \text{var}(Y^c) \leq |T|^2 p_c.$$

□

Demostración del Teorema 2.4: Sea $x \in \mathbb{R}$ fijo un punto de Lebesgue de $f_{\mathcal{X}}$. Por definición de convergencia casi completa necesitamos probar que para todo $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P\left(\left|\hat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right| > \epsilon\right) < \infty.$$

Sea

$$C_n = C_n(x) \doteq \left\{ \left| \widehat{f_{\mathcal{X}}}(x) - f_{\mathcal{X}}(x) \right| > \epsilon \right\}.$$

Escribimos, $C_n = A_n \cup B_n$ con

$$A_n = A_n(x) = \left\{ h_n^{\mathcal{X}}(x) < \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon)} \right\}$$

y

$$B_n = B_n(x) = \begin{cases} \left\{ h_n^{\mathcal{X}}(x) > \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) - \epsilon)} \right\} & \text{si } f_{\mathcal{X}}(x) > \epsilon \\ \emptyset & \text{si } f_{\mathcal{X}}(x) \leq \epsilon. \end{cases}$$

Ya que $P(C_n) \leq P(A_n) + P(B_n)$ será suficiente probar que

$$\sum_{n=1}^{\infty} P(A_n) < \infty \quad \text{y} \quad \sum_{n=1}^{\infty} P(B_n) < \infty.$$

Para ello definimos

$$a_n = a_n(x) \doteq \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon)} \quad \text{y} \quad b_n = b_n(x) \doteq \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) - \epsilon)}$$

luego, podemos escribir

$$A_n = \{h_n^{\mathcal{X}} < a_n\} \quad \text{y} \quad B_n \subseteq \{h_n^{\mathcal{X}} > b_n, f_{\mathcal{X}}(x) > \epsilon\}.$$

Observemos que

$$h_n^{\mathcal{X}} < a_n \Leftrightarrow \sum_{i=1}^n \underbrace{\int_T \mathbb{I}_{I(x, a_n)}(\mathcal{X}_i(t)) dt}_{\doteq Y_i^{a_n}} > k_n \quad \text{y} \quad h_n^{\mathcal{X}} > b_n \Leftrightarrow \sum_{i=1}^n \underbrace{\int_T \mathbb{I}_{I(x, b_n)}(\mathcal{X}_i(t)) dt}_{\doteq Y_i^{b_n}} < k_n$$

donde $\{Y_i^{a_n}\}_{i=1}^n$ y $\{Y_i^{b_n}\}_{i=1}^n$ son variables aleatorias independientes pues los procesos $\{\mathcal{X}_i(t)\}_{i=1}^n$ lo son y además,

$$P(A_n) = P(h_n^{\mathcal{X}} < a_n) = P\left(\sum_{i=1}^n Y_i^{a_n} > k_n\right) \quad (2.6)$$

y

$$P(B_n) \leq P(h_n^{\mathcal{X}} > b_n, f_{\mathcal{X}}(x) > \epsilon) = P\left(\sum_{i=1}^n Y_i^{b_n} < k_n, f_{\mathcal{X}}(x) > \epsilon\right). \quad (2.7)$$

Comencemos probando que $\sum_{n=1}^{\infty} P(A_n) < \infty$. El Lema 2.6 con $c = a_n$ implica que $\mathbb{E}(Y_i^{a_n}) = |T|p_{a_n}$. Luego, en (2.6) tenemos

$$P(A_n) = P(h_n^{\mathcal{X}} < a_n) = P\left(\sum_{i=1}^n Y_i^{a_n} > k_n\right) = P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n - n|T|p_{a_n}\right). \quad (2.8)$$

La hipótesis H3 implica que

$$a_n = \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon)} = \frac{1}{2|T|(f_{\mathcal{X}}(x) + \epsilon)} \frac{k_n}{n} \xrightarrow{n \rightarrow \infty} 0$$

y dado que x es un punto de Lebesgue de $f_{\mathcal{X}}$,

$$\frac{p_{a_n}}{2a_n} = \frac{1}{2a_n} \int_{\{u:|u-x|\leq a_n\}} f_{\mathcal{X}}(u) du \xrightarrow{n \rightarrow \infty} f_{\mathcal{X}}(x).$$

Es decir, existe $N_1 = N_1(x)$ tal que si $n \geq N_1$,

$$\left| \frac{p_{a_n}}{2a_n} - f_{\mathcal{X}}(x) \right| < \epsilon/2 \quad (2.9)$$

o equivalentemente, usando la definición de a_n

$$p_{a_n} < 2a_n (f_{\mathcal{X}}(x) + \epsilon/2) = 2 \left(\frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon)} \right) (f_{\mathcal{X}}(x) + \epsilon/2) = \frac{k_n}{n} \frac{1}{|T|} \left(\frac{f_{\mathcal{X}}(x) + \epsilon/2}{f_{\mathcal{X}}(x) + \epsilon} \right).$$

Luego

$$n|T|p_{a_n} < k_n \left(\frac{f_{\mathcal{X}}(x) + \epsilon/2}{f_{\mathcal{X}}(x) + \epsilon} \right).$$

Con esta desigualdad en (2.8) tenemos que

$$\begin{aligned} P(A_n) &\leq P \left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n - k_n \left(\frac{f_{\mathcal{X}}(x) + \epsilon/2}{f_{\mathcal{X}}(x) + \epsilon} \right) \right) \\ &= P \left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n \left(\frac{\epsilon}{2(f_{\mathcal{X}}(x) + \epsilon)} \right) \right) \quad (2.10) \\ &= P \left(\sum_{i=1}^n \bar{Y}_i^{a_n} > n \left(\frac{k_n C_1}{n} \right) \right) \quad \left(C_1 = C_1(x) = \frac{\epsilon}{2(f_{\mathcal{X}}(x) + \epsilon)} > 0 \right) \\ &\leq 2 \exp \left(- \frac{\frac{k_n^2 C_1^2}{n^2} n}{2|T|^2 p_{a_n} \left(1 + \frac{k_n C_1}{n} \frac{2|T|}{|T|^2 p_{a_n}} \right)} \right). \end{aligned}$$

En la última desigualdad hemos usado la desigualdad de Bernstein (Corolario A.32) con $M = 2|T|$, $\sigma^2 = |T|^2 p_{a_n}$ y $\epsilon \doteq \frac{k_n C_1}{n} > 0$. Trabajando el exponente de esta última expresión resulta,

$$\begin{aligned} \frac{\frac{k_n^2 C_1^2}{n^2} n}{2|T|^2 p_{a_n} \left(1 + \frac{k_n C_1}{n} \frac{2|T|}{|T|^2 p_{a_n}} \right)} &= \frac{k_n^2 C_1^2}{2n|T|^2 p_{a_n} \left(1 + \frac{k_n C_1}{n} \frac{2|T|}{|T|^2 p_{a_n}} \right)} \\ &= \frac{k_n^2 C_1^2}{2n|T|^2 p_{a_n} + 4k_n C_1 |T|} \\ &> \frac{k_n^2 C_1^2}{4n|T|^2 p_{a_n} (f_{\mathcal{X}}(x) + \epsilon) + 4|T| k_n C_1} \quad (\text{de (2.9), } p_{a_n} < 2a_n (f_{\mathcal{X}}(x) + \epsilon)) \\ &= \frac{k_n^2 C_1^2}{2|T| k_n + 4|T| k_n C_1} \quad (k_n = 2a_n n |T| (f_{\mathcal{X}}(x) + \epsilon)) \\ &= k_n C_2. \quad \left(C_2 = C_2(x, \epsilon) \doteq \frac{C_1^2}{2|T|(1 + 2C_1)} \right) \end{aligned}$$

Luego, en (2.10) para $n \geq N_1$ tenemos que

$$P(A_n) \leq 2 \exp(-C_2 k_n).$$

Finalmente, la hipótesis H4 implica que

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{N_1(x)-1} P(A_n) + \sum_{n=N_1(x)}^{\infty} P(A_n) \leq N_1(x) + 2 \sum_{n=N_1(x)}^{\infty} \exp(-C_2 k_n) < \infty. \quad (2.11)$$

Continuemos probando que $\sum_{n=1}^{\infty} P(B_n) < \infty$. El Lema 2.6, ahora con $c = b_n$ implica que $\mathbb{E}(Y_i^{b_n}) = |T|p_{b_n}$ luego, en (2.7) tenemos que

$$\begin{aligned} P(B_n) &\leq P(h_n^{\mathcal{X}} > b_n, f_{\mathcal{X}}(x) > \epsilon) \\ &= P\left(\sum_{i=1}^n Y_i^{b_n} < k_n, f_{\mathcal{X}}(x) > \epsilon\right) \\ &= P\left(\sum_{i=1}^n \bar{Y}_i^{b_n} < k_n - n|T|p_{b_n}, f_{\mathcal{X}}(x) > \epsilon\right). \end{aligned} \quad (2.12)$$

De la hipótesis H4 y dado que $f_{\mathcal{X}}(x) > \epsilon$,

$$b_n = \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) - \epsilon)} = \frac{1}{2|T|(f_{\mathcal{X}}(x) - \epsilon)} \frac{k_n}{n} \xrightarrow{n \rightarrow \infty} 0,$$

por lo tanto, dado que x es un punto de Lebesgue de $f_{\mathcal{X}}$,

$$\frac{p_{b_n}}{2b_n} = \frac{1}{2b_n} \int_{\{u: |u-x| \leq b_n\}} f_{\mathcal{X}}(u) du \xrightarrow{n \rightarrow \infty} f_{\mathcal{X}}(x).$$

Es decir, existe $M_1 = M_1(x)$ tal que si $n \geq M_1$,

$$\left| \frac{p_{b_n}}{2b_n} - f_{\mathcal{X}}(x) \right| < \epsilon/2 \quad (2.13)$$

o, usando la definición de b_n ,

$$p_{b_n} > 2b_n(f_{\mathcal{X}}(x) - \epsilon/2) = 2 \left(\frac{k_n}{2n|T|(f_{\mathcal{X}}(x) - \epsilon)} \right) (f_{\mathcal{X}}(x) - \epsilon/2) = \frac{k_n}{n} \frac{1}{|T|} \left(\frac{f_{\mathcal{X}}(x) - \epsilon/2}{f_{\mathcal{X}}(x) - \epsilon} \right).$$

Luego

$$n|T|p_{b_n} > k_n \left(\frac{f_{\mathcal{X}}(x) - \epsilon/2}{f_{\mathcal{X}}(x) - \epsilon} \right).$$

Con esta desigualdad en (2.12) tenemos

$$\begin{aligned}
P(B_n) &\leq P\left(\sum_{i=1}^n \bar{Y}_i^{b_n} < k_n - k_n \left(\frac{f_{\mathcal{X}}(x) - \epsilon/2}{f_{\mathcal{X}}(x) - \epsilon}\right), f_{\mathcal{X}}(x) > \epsilon\right) \\
&= P\left(\sum_{i=1}^n \bar{Y}_i^{b_n} < k_n \left(\frac{-\epsilon}{2(f_{\mathcal{X}}(x) - \epsilon)}\right), f_{\mathcal{X}}(x) > \epsilon\right) \\
&= P\left(\sum_{i=1}^n (-\bar{Y}_i^{b_n}) > k_n \left(\frac{\epsilon}{2(f_{\mathcal{X}}(x) - \epsilon)}\right), f_{\mathcal{X}}(x) > \epsilon\right) \\
&= P\left(\sum_{i=1}^n (-\bar{Y}_i^{b_n}) > n \left(\frac{k_n D_1}{n}\right), f_{\mathcal{X}}(x) > \epsilon\right) \quad \left(D_1 = D_1(x) \doteq \frac{\epsilon}{2(f_{\mathcal{X}}(x) - \epsilon)} > 0\right) \\
&\leq 2 \exp\left(-\frac{\frac{k_n^2 D_1^2}{n^2} n}{2|T|^2 p_{b_n} \left(1 + \frac{k_n D_1}{n} \frac{2|T|}{|T|^2 p_{b_n}}\right)}\right),
\end{aligned} \tag{2.14}$$

con $D_1 = D_1(x, \epsilon) \doteq \frac{\epsilon}{2(f_{\mathcal{X}}(x) - \epsilon)} > 0$. En la última desigualdad hemos usado la desigualdad de Bernstein (Corolario A.32) con $M = 2|T|$, $\sigma^2 = |T|^2 p_{b_n}$ y $\epsilon \doteq \frac{k_n D_1}{n} > 0$. rabajando el exponente,

$$\begin{aligned}
\frac{\frac{D_1^2 k_n^2}{n^2} n}{2|T|^2 p_{b_n} \left(1 + \frac{k_n}{n} \frac{2|T|}{|T|^2 p_{b_n}}\right)} &= \frac{D_1^2 k_n^2}{2n|T|^2 p_{b_n} \left(1 + \frac{k_n}{n} \frac{2|T|}{|T|^2 p_{b_n}}\right)} \\
&= \frac{D_1^2 k_n^2}{2n|T|^2 p_{b_n} + 4D_1 k_n |T|} \\
&> \frac{D_1^2 k_n^2}{4n|T|^2 b_n (f_{\mathcal{X}}(x) + \epsilon) + 4D_1 k_n |T|} \quad (\text{de (2.13), } p_{b_n} < 2b_n(f_{\mathcal{X}}(x) + \epsilon)) \\
&= \frac{D_1^2 k_n^2}{2|T| \frac{k_n}{(f_{\mathcal{X}}(x) - \epsilon)} (f_{\mathcal{X}}(x) + \epsilon) + 4D_1 k_n |T|} \quad \left(b_n = \frac{k_n}{2n|T| (f_{\mathcal{X}}(x) - \epsilon)}\right) \\
&= k_n \frac{D_1^2}{|T| \left(2 \left(\frac{f_{\mathcal{X}}(x) + \epsilon}{f_{\mathcal{X}}(x) - \epsilon}\right) + 4D_1\right)} \\
&= D_2 k_n,
\end{aligned}$$

con $D_2 = D_2(x, \epsilon) \doteq \frac{D_1^2}{|T| \left(2 \left(\frac{f_{\mathcal{X}}(x) + \epsilon}{f_{\mathcal{X}}(x) - \epsilon}\right) + 4D_1\right)} > 0$. Luego en (2.14) para todo $n \geq M_1$ tenemos que

$$P(B_n) \leq 2 \exp(-D_2 k_n).$$

La hipótesis H4 implica que

$$\sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{M_1(x)-1} P(B_n) + \sum_{n=M_1(x)}^{\infty} P(B_n) \leq M_1(x) + 2 \sum_{n=1}^{\infty} \exp(-D_2 k_n) < \infty. \tag{2.15}$$

Finalmente, de (2.11) y (2.15) tenemos el resultado.

□

El Teorema 2.4 asegura la consistencia de nuestro estimador, es decir, que converge a la función de densidad marginal cuando el tamaño de la muestra aumenta. Lo que nos gustaría saber es cuán rápido lo hace, es por ello que en el teorema siguiente, hallamos tasas de convergencia de $\widehat{f}_{\mathcal{X}}$.

Teorema 2.7. Velocidades de Convergencia. *Supongamos que H1-H3 se cumplen. Elijamos dos sucesiones $\{k_n\}$ y $\{v_n\}$ de números reales positivos que tienden a infinito tales que $(k_n/n)v_n = o(1)$ y $\sum_{n=1}^{\infty} \exp(-a(k_n/v_n)) < \infty$ para cada $a > 0$. Para ese k_n supongamos que H5 se verifica y además que $f_{\mathcal{X}}$ es estrictamente positiva. Entonces, para todo $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} v_n(\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)) = 0 \quad a.co.$$

Observación 2.8. Nuestras hipótesis implican que podemos elegir k_n tal que $v_n = n^\gamma$ para cualquier $\gamma < \frac{1}{2}$, este es el tipo de resultados que se obtienen en el marco paramétrico. Más precisamente, si $k_n = n^\beta$ y $v_n = n^\gamma$, $\gamma < \frac{1}{2}$ para que las condiciones $(k_n/n)v_n = o(1)$ y $k_n/v_n \rightarrow \infty$ se cumplan, es suficiente que $\beta - 1 + \gamma < 0$ y $\beta - \gamma > 0$ o equivalentemente, $\beta < 1 - \gamma$ y $\beta > \gamma$. Luego, dado $\gamma < \frac{1}{2}$ podemos elegir β de manera que las hipótesis se cumplen (ver Figura 1).

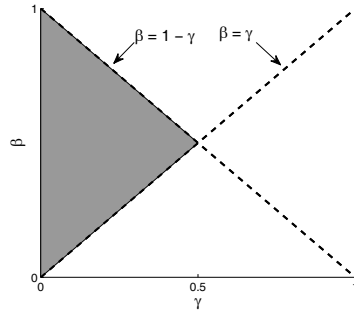


FIGURA 1. El área sombreada muestra la intersección de las regiones $\beta < 1 - \gamma$ y $\beta > \gamma$.

Antes de probar el resultado necesitamos probar el siguiente lema.

Lema 2.9. *Dado un proceso estocástico estacionario $\mathcal{X}(t)$, consideremos la variable aleatoria, $Y^c \doteq \int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) dt$ y la variable aleatoria centrada $\overline{Y}^c \doteq Y^c - \mathbb{E}(Y^c)$ donde c puede depender de x . Entonces,*

$$\text{var}(Y^c) = \text{var}(\overline{Y}^c) = \int_T \int_T \int_{\{u:|u-x|\leq c\}} \int_{\{v:|v-x|\leq c\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) du dv dt ds.$$

Demostración. Dado que

$$\begin{aligned}
\mathbb{E}(Y^c) &= \mathbb{E}\left(\int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) dt\right) \\
&= \int_T \mathbb{E}\left(\mathbb{I}_{I(x,c)}(\mathcal{X}(t))\right) dt && \text{(Teorema de Tonelli)} \\
&= \int_T P(\mathcal{X}(t) \in I(x,c)) dt \\
&= \int_T \int_{\{x:|u-x|\leq c\}} f_{\mathcal{X}}(u) du dt
\end{aligned}$$

y

$$\begin{aligned}
\mathbb{E}((Y^c)^2) &= \mathbb{E}\left(\left(\int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) dt\right)^2\right) \\
&= \mathbb{E}\left(\int_T \int_T \mathbb{I}_{I(x,c)}(\mathcal{X}(t)) \mathbb{I}_{I(x,c)}(\mathcal{X}(s)) dt ds\right) \\
&= \int_T \int_T \mathbb{E}\left(\mathbb{I}_{I(x,c)}(\mathcal{X}(t)) \mathbb{I}_{I(x,c)}(\mathcal{X}(s))\right) dt ds && \text{(Teorema de Tonelli)} \\
&= \int_T \int_T P(\mathcal{X}(t) \in I(x,c), \mathcal{X}(s) \in I(x,c)) dt ds \\
&= \int_T \int_T \int_{\{u:|u-x|\leq c\}} \int_{\{v:|v-x|\leq c\}} f_{st}(u,v) du dv dt ds
\end{aligned}$$

entonces,

$$\begin{aligned}
\text{var}(\overline{Y^c}) &= \text{var}(Y^c) = \mathbb{E}((Y^c)^2) - \mathbb{E}(Y^c)^2 \\
&= \int_T \int_T \int_{\{u:|u-x|\leq c\}} \int_{\{v:|v-x|\leq c\}} f_{st}(u,v) du dv dt ds \\
&\quad - \int_T \int_T \int_{\{u:|u-x|\leq c\}} \int_{\{v:|v-x|\leq c\}} f_{\mathcal{X}}(u) f_{\mathcal{X}}(v) du dv dt ds \\
&= \int_T \int_T \int_{\{u:|u-x|\leq c\}} \int_{\{v:|v-x|\leq c\}} (f_{st}(u,v) - f_{\mathcal{X}}(u) f_{\mathcal{X}}(v)) du dv dt ds.
\end{aligned}$$

□

Demostración del Teorema 2.7: Por definición de convergencia casi completa necesitamos probar que

$$\forall \epsilon > 0, \quad \sum_{n=1}^{\infty} P\left(v_n \left|\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right| > \epsilon\right) < \infty.$$

Sean $\epsilon_n \doteq \frac{\epsilon}{v_n}$ y

$$C_n = C_n(x) \doteq \left\{v_n \left|\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right| > \epsilon\right\} = \left\{\left|\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right| > \epsilon_n\right\}.$$

Análogamente a lo hecho en la demostración del Teorema 2.4, podemos escribir

$$C_n = A_n \cup B_n$$

con

$$A_n = \left\{ h_n^{\mathcal{X}}(x) < \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon_n)} \right\}$$

y

$$B_n = \begin{cases} \left\{ h_n^{\mathcal{X}}(x) > \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) - \epsilon_n)} \right\} & \text{si } f_{\mathcal{X}}(x) > \epsilon_n \\ \emptyset & \text{si } f_{\mathcal{X}}(x) \leq \epsilon_n. \end{cases}$$

Definiendo

$$a_n = a_n(x) \doteq \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon_n)} \quad \text{y} \quad b_n = b_n(x) \doteq \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) - \epsilon_n)}$$

tenemos que

$$A_n = \{h_n^{\mathcal{X}} < a_n\} \quad \text{y} \quad B_n \subseteq \{h_n^{\mathcal{X}} > b_n, f_{\mathcal{X}}(x) > \epsilon_n\}.$$

Además,

$$h_n^{\mathcal{X}} < a_n \Leftrightarrow \sum_{i=1}^n \underbrace{\int_T \mathbb{I}_{I(x, a_n)}(\mathcal{X}_i(t)) dt}_{\doteq Y_i^{a_n}} > k_n \quad \text{y} \quad h_n^{\mathcal{X}} > b_n \Leftrightarrow \sum_{i=1}^n \underbrace{\int_T \mathbb{I}_{I(x, b_n)}(\mathcal{X}_i(t)) dt}_{\doteq Y_i^{b_n}} < k_n,$$

de donde se sigue que

$$P(A_n) = P(h_n^{\mathcal{X}} < a_n) = P\left(\sum_{i=1}^n Y_i^{a_n} > k_n\right) \quad (2.16)$$

y

$$P(B_n) \leq P(h_n^{\mathcal{X}} > b_n) = P\left(\sum_{i=1}^n Y_i^{b_n} < k_n, f_{\mathcal{X}}(x) > \epsilon_n\right),$$

con $\{Y_i^{a_n}\}_{i=1}^n$ y $\{Y_i^{b_n}\}_{i=1}^n$ variables aleatorias independientes pues los procesos $\{\mathcal{X}_i(t)\}_{i=1}^n$ lo son. Por lo tanto, dado que $P(C_n) \leq P(A_n) + P(B_n)$, será suficiente probar que

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P\left(\sum_{i=1}^n Y_i^{a_n} > k_n\right) < \infty \quad (2.17)$$

y

$$\sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{\infty} P\left(\sum_{i=1}^n Y_i^{b_n} < k_n, f_{\mathcal{X}}(x) > \epsilon_n\right) < \infty. \quad (2.18)$$

La demostración de (2.18) es análoga a la de (2.17) y por lo tanto será omitida. Para probar (2.17), en el Lema 2.6 tomamos $c = a_n$ de donde resulta que $\mathbb{E}(Y_i^{a_n}) = |T|p_{a_n}$. Luego, en (2.16) tenemos que

$$P(A_n) = P\left(\sum_{i=1}^n Y_i^{a_n} > k_n\right) = P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n - n|T|p_{a_n}\right). \quad (2.19)$$

El Teorema del valor medio y la hipótesis H2 nos aseguran la existencia de $x_n \in I_{(x, a_n)}$ para el cual

$$\frac{p_{a_n}}{2a_n} = \frac{1}{2a_n} \int_{\{u: |u-x| \leq a_n\}} f_{\mathcal{X}}(u) du = f_{\mathcal{X}}(x_n).$$

Usando nuevamente la hipótesis H2 junto con la igualdad anterior tenemos que

$$\left| \frac{p_{a_n}}{2a_n} - f_{\mathcal{X}}(x) \right| = |f_{\mathcal{X}}(x_n) - f_{\mathcal{X}}(x)| \leq K |x_n - x| \leq K a_n.$$

Luego, $\frac{p_{a_n}}{2a_n} - f_{\mathcal{X}}(x) \leq K a_n$, lo que implica que

$$p_{a_n} \leq 2a_n(f_{\mathcal{X}}(x) + K a_n). \quad (2.20)$$

Observemos que

$$\begin{aligned} K a_n &= K \left(\frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon_n)} \right) && \text{(definición de } a_n) \\ &= K \left(\frac{1}{2|T|(f_{\mathcal{X}}(x) + \epsilon_n)} \right) \left(\frac{k_n}{n} \right) && (2.21) \\ &< K \left(\frac{1}{2|T|f_{\mathcal{X}}(x)} \right) \left(\frac{k_n}{n} \right) && (\epsilon_n > 0, \forall n) \\ &= C_1 \left(\frac{k_n}{n} \right). && \left(C_1 = C_1(x) \doteq K \left(\frac{1}{2|T|f_{\mathcal{X}}(x)} \right) \right). \end{aligned}$$

Dado que por hipótesis $v_n \left(\frac{k_n}{n} \right) \rightarrow 0$, $\forall \epsilon > 0$ existe N_0 tal que si $n \geq N_0$, $v_n \left(\frac{k_n}{n} \right) \leq \frac{\epsilon}{2C_1}$. Por lo tanto, en (2.21) para todo $n \geq N_0$ tenemos que

$$K a_n \leq C_1 \left(\frac{k_n}{n} \right) < \frac{\epsilon}{2v_n}. \quad (2.22)$$

Reemplazando (2.22) y la definición de a_n en (2.20) tenemos

$$p_{a_n} \leq 2a_n \left(f_{\mathcal{X}}(x) + \frac{\epsilon}{2v_n} \right) = \frac{k_n}{n} \frac{1}{|T|} \left(\frac{f_{\mathcal{X}}(x) + \epsilon_n/2}{f_{\mathcal{X}}(x) + \epsilon_n} \right).$$

Sustituyendo esta desigualdad en (3.2.3) resulta

$$\begin{aligned} P(A_n) &= P \left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n - n|T|p_{a_n} \right) \\ &\leq P \left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n - k_n \left(\frac{f_{\mathcal{X}}(x) + \epsilon_n/2}{f_{\mathcal{X}}(x) + \epsilon_n} \right) \right) \\ &= P \left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n \left(\frac{\epsilon_n}{2(f_{\mathcal{X}}(x) + \epsilon_n)} \right) \right). \end{aligned}$$

Dado que $\epsilon_n \rightarrow 0$ pues $v_n \rightarrow \infty$, existe $N_1 = N_1(x)$ tal que si $n \geq N_1$, $\epsilon_n < f_{\mathcal{X}}(x)/2$. Luego,

$$\begin{aligned}
P(A_n) &= P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n \left(\frac{\epsilon_n}{2(f_{\mathcal{X}}(x) + \epsilon_n)}\right)\right) \\
&\leq P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n \left(\frac{\epsilon_n}{2(f_{\mathcal{X}}(x) + f_{\mathcal{X}}(x)/2)}\right)\right) \\
&= P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n \left(\frac{\epsilon_n}{3f_{\mathcal{X}}(x)}\right)\right) \tag{2.23} \\
&= P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > k_n \left(\frac{\epsilon}{3f_{\mathcal{X}}(x)v_n}\right)\right) \quad \left(\epsilon_n = \frac{\epsilon}{v_n}\right) \\
&= P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > C_2 \left(\frac{k_n}{v_n}\right)\right) \quad \left(C_2 = C_2(x) = \frac{\epsilon}{3f_{\mathcal{X}}(x)}\right) \\
&= P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > \left(\frac{C_2 k_n}{nv_n}\right) n\right).
\end{aligned}$$

Con el fin de aplicar la desigualdad de Bernstein, necesitamos acotar $\text{var}(\bar{Y}_i^{a_n})$. Del Lema 2.9 con $c = a_n$ resulta que

$$\text{var}(\bar{Y}_i^{a_n}) = \int_T \int_T \int_{\{u:|u-x|\leq a_n\}} \int_{\{v:|v-x|\leq a_n\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, du \, dv \, dt \, ds.$$

Para poder aplicar la hipótesis H5 escribimos

$$\frac{1}{a_n^2} \text{var}(\bar{Y}_i^{a_n}) = \frac{1}{a_n^2} \int_T \int_T \int_{\{u:|u-x|\leq a_n\}} \int_{\{v:|v-x|\leq a_n\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, du \, dv \, dt \, ds. \tag{2.24}$$

Notemos que

$$a_n = \frac{k_n}{2n|T|(f_{\mathcal{X}}(x) + \epsilon_n)} \doteq r_n \left(\frac{k_n}{n}\right)$$

dado que $\epsilon_n \rightarrow 0$,

$$r_n = \left(\frac{1}{2|T|(f_{\mathcal{X}}(x) + \epsilon_n)}\right) \rightarrow \left(\frac{1}{2|T|f_{\mathcal{X}}(x)}\right) \doteq C_3(x) = C_3 > 0.$$

Es decir, existe $N_2 = N_2(x)$ tal que si $n \geq N_2$, $C_4 \leq r_n \leq C_5$ para algunas constantes C_4 y C_5 .

Luego,

$$C_4 \frac{k_n}{n} \leq a_n \leq C_5 \frac{k_n}{n}. \tag{2.25}$$

Por lo tanto en (2.24) tenemos que

$$\begin{aligned}
\frac{1}{a_n^2} \text{var}(\bar{Y}_i^{a_n}) &\leq \frac{1}{(C_4 \frac{k_n}{n})^2} \int_T \int_T \int_{\{u:|u-x|\leq C_5 \frac{k_n}{n}\}} \int_{\{v:|v-x|\leq C_5 \frac{k_n}{n}\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, dudvdt ds \\
&= \left(\frac{C_5}{C_4}\right)^2 \frac{1}{(C_5 \frac{k_n}{n})^2} \int_T \int_T \int_{\{u:|u-x|\leq C_5 \frac{k_n}{n}\}} \int_{\{v:|v-x|\leq C_5 \frac{k_n}{n}\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, dudvdt ds
\end{aligned}$$

y

$$\begin{aligned} \frac{1}{a_n^2} \text{var} (\bar{Y}_i^{a_n}) &\geq \frac{1}{(C_5 \frac{k_n}{n})^2} \int_T \int_T \int_{\{u: |u-x| \leq C_4 \frac{k_n}{n}\}} \int_{\{v: |v-x| \leq C_4 \frac{k_n}{n}\}} (f_{st}(u, v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, dudvdt ds \\ &= \left(\frac{C_4}{C_5}\right)^2 \frac{1}{(C_4 \frac{k_n}{n})^2} \int_T \int_T \int_{\{u: |u-x| \leq C_4 \frac{k_n}{n}\}} \int_{\{v: |v-x| \leq C_4 \frac{k_n}{n}\}} (f_{st}(u, v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, dudvdt ds \end{aligned}$$

De estas dos desigualdades y de la hipótesis H5 para $C = C_4$ y $C = C_5$ se sigue que existe $N_3 = N_3(x)$ tal que si $n \geq N_3$,

$$C_6 \doteq \left(\frac{C_4}{C_5} C_0\right)^2 \leq \frac{1}{a_n^2} \text{var} (\bar{Y}_i^{a_n}) \leq \left(\frac{C_5}{C_4} C_0\right)^2 \doteq C_7,$$

o equivalentemente,

$$C_6 a_n^2 \leq \text{var} (\bar{Y}_i^{a_n}) \leq C_7 a_n^2,$$

Luego, aplicando la desigualdad de Bernstein (A.32) en (2.23) con $\epsilon = \frac{C_2 k_n}{n v_n} > 0$, $M = 2|T|$ y $\sigma^2 \leq C_7 a_n^2$, para $n \geq N_4 = \max\{N_0, N_1, N_2, N_3\}$, tenemos que

$$P(A_n) \leq P\left(\sum_{i=1}^n \bar{Y}_i^{a_n} > \left(\frac{C_2 k_n}{n v_n}\right) n\right) \leq 2 \exp\left\{-\frac{\frac{C_2^2 k_n^2}{n^2 v_n^2} n}{2C_7 a_n^2 \left(1 + \frac{C_2 k_n}{n v_n} \frac{2|T|}{C_7 a_n^2}\right)}\right\}. \quad (2.26)$$

Reescribiendo el exponente,

$$\begin{aligned} \frac{\frac{C_2^2 k_n^2}{n^2 v_n^2} n}{2C_7 a_n^2 \left(1 + \frac{C_2 k_n}{n v_n} \frac{2|T|}{C_7 a_n^2}\right)} &= \frac{C_2^2 k_n^2}{2n v_n^2 C_7 a_n^2 + 4C_2 k_n |T| v_n} \\ &= \frac{C_2^2 k_n}{2\frac{n}{k_n} v_n^2 C_7 a_n^2 + 4C_2 |T| v_n} \\ &\geq \frac{C_2^2 k_n}{2v_n^2 C_7 C_5^2 \frac{k_n}{n} + 4C_2 |T| v_n} \quad \left(\text{de (2.25), } a_n \leq C_5 \frac{k_n}{n}\right) \\ &\geq \frac{C_2^2 k_n}{2v_n C_7 C_5^2 \frac{\epsilon}{2C_1} + 4C_2 |T| v_n} \quad \left(\text{si } n \geq N_0, v_n \left(\frac{k_n}{n}\right) \leq \frac{\epsilon}{2C_1}\right) \\ &= C_8 \frac{k_n}{v_n}. \quad \left(C_8 = C_8(x) \doteq \frac{C_2^2}{2C_7 C_5^2 \frac{\epsilon}{2C_1} + 4C_2 |T|}\right) \end{aligned}$$

Reemplazando esta desigualdad en (2.26) resulta que

$$P(A_n) \leq 2 \exp\left\{-C_8 \frac{k_n}{v_n}\right\}.$$

Finalmente, usando que por hipótesis $\sum_{n=1}^{\infty} \exp\left\{-a \frac{k_n}{v_n}\right\} < \infty$ para todo $a > 0$, concluimos que

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{N_4(x)-1} P(A_n) + \sum_{n=N_4(x)}^{\infty} P(A_n) \leq N_4(x) + 2 \sum_{n=N_4(x)}^{\infty} \exp\left\{-C_8 \frac{k_n}{v_n}\right\} < \infty.$$

□

Observación 2.10 . Supongamos que existen constantes F_1 y F_2 positivas tales que $F_1 \leq f_{\mathcal{X}}(x) \leq F_2$ para todo x . En este caso $C_1(x) = K\left(\frac{1}{2|T|f_{\mathcal{X}}(x)}\right) \leq K\left(\frac{1}{2|T|F_1}\right)$, $C_2(x) = \frac{\epsilon}{3f_{\mathcal{X}}(x)} \geq \frac{\epsilon}{3F_2}$ y $\frac{1}{2|T|F_2} \leq C_3(x) = \frac{1}{2|T|f_{\mathcal{X}}(x)} \leq \frac{1}{2|T|F_1}$ y por lo tanto C_4, C_5, C_6, C_7 , y C_8 no dependen de x . De esto se sigue que N_1, N_2 (y por lo tanto N_3 y N_4) tampoco dependen de x .

Teorema 2.11. Normalidad Asintótica. Supongamos que $H1$ y $H3$ se verifican y que la densidad $f_{\mathcal{X}}$ es estrictamente positiva y tiene dos derivadas acotadas. Elijamos una sucesión k_n de números reales positivos tal que $\sqrt{n}/k_n = o(1)$ y $k_n/n^{3/4} = o(1)$. Para ese k_n , supongamos también que $H5$ se cumple. Entonces, para todo $x \in \mathbb{R}$,

$$\sqrt{n}(\hat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)) \rightarrow \mathcal{N}\left(0, \left(\frac{2|T|}{C_0(x)}\right)^2\right).$$

Demostración. Por definición de convergencia en distribución tenemos que demostrar que $\forall a > 0$,

$$P\left(\sqrt{n}\left(\hat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right) \leq a\right) \rightarrow \Phi\left(a\frac{2|T|}{C_0}\right)$$

donde Φ es la Función de Distribución Acumulada Normal Standard. Sea

$$a_n \doteq \frac{k_n}{2n|T|\left(f_{\mathcal{X}}(x) + \frac{a}{\sqrt{n}}\right)}.$$

Dado que

$$\begin{aligned} \sqrt{n}\left(\hat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right) \leq a &\Leftrightarrow \sqrt{n}\left(\frac{k_n}{2n|T|h_n^{\mathcal{X}}(x)} - f_{\mathcal{X}}(x)\right) \leq a \\ &\Leftrightarrow \frac{k_n}{2n|T|h_n^{\mathcal{X}}(x)} - f_{\mathcal{X}}(x) \leq \frac{a}{\sqrt{n}} \\ &\Leftrightarrow \frac{k_n}{2n|T|h_n^{\mathcal{X}}(x)} \leq f_{\mathcal{X}}(x) + \frac{a}{\sqrt{n}} \\ &\Leftrightarrow h_n^{\mathcal{X}}(x) \geq \frac{k_n}{2n|T|\left(f_{\mathcal{X}}(x) + \frac{a}{\sqrt{n}}\right)} \\ &\Leftrightarrow h_n^{\mathcal{X}}(x) \geq a_n, \end{aligned}$$

entonces

$$P\left(\sqrt{n}\left(\hat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right) \leq a\right) = P\left(h_n^{\mathcal{X}}(x) \geq a_n\right).$$

Pero

$$h_n^{\mathcal{X}} \geq a_n \Leftrightarrow \sum_{i=1}^n \underbrace{\int_T \mathbb{I}_{I(x, a_n)}(\mathcal{X}_i(t)) dt}_{\doteq Y_i^{a_n}} < k_n$$

con $\{Y_i^{a_n}\}_{i=1}^n$ variables aleatorias independientes pues los procesos $\{\mathcal{X}_i(t)\}_{i=1}^n$ lo son. Por lo tanto

$$\begin{aligned}
P\left(\sqrt{n}\left(\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x)\right) \leq a\right) &= P\left(h_n^{\mathcal{X}}(x) \geq a_n\right) \\
&= P\left(\sum_{i=1}^n Y_i^{a_n} \leq k_n\right) \\
&= P(S_n \leq k_n) \\
&= P\left(\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}} \leq \frac{k_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}}\right)
\end{aligned} \tag{2.27}$$

$\left(S_n \doteq \sum_{i=1}^n Y_i^{a_n}\right)$

donde

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^n Y_i^{a_n}\right) = \sum_{i=1}^n \mathbb{E}(Y_i^{a_n}),$$

y dado que las $Y_i^{a_n}$ son independientes,

$$s_n^2 \doteq \text{var}(S_n) = \text{var}\left(\sum_{i=1}^n Y_i^{a_n}\right) = \sum_{i=1}^n \text{var}(Y_i^{a_n}).$$

Del Lema 2.9 con $c = a_n$ tenemos que

$$\begin{aligned}
s_n^2 &= \sum_{i=1}^n \text{var}(Y_i^{a_n}) \\
&= n \int_T \int_T \int_{\{u:|u-x|\leq a_n\}} \int_{\{v:|v-x|\leq a_n\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, du \, dv \, dt \, ds \\
&= (na_n^2) \frac{1}{a_n^2} \int_T \int_T \int_{\{u:|u-x|\leq a_n\}} \int_{\{v:|v-x|\leq a_n\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, du \, dv \, dt \, ds \\
&= C_1^2 \frac{k_n^2}{n} \frac{1}{a_n^2} \int_T \int_T \int_{\{u:|u-x|\leq a_n\}} \int_{\{v:|v-x|\leq a_n\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, du \, dv \, dt \, ds.
\end{aligned} \tag{2.28}$$

donde $C_1 = C_1(x) \doteq \frac{1}{2|T|(f_{\mathcal{X}}(x)+\epsilon)}$. Ahora, de la hipótesis H5 con $C = C_1$ tenemos

$$\lim_{n \rightarrow \infty} \frac{1}{a_n^2} \int_T \int_T \int_{\{x:|u-x|\leq a_n\}} \int_{\{x:|v-x|\leq a_n\}} (f_{st}(u,v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) \, du \, dv \, ds \, dt = C_0^2 < \infty,$$

además, por hipótesis $\frac{k_n^2}{n} \rightarrow \infty$, luego en (2.28) tenemos que $s_n^2 \rightarrow \infty$ y por lo tanto la condición de Lindeberg (A.5) se cumple luego la Observación (A.27) implica que

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1). \tag{2.29}$$

Resta ver que ocurre con $\frac{k_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}}$. Dado que $k_n = 2a_n n |T| \left(f_{\mathcal{X}}(x) + \frac{a}{\sqrt{n}} \right)$ y además $2a_n f_{\mathcal{X}}(x) = \int_{x-a_n}^{x+a_n} f_{\mathcal{X}}(u) du$ entonces,

$$\begin{aligned}
\frac{k_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}} &= \frac{k_n - n|T|p_{a_n}}{s_n} \\
&= \frac{2a_n n |T| \left(f_{\mathcal{X}}(x) + \frac{a}{\sqrt{n}} \right) - n|T| \int_{\{x: |u-x| \leq a_n\}} f_{\mathcal{X}}(u) du}{s_n} \\
&= \frac{n|T| \int_{x-a_n}^{x+a_n} f_{\mathcal{X}}(u) du + 2a_n n |T| \frac{a}{\sqrt{n}} - n|T| \int_{x-a_n}^{x+a_n} f_{\mathcal{X}}(u) du}{s_n} \quad (2.30) \\
&= s_n^{-1} n |T| \int_{x-a_n}^{x+a_n} (f_{\mathcal{X}}(x) - f_{\mathcal{X}}(u)) du + \frac{2a_n n |T| a}{s_n \sqrt{n}} \\
&= s_n^{-1} n |T| \int_{x-a_n}^{x+a_n} (f_{\mathcal{X}}(x) - f_{\mathcal{X}}(u)) du + 2|T| a \frac{a_n \sqrt{n}}{s_n}.
\end{aligned}$$

Por el teorema de Taylor, existe un número x^* entre u y x tal que

$$f_{\mathcal{X}}(x) - f_{\mathcal{X}}(u) = -f'_{\mathcal{X}}(x)(u-x) - \frac{1}{2} f''_{\mathcal{X}}(x^*)(u-x^*)^2.$$

Luego,

$$\begin{aligned}
\int_{x-a_n}^{x+a_n} (f_{\mathcal{X}}(x) - f_{\mathcal{X}}(u)) du &= -f'_{\mathcal{X}}(x) \int_{x-a_n}^{x+a_n} (u-x) du - \frac{1}{2} \int_{x-a_n}^{x+a_n} f''_{\mathcal{X}}(x^*)(u-x^*)^2 du \\
&= -\frac{1}{2} \int_{x-a_n}^{x+a_n} f''_{\mathcal{X}}(x^*)(u-x^*)^2 du
\end{aligned}$$

por lo que,

$$\begin{aligned}
\left| s_n^{-1} n |T| \int_{x-a_n}^{x+a_n} (f_{\mathcal{X}}(x) - f_{\mathcal{X}}(u)) du \right| &= \left| -s_n^{-1} n |T| \frac{1}{2} \int_{x-a_n}^{x+a_n} f''_{\mathcal{X}}(x^*)(u-x^*)^2 du \right| \\
&\leq s_n^{-1} n \frac{|T|}{2} \int_{x-a_n}^{x+a_n} |f''_{\mathcal{X}}(x^*)| |(u-x^*)^2| du \\
&\leq s_n^{-1} n \frac{|T| C_1}{2} \int_{x-a_n}^{x+a_n} (u-x^*)^2 du \quad (|f''_{\mathcal{X}}(x)| \leq C_1, \forall x) \\
&\leq s_n^{-1} n \frac{|T| C_1}{2} \int_{x-a_n}^{x+a_n} (2a_n)^2 du \quad (x^* \in (x, u)) \\
&= 4C_1 s_n^{-1} n a_n^3.
\end{aligned}$$

Luego en (2.30) tenemos que

$$\begin{aligned}
\frac{k_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}} &= O\left(a_n^3 s_n^{-1} n |T|\right) + 2|T| a \frac{a_n \sqrt{n}}{s_n} \\
&\leq 4C_1 \frac{k_n^2}{4n^2 |T|^2 f_e^2(x)} \sqrt{n} |T| + 2|T| a \frac{a_n \sqrt{n}}{s_n} \quad \left(a_n \leq \frac{k_n}{2n|T| f_{\mathcal{X}}(x)} \right) \quad (2.31) \\
&= \frac{C_1}{|T|} \frac{k_n^2}{n^{3/2}} + 2|T| a \frac{a_n \sqrt{n}}{s_n}.
\end{aligned}$$

Finalmente, de (2.28) resulta que

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{s_n^2}{na_n^2} &= \lim_{n \rightarrow \infty} \frac{1}{a_n^2} \int_T \int_T \int_{\{u: |u-x| \leq a_n\}} \int_{\{v: |v-x| \leq a_n\}} (f_{st}(u, v) - f_{\mathcal{X}}(u)f_{\mathcal{X}}(v)) du dv dt ds \\ &= C_0^2, \end{aligned}$$

de donde se sigue que $\frac{a_n \sqrt{n}}{s_n} \rightarrow \frac{1}{C_0}$. Además por hipótesis $\frac{k_n^2}{n^{3/2}} \rightarrow 0$, en (2.31) tenemos que

$$\lim_{n \rightarrow \infty} \frac{k_n - \mathbb{E}(S_n)}{\sqrt{\text{var}(S_n)}} = \frac{2|T|a}{C_0}.$$

Usando esta última expresión junto con (2.27) y (2.29) resulta

$$\sqrt{n} \left(\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \left(\frac{2|T|}{C_0} \right)^2 \right).$$

□

Teorema 2.12. Velocidades de convergencia uniforme. *Sea M compacto. Supongamos que $H1$ – $H3$ se cumplen. Elijamos dos sucesiones $\{k_n\}$ y $\{v_n\}$ de números reales positivos que tienden a infinito tales que $k_n = n^\beta$, $\beta < 1$, $v_n \left(\frac{k_n}{n}\right) = o(1)$ y $\frac{k_n}{v_n \log n} \rightarrow \infty$. Para ese k_n supongamos que $H5$ se verifica y que además $f_{\mathcal{X}}$ es estrictamente positiva. Entonces,*

$$\sup_{x \in M} v_n \left| \widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right| \xrightarrow{a.co.} 0.$$

Observación 2.13. Nuestras hipótesis implican que k_n puede elegirse de manera que $v_n = \frac{n^\gamma}{\log n}$ con $\gamma < 1/2$. Más precisamente, sean $k_n = n^\beta$ y $v_n = \frac{n^\gamma}{\log n}$, $\gamma < 1/2$. Para que $v_n \left(\frac{k_n}{n}\right) = \frac{n^{\gamma+\beta-1}}{\log n} \rightarrow 0$ y $\frac{k_n}{v_n \log n} = \frac{n^\beta}{\frac{n^\gamma}{\log n} \log n} = n^{\beta-\gamma} \rightarrow \infty$ se verifiquen, debe ser $\gamma + \beta - 1 \leq 0$ y $\beta - \gamma > 0$ o equivalentemente, $\beta \leq 1 - \gamma$ y $\beta > \gamma$. Por lo tanto, dado $\gamma < \frac{1}{2}$ siempre podemos elegir β de manera que verifique estas hipótesis (ver Figura 2).

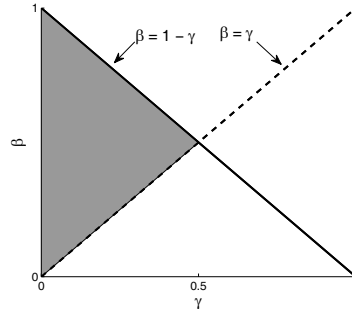


FIGURA 2. El área sombreada muestra la intersección de las regiones $\beta \leq 1 - \gamma$ y $\beta > \gamma$.

Antes de comenzar con la demostración necesitamos el siguiente lema.

Lema 2.14. Para cada x, y , para cada n se tiene que

$$|h_n^{\mathcal{X}}(x) - h_n^{\mathcal{X}}(y)| \leq |x - y|.$$

Demostración. Sean x, y , y t fijos. Observemos que si $|\mathcal{X}_i(t) - x| \leq h_n^{\mathcal{X}}(x)$ entonces

$$|\mathcal{X}_i(t) - y| \leq |\mathcal{X}_i(t) - x| + |x - y| \leq h_n^{\mathcal{X}}(x) + |x - y|.$$

Es decir,

$$\{t : |\mathcal{X}_i(t) - x| \leq h_n^{\mathcal{X}}(x)\} \subset \{t : |\mathcal{X}_i(t) - y| \leq h_n^{\mathcal{X}}(x) + |x - y|\}$$

de donde se sigue que

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{\{t: |\mathcal{X}_i(t) - x| \leq h_n^{\mathcal{X}}(x)\}}(\mathcal{X}_i(t)) dt \leq \sum_{i=1}^n \int_T \mathbb{I}_{\{t: |\mathcal{X}_i(t) - y| \leq h_n^{\mathcal{X}}(x) + |x - y|\}}(\mathcal{X}_i(t)) dt. \quad (2.32)$$

Por otro lado, por definición de k_n ,

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{\{t: |\mathcal{X}_i(t) - y| \leq h_n^{\mathcal{X}}(y)\}}(\mathcal{X}_i(t)) dt. \quad (2.33)$$

Luego, de (2.32) y (2.33) resulta que

$$h_n^{\mathcal{X}}(y) \leq h_n^{\mathcal{X}}(x) + |x - y|. \quad (2.34)$$

Análogamente podemos probar que

$$h_n^{\mathcal{X}}(x) \leq h_n^{\mathcal{X}}(y) + |y - x| \quad (2.35)$$

por lo que, de (2.34) y (2.35) tenemos el resultado. \square

Demostración del Teorema 2.12: Por definición de convergencia casi completa necesitamos probar que si

$$C_n = \left\{ \sup_{x \in M} v_n \left| \widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right| > \epsilon \right\},$$

entonces para todo $\epsilon > 0$, $\sum_{n=1}^{\infty} P(C_n) < \infty$.

Sea $\epsilon > 0$, dado que M es compacto podemos considerar una colección finita de intervalos que lo cubren. Es decir, sea $\{I_j\}_{j=1}^N$, $I_j = I(x_j, b_n)$ de centro x_j y radios b_n con $b_n = \left(\frac{k_n}{n}\right)^2$, tales que $M \subset \bigcup_{j=1}^N I(x_j, b_n)$. Luego, $N = O(b_n^{-1})$.

Para cada $x \in M$, existe un j tal que $x \in I(x_j, b_n)$ para algún j . Luego, para ese j podemos escribir

$$v_n \left| \widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right| \leq v_n \left| \widehat{f}_{\mathcal{X}}(x) - \widehat{f}_{\mathcal{X}}(x_j) \right| + v_n \left| \widehat{f}_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x_j) \right| + v_n |f_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x)|. \quad (2.36)$$

Dado que f es Lipschitz y que $x \in I(x_j, b_n)$,

$$v_n |f_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x)| \leq K v_n |x_j - x| \leq K v_n b_n = K v_n \left(\frac{k_n}{n}\right)^2.$$

Ahora, dado que $\frac{k_n}{n} \rightarrow 0$ existe N_0 tal que si $n \geq N_0$ $\frac{k_n}{n} < 1$. Además, dado que $v_n \left(\frac{k_n}{n}\right) \rightarrow 0$ existe N_1 tal que si $n \geq N_1$, $v_n \left(\frac{k_n}{n}\right) < \frac{\epsilon}{3K}$. Por lo tanto, si $n \geq \max\{N_0, N_1\}$ entonces, $v_n \left(\frac{k_n}{n}\right)^2 < \frac{\epsilon}{3K}$ de donde se sigue que

$$v_n |f_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x)| < \epsilon/3.$$

Con esta desigualdad en (2.36) y tomando supremo obtenemos

$$\sup_{x \in M} v_n \left| \widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right| \leq \max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \widehat{f}_{\mathcal{X}}(x) - \widehat{f}_{\mathcal{X}}(x_j) \right| + \max_{1 \leq j \leq N} v_n \left| \widehat{f}_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x_j) \right| + \epsilon/3.$$

Luego, para probar que $\sum_{n=1}^{\infty} P(C_n) < \infty$, es suficiente probar que $\sum_{n=1}^{\infty} P(A_n) < \infty$ y $\sum_{n=1}^{\infty} P(B_n) < \infty$ donde

$$A_n = \left\{ \max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \widehat{f}_{\mathcal{X}}(x) - \widehat{f}_{\mathcal{X}}(x_j) \right| > \epsilon/3 \right\}$$

y

$$B_n = \left\{ \max_{1 \leq j \leq N} v_n \left| \widehat{f}_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x_j) \right| > \epsilon/3 \right\}.$$

Comencemos probando que $\sum_{n=1}^{\infty} P(B_n) < \infty$. Dado que

$$P(B_n) \leq N \max_{1 \leq j \leq N} P\left(v_n \left| \widehat{f}_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x_j) \right| > \epsilon/3\right), \quad (2.37)$$

deberíamos usar el Teorema 2.7 para encontrar una cota de $P\left(v_n \left| \widehat{f}_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x_j) \right| > \epsilon/3\right)$ que no dependa de x_j . Para ello primero observemos que dado que por H2 $f_{\mathcal{X}}$ es Lipschitz, es uniformemente continua y dado que K es compacto, entonces $f_{\mathcal{X}}$ es acotada en K . Es decir, existen constantes F_1 y F_2 positivas (pues por hipótesis $f_{\mathcal{X}}$ es estrictamente positiva) tales que $0 < F_1 \leq f_{\mathcal{X}}(x_j) \leq F_2$ para todo x_j . Por lo tanto, del Teorema 2.7 y de la Observación 2.10, para todo $n \geq N_2$ tenemos que,

$$P\left(v_n \left| \widehat{f}_{\mathcal{X}}(x_j) - f_{\mathcal{X}}(x_j) \right| > \epsilon/3\right) \leq 2 \exp\left(-C_1 \frac{k_n}{v_n}\right) = 2n^{-C_1 \frac{k_n}{v_n \log n}}, \quad (2.38)$$

con C_1 y N_2 que no dependen de x_j . Además, dado que $N = O(b_n^{-1})$, existen C_2 y N_3 tales que para todo $n \geq N_3$,

$$N \leq C_2 b_n^{-1} = C_2 \left(\frac{k_n}{n}\right)^{-2}. \quad (2.39)$$

Luego, con (2.38) y (2.39) en (2.37), para todo $n \geq N_4 = \max\{N_2, N_3\}$ tenemos

$$\begin{aligned} P(B_n) &\leq C_2 \left(\frac{k_n}{n}\right)^{-2} 2n^{-C_1 \frac{k_n}{v_n \log n}} \\ &= 2C_2 k_n^{-2} n^{2-C_1 \frac{k_n}{v_n \log n}} \\ &= 2C_2 n^{-2\beta+2-C_1 \frac{k_n}{v_n \log n}}. \end{aligned} \quad \left(k_n = n^\beta, \beta < 1\right)$$

Sea C_3 una constante tal que $C_3 > \frac{-2\beta+3}{C_1} > 0$. Como por hipótesis $\frac{k_n}{v_n \log n} \rightarrow \infty$, existe N_5 tal que si $n \geq N_5$, $\frac{k_n}{v_n \log n} > C_3$. Luego, para todo $n \geq N_5$,

$$P(B_n) \leq C_2 n^{-2\beta+2-C_1 C_3}.$$

Como hemos elegido C_3 de manera que $C_3 > \frac{-2\beta+3}{C_1}$ entonces, $-2\beta+2-C_1 C_3 < -1$. En efecto,

$$C_3 > \frac{-2\beta+3}{C_1} \Rightarrow C_1 C_3 > -2\beta+3 \Rightarrow -1 > -2\beta+2-C_1 C_3.$$

Por lo tanto, si $N_6 = \max\{N_4, N_5\}$,

$$\sum_{n=N_6}^{\infty} P(B_n) \leq C_2 \sum_{n=N_6}^{\infty} n^{-2\beta+2-C_1 C_3} < \infty.$$

Finalmente,

$$\sum_{n=1}^{\infty} P(B_n) \leq N_6 + 2C_2 \sum_{n=N_6}^{\infty} n^{-2\beta+2-C_1 C_3} < \infty. \quad (2.40)$$

Resta probar entonces que $\sum_{n=1}^{\infty} P(A_n) < \infty$. Dado que $C_n = A_n \cup B_n = (A_n \cap B_n^c) \cup B_n$ y $\sum_{n=1}^{\infty} P(B_n) < \infty$, es suficiente probar que $\sum_{n=1}^{\infty} P(A_n \cap B_n^c) < \infty$. En B_n^c escribimos

$$\begin{aligned} \max_{1 \leq j \leq N} |\widehat{f_{\mathcal{X}}}(x_j)| &\leq \max_{1 \leq j \leq N} |\widehat{f_{\mathcal{X}}}(x_j) - f_{\mathcal{X}}(x_j)| + \max_{1 \leq j \leq N} |f_{\mathcal{X}}(x_j)| \\ &\leq \epsilon/3 + \max_{1 \leq j \leq N} |f_{\mathcal{X}}(x_j)| \\ &\leq \epsilon/3 + F \doteq C_4. \end{aligned} \quad (2.41)$$

Luego en A_n ,

$$\begin{aligned}
\max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \widehat{f_{\mathcal{X}}}(x) - \widehat{f_{\mathcal{X}}}(x_j) \right| &= \max_{1 \leq j \leq N} \left| \widehat{f_{\mathcal{X}}}(x_j) \right| \max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \frac{\widehat{f_{\mathcal{X}}}(x)}{\widehat{f_{\mathcal{X}}}(x_j)} - 1 \right| \\
&= \max_{1 \leq j \leq N} \left| \widehat{f_{\mathcal{X}}}(x_j) \right| \max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \frac{h_n^{\mathcal{X}}(x_j)}{h_n^{\mathcal{X}}(x)} - 1 \right| \quad (2.42) \\
&\leq C_4 \max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \frac{h_n^{\mathcal{X}}(x_j)}{h_n^{\mathcal{X}}(x)} - 1 \right| \quad (\text{de (2.41)}) \\
&\leq C_4 \max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \frac{|x - x_j|}{h_n^{\mathcal{X}}(x)} \quad (\text{Lema 2.14}) \\
&\leq C_4 \max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \frac{b_n}{h_n^{\mathcal{X}}(x)} \quad (x \in I_j) \\
&= C_4 v_n \left(\frac{k_n}{n} \right)^2 \max_{1 \leq j \leq N} \sup_{x \in I_j} \frac{1}{h_n^{\mathcal{X}}(x)}. \quad \left(b_n = \left(\frac{k_n}{n} \right)^2 \right)
\end{aligned}$$

Ahora,

$$\begin{aligned}
h_n^{\mathcal{X}}(x) &\geq h_n^{\mathcal{X}}(x_j) - |x - x_j| \quad (\text{Lema 2.14}) \\
&\geq h_n^{\mathcal{X}}(x_j) - b_n \quad (x \in I_j \Rightarrow |x - x_j| \leq b_n) \\
&= h_n^{\mathcal{X}}(x_j) - \left(\frac{k_n}{n} \right)^2 \quad \left(b_n = \left(\frac{k_n}{n} \right)^2 \right) \\
&= \frac{k_n}{2n|T|\widehat{f_{\mathcal{X}}}(x_j)} - \left(\frac{k_n}{n} \right)^2 \quad (\text{Def. de } \widehat{f_{\mathcal{X}}}) \\
&\geq \frac{k_n}{2n|T|C_4} - \left(\frac{k_n}{n} \right)^2 \quad \left(\text{de (2.41), ya que } \widehat{f_{\mathcal{X}}}(x_j) \leq \max_{1 \leq j \leq N} \left| \widehat{f_{\mathcal{X}}}(x_j) \right| \right) \\
&= \frac{k_n}{n} \left(\frac{1}{2|T|C_4} - \frac{k_n}{n} \right) \\
&> \frac{k_n}{n} C_5,
\end{aligned}$$

donde en la última desigualdad hemos usado que, como $\frac{k_n}{n} \rightarrow 0$ existe N_7 tal que para todo $n \geq N_7$, $\frac{k_n}{n} < \frac{1}{2|T|C_4}$. Aquí C_5 es una constante menor que $\frac{1}{2|T|C_4}$. Luego

$$\max_{1 \leq j \leq N} \sup_{x \in I_j} \frac{1}{h_n^{\mathcal{X}}(x)} \leq C_5^{-1} \left(\frac{k_n}{n} \right)^{-1}.$$

Con esta desigualdad en (2.42) resulta

$$\max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \widehat{f_{\mathcal{X}}}(x) - \widehat{f_{\mathcal{X}}}(x_j) \right| \leq C_4 C_5^{-1} v_n \left(\frac{k_n}{n} \right) < \epsilon/3,$$

donde en la última desigualdad hemos usado que, como $v_n \left(\frac{k_n}{n}\right) \rightarrow 0$ existe N_8 tal que para todo $n \geq N_8$, $v_n \left(\frac{k_n}{n}\right) < \frac{\epsilon}{3C_4C_5^{-1}}$. Luego, para todo $n \geq N_8$ tenemos que

$$P(A_n \cap B_n^c) \leq P\left(\max_{1 \leq j \leq N} \sup_{x \in I_j} v_n \left| \widehat{f}_{\mathcal{X}}(x) - \widehat{f}_{\mathcal{X}}(x_j) \right| > \epsilon/3\right) = 0.$$

Lo que implica que

$$\sum_{n=1}^{\infty} P(A_n \cap B_n^c) = \sum_{n=1}^{N_8-1} P(A_n \cap B_n^c) \leq N_8 < \infty. \quad (2.43)$$

Finalmente, de (2.40) y (2.43) tenemos el resultado. □

2.2. Estimando la densidad de un proceso no estacionario

En esta Sección extendemos el estimador dado en la Sección 2.1 a una familia particular de procesos estocásticos no estacionarios y estudiamos sus propiedades asintóticas. Para ello, supongamos que en el modelo (2.2) la función media $\mu(t)$ es no constante con respecto al tiempo, es decir, supongamos que $\mathcal{X}(t)$ es dado por

$$\mathcal{X}(t) = \mu(t) + e(t),$$

donde $\mu(t)$ es una la función media (determinística) y $e(t)$ es un proceso estocástico estacionario de primer orden, con media cero y función de densidad no conocida f_e . La función de densidad de $\mathcal{X}(t)$ será denotada por $f_{\mathcal{X}(t)} \doteq f_{\mathcal{X}_t}$.

2.2.1. Extensión del estimador

Sean $\{\mathcal{X}_1(t), \dots, \mathcal{X}_n(t)\}$ trayectorias independientes con la misma distribución que $\mathcal{X}(t)$. Definimos el estimador de la función de densidad $f_{\mathcal{X}_t}$ como sigue

$$\widehat{f}_{\mathcal{X}_t}(x) = \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)), \quad (2.44)$$

donde

$$\widehat{f}_u(x) \doteq \frac{k_n}{2n|T|h_n^u(x)},$$

con $u = \{\mathcal{U}_{n1}, \dots, \mathcal{U}_{nn}\}$ dado por

$$\mathcal{U}_{ni}(t) = \mathcal{X}_i(t) - \bar{\mathcal{X}}_n(t) = e_i(t) - \bar{e}_n(t). \quad (2.45)$$

Aquí $\{e_1(t), \dots, e_n(t)\}$ es una muestra aleatoria de $e(t)$, $\bar{e}_n(t) = \frac{1}{n} \sum_{i=1}^n e_i(t)$ y h_n^u es definida como en (2.3) reemplazando $\{e_i(t)\}_{i=1}^n$ por u .

Para todo t fijo, las variables aleatorias $\{\mathcal{U}_{n1}(t), \dots, \mathcal{U}_{nn}(t)\}$ son idénticamente distribuidas con $\mathbb{E}(\mathcal{U}_{ni}(t)) = 0$ pero no necesariamente independientes. Por lo tanto, no podemos usar directamente los resultados de la sección 2.1. Sin embargo, todavía podemos probar la convergencia completa del estimador de $f_{\mathcal{X}_t}$ y obtener velocidades de convergencia.

2.2.2. Resultados asintóticos

Teorema 2.15. Velocidades de convergencia. Para t fijo, supongamos que $H1$, $H2$ y $H4$ se verifican para $e(t)$ y que f_e es estrictamente positiva. Elijamos dos sucesiones $\{k_n\}$ y $\{v_n\}$ de números reales positivos que tienden a infinito tales que $(k_n/n)v_n = o(1)$, $\sum_{n=1}^{\infty} \exp(-a(k_n/v_n)) < \infty$ para cada $a > 0$ y $v_n(n/k_n)|\bar{e}_n(t)| \rightarrow 0$ a.co. Supongamos que la sucesión k_n verifica $H5$. Entonces, para todo $x \in \mathbb{R}$,

$$\lim_{n \rightarrow \infty} v_n \left(\widehat{f}_{\mathcal{X}_t}(x) - f_{\mathcal{X}_t}(x) \right) = 0 \quad \text{a.co.}$$

Observación 2.16. En este caso, nuestras hipótesis implican que podemos elegir k_n tal que $v_n = n^\gamma$ para todo $\gamma < \frac{1}{4}$. Más precisamente, sean $k_n = n^\beta$, $v_n = n^\gamma$ y t fijo. Si $\text{var}(e(t)) < \infty$, del Teorema central del límite sabemos que $\bar{e}_n(t) = o(n^{-\alpha})$ con $\alpha < 1/2$. Para que las condiciones $v_n \frac{k_n}{n} \rightarrow 0$, $\frac{k_n}{v_n} \rightarrow \infty$ y $v_n \frac{n}{k_n} |\bar{e}_n(t)| \rightarrow 0$ se cumplan, es suficiente que $\gamma + \beta - 1 < 0$ y $\gamma + 1 - \beta - \alpha < 0$ con $\alpha < 1/2$, o equivalentemente $\beta < 1 - \gamma$ y $\beta > \gamma + \frac{1}{2}$ y así, dado $\gamma < \frac{1}{4}$ siempre podemos elegir β de manera que las hipótesis se cumplan (ver Figura 3).

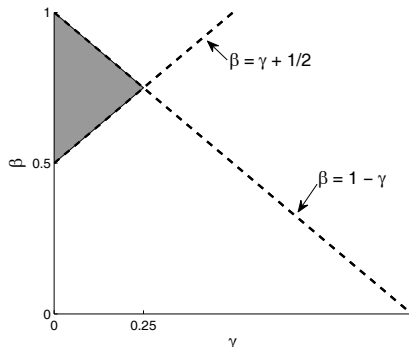


FIGURA 3. El área sombreada muestra la intersección de las regiones $\beta < 1 - \gamma$ y $\beta > \gamma + \frac{1}{2}$.

La demostración de este teorema será consecuencia inmediata del Teorema 2.7 y del siguiente lema:

Lema 2.17. *Supongamos que H1-H3 se verifican para $e(t)$ y que $v_n(n/k_n)|\bar{e}_n(t)| = o(1)$ a.co. para todo t fijo. Entonces, para todo $x \in \mathbb{R}$,*

$$\lim_{n \rightarrow \infty} v_n \left(\widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - \widehat{f}_e(x - \mu(t)) \right) = 0, \quad a.co.$$

donde \widehat{f}_e es el estimador de f_e .

Para probar este lema necesitamos un resultado auxiliar.

Lema 2.18. *Para t fijo, sean h_n^u y $\bar{e}_n(t)$ como en (2.45) donde $u = \{\mathcal{U}_{n1}, \dots, \mathcal{U}_{nn}\}$ con $\mathcal{U}_{ni}(t) = \mathcal{X}_i(t) - \bar{\mathcal{X}}_n(t) = e_i(t) - \bar{e}_n(t)$ y h_n^e como definida en (2.3) para $e(t)$. Entonces, para cada n y x se tiene que,*

$$|h_n^u(x - \bar{\mathcal{X}}_n(t)) - h_n^e(x - \mu(t))| \leq 2|\bar{e}_n(t)|.$$

Demostración del Lema 2.18: Es consecuencia inmediata de las siguientes desigualdades:

$$(i) \quad |h_n^u(x - \bar{\mathcal{X}}_n(t)) - h_n^u(x - \mu(t))| \leq |\bar{e}_n(t)|;$$

$$(ii) \quad |h_n^u(x - \mu(t)) - h_n^e(x - \mu(t))| \leq |\bar{e}_n(t)|.$$

(i) Sean x y t fijos. Tomando $e = u$, $x = x - \bar{\mathcal{X}}_n(t)$ e $y = x - \mu(t)$ en el Lema 2.14 la parte

(i) queda demostrada ya que

$$|h_n^u(x - \bar{\mathcal{X}}_n(t)) - h_n^u(x - \mu(t))| \leq |(x - \bar{\mathcal{X}}_n(t)) - (x - \mu(t))| = |\bar{\mathcal{X}}_n(t) - \mu(t)| = |\bar{e}_n(t)|.$$

(ii) Sean x y t fijos. Observemos que si $|\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t))$ entonces

$$\begin{aligned} |e_i(t) - (x - \mu(t))| &= |(\mathcal{U}_{ni}(t) + \bar{e}_n(t)) - (x - \mu(t))| && \text{(de (2.45))} \\ &\leq |\mathcal{U}_{ni}(t) - (x - \mu(t))| + |\bar{e}_n(t)| \\ &\leq h_n^u(x - \mu(t)) + |\bar{e}_n(t)|. \end{aligned}$$

Por lo tanto,

$$\{t : |\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t))\} \subset \{t : |e_i(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t)) + |\bar{e}_n(t)|\},$$

de donde se sigue que

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{\{|\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t))\}}(t) dt \leq \sum_{i=1}^n \int_T \mathbb{I}_{\{|e_i(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t)) + |\bar{e}_n(t)|\}}(t) dt.$$

Por otro lado, por definición de k_n ,

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{\{|e_i(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t))\}}(t) dt$$

luego,

$$\sum_{i=1}^n \int_T \mathbb{I}_{\{|e_i(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t))\}}(t) dt \leq \sum_{i=1}^n \int_T \mathbb{I}_{\{|e_i(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t)) + |\bar{e}_n(t)|\}}(t) dt,$$

de donde se sigue que

$$h_n^e(x - \mu(t)) \leq h_n^u(x - \mu(t)) + |\bar{e}_n(t)|. \quad (2.46)$$

Similarmente, observemos que si $|e_i(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t))$ entonces

$$\begin{aligned} |\mathcal{U}_{ni}(t) - (x - \mu(t))| &= |(e_i(t) - \bar{e}_n(t)) - (x - \mu(t))| && \text{(de (2.45))} \\ &\leq |e_i(t) - (x - \mu(t))| + |\bar{e}_n(t)| \\ &\leq h_n^e(x - \mu(t)) + |\bar{e}_n(t)|, \end{aligned}$$

de manera que

$$\{t : |e_i(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t))\} \subset \{t : |\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t)) + |\bar{e}_n(t)|\},$$

por lo tanto

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{\{|e_i(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t))\}}(t) dt \leq \sum_{i=1}^n \int_T \mathbb{I}_{\{|\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t)) + |\bar{e}_n(t)|\}}(t) dt.$$

Pero por definición de k_n ,

$$k_n = \sum_{i=1}^n \int_T \mathbb{I}_{\{|\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t))\}}(t) dt,$$

luego

$$\sum_{i=1}^n \int_T \mathbb{I}_{\{|\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^u(x - \mu(t))\}}(t) dt \leq \sum_{i=1}^n \int_T \mathbb{I}_{\{|\mathcal{U}_{ni}(t) - (x - \mu(t))| \leq h_n^e(x - \mu(t)) + |\bar{e}_n(t)|\}}(t) dt,$$

de donde se sigue que

$$h_n^u(x - \mu(t)) \leq h_n^e(x - \mu(t)) + |\bar{e}_n(t)|. \quad (2.47)$$

Finalmente, de (2.46) y (2.47) tenemos el resultado. □

Demostración del Lema 2.17: Sea $\epsilon > 0$ y sean x, t fijos.

$$\begin{aligned} v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - \widehat{f}_e(x - \mu(t)) \right| &= \frac{k_n v_n}{2n|T|} \frac{|h_n^e(x - \mu(t)) - h_n^u(x - \bar{\mathcal{X}}_n(t))|}{h_n^u(x - \bar{\mathcal{X}}_n(t)) h_n^e(x - \mu(t))} \\ &\leq \frac{k_n v_n}{2n|T|} \frac{2|\bar{e}_n(t)|}{h_n^u(x - \bar{\mathcal{X}}_n(t)) h_n^e(x - \mu(t))}. \end{aligned} \quad (2.48) \quad \text{(Lema 2.18)}$$

Aplicando el Teorema 2.4 a $e(t)$ tenemos que existe $N_1 = N_1(x)$ tal que si $n \geq N_1$, $\widehat{f}_e(x - \mu(t)) \leq f_e(x - \mu(t)) + \epsilon$. Para este N_1 escribimos

$$h_n^e(x - \mu(t)) = \frac{k_n}{2n|T|\widehat{f}_e(x - \mu(t))} \geq \frac{1}{2|T|(f_e(x - \mu(t)) + \epsilon)} \frac{k_n}{n} \doteq C_1(x) \frac{k_n}{n}. \quad (2.49)$$

Por otro lado, del Lema 2.18 tenemos que $h_n^e(x - \mu(t)) \leq h_n^u(x - \bar{\mathcal{X}}_n(t)) + 2|\bar{e}_n(t)|$, luego

$$h_n^u(x - \bar{\mathcal{X}}_n(t)) + 2|\bar{e}_n(t)| \geq h_n^e(x - \mu(t)) \geq C_1 \frac{k_n}{n}$$

o equivalentemente,

$$h_n^u(x - \bar{\mathcal{X}}_n(t)) \geq C_1 \frac{k_n}{n} - 2|\bar{e}_n(t)|. \quad (2.50)$$

Ahora, dado que $v_n \frac{n}{k_n} |\bar{e}_n(t)| \xrightarrow{a.co.} 0$ y $v_n \rightarrow \infty$ entonces, $\frac{n}{k_n} |\bar{e}_n(t)| \xrightarrow{a.co.} 0$. Por lo tanto, existe N_2 tal que si $n \geq N_2$,

$$\frac{n}{k_n} |\bar{e}_n(t)| \leq \frac{1}{4} C_1.$$

Con esta desigualdad en (2.50) resulta,

$$h_n^u(x - \bar{\mathcal{X}}_n(t)) \geq C_1 \frac{k_n}{n} - 2|\bar{e}_n(t)| \geq \frac{1}{2} C_1 \frac{k_n}{n}. \quad (2.51)$$

Finalmente, con (2.49) y (2.51) en (2.48), para $n \geq N_3 = \max\{N_1, N_2\}$ tenemos que,

$$\begin{aligned} v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - \widehat{f}_e(x - \mu(t)) \right| &\leq \frac{1}{|T|} \frac{k_n v_n}{n} \frac{|\bar{e}_n(t)|}{\frac{1}{2} C_1 \frac{k_n}{n} C_1 \frac{k_n}{n}} \\ &= \frac{2}{|T| C_1^2} v_n \frac{n}{k_n} |\bar{e}_n(t)| \\ &\doteq C_2 v_n \frac{n}{k_n} |\bar{e}_n(t)|. \quad \left(C_2 = \frac{2}{|T| C_1^2} \right) \end{aligned}$$

Luego

$$\begin{aligned} \sum_{n=1}^{\infty} P\left(v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - \widehat{f}_e(x - \mu(t)) \right| \geq \epsilon\right) &= \sum_{n=1}^{N_3-1} P\left(v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - \widehat{f}_e(x - \mu(t)) \right| \geq \epsilon\right) \\ &\quad + \sum_{n=N_3}^{\infty} P\left(v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - \widehat{f}_e(x - \mu(t)) \right| \geq \epsilon\right) \\ &\leq N_3 + \sum_{n=N_3}^{\infty} P\left(C_2 v_n \frac{n}{k_n} |\bar{e}_n(t)| \geq \epsilon\right) \\ &\leq N_3 + \sum_{n=1}^{\infty} P\left(v_n \frac{n}{k_n} |\bar{e}_n(t)| \geq \frac{\epsilon}{C_2}\right) < \infty \end{aligned}$$

pues por hipótesis, $v_n \frac{n}{k_n} |\bar{e}_n(t)| \xrightarrow{a.co.} 0$.

□

Demostración del Teorema 2.15: Dado que $\mathcal{X}(t)$ satisface el modelo (2.2), $f_{\mathcal{X}_t}(x) = f_e(x - \mu(t))$. Además, de la definición (2.44), $\widehat{f}_{\mathcal{X}_t}(x) = \widehat{f}_u(x - \bar{\mathcal{X}}_n(t))$. Luego, para probar el teorema es suficiente probar que

$$\lim_{n \rightarrow \infty} v_n \left(\widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - f_e(x - \mu(t)) \right) = 0, \quad a.co.$$

Para ello sean $\epsilon > 0$, x, t fijos. Dado que,

$$\begin{aligned} \sum_{n=1}^{\infty} P \left(v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - f_e(x - \mu(t)) \right| > \epsilon \right) &\leq \sum_{n=1}^{\infty} P \left(v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - \widehat{f}_e(x - \mu(t)) \right| > \epsilon/2 \right) \\ &+ \sum_{n=1}^{\infty} P \left(v_n \left| \widehat{f}_e(x - \mu(t)) - f_e(x - \mu(t)) \right| > \epsilon/2 \right), \end{aligned}$$

del Lema 2.17 y del Teorema 2.7 aplicado a $e(t)$ se sigue que

$$\sum_{n=1}^{\infty} P \left(v_n \left| \widehat{f}_u(x - \bar{\mathcal{X}}_n(t)) - f_e(x - \mu(t)) \right| > \epsilon \right) < \infty.$$

□

2.3. Una nueva regla de clasificación para datos funcionales

Sea $\{(\mathcal{X}_i(t), Y_i)\}_{i=1}^n$ una muestra aleatoria de $(\mathcal{X}(t), Y) \in \mathcal{F} \times \bar{G} = \{1, \dots, G\}$. El objetivo de esta sección es, dada una nueva curva $\mathcal{X}(t) \in \mathcal{F}$, asignarle una clase g de \bar{G} . Para ello, supondremos que para cada clase g el modelo (2.2) se cumple. Esto es,

$$\mathcal{X}^g(t) = \mu^g(t) + e^g(t), \quad g \in \bar{G},$$

donde $\mu^g(t)$ es la función media de la clase g y $e^g(t)$ es un proceso estocástico estacionario con función de densidad desconocida f_e^g .

Recordemos que la Regla de Bayes elige como clase aquella para la cual la probabilidad a posteriori es máxima. Por lo tanto, para cada t fijo, la Regla de Bayes asignará a \mathcal{X} la clase g_0 si y sólo si $\widehat{f}_{\mathcal{X}_t}^{g_0}(\mathcal{X}(t)) = \arg \max_{g \in \bar{G}} \widehat{f}_{\mathcal{X}_t}^g(\mathcal{X}(t))$. Esto nos motiva a definir la siguiente regla de clasificación: clasificaremos a $\mathcal{X}(t)$ en la clase g_0 (y definiremos $\widehat{Y} = g_0$) si y sólo si

$$\lambda \left(\left\{ t : \widehat{f}_{\mathcal{X}_t}^{g_0}(\mathcal{X}(t)) > \widehat{f}_{\mathcal{X}_t}^g(\mathcal{X}(t)) \right\} \right) > \lambda \left(\left\{ t : \widehat{f}_{\mathcal{X}_t}^{g_0}(\mathcal{X}(t)) \leq \widehat{f}_{\mathcal{X}_t}^g(\mathcal{X}(t)) \right\} \right), \quad \forall g \neq g_0 \quad (2.52)$$

donde, $\widehat{f}_{\mathcal{X}_t}^g$ es el estimador de $f_{\mathcal{X}_t}^g$.

2.4. Estudios de simulación

En esta sección presentamos algunos estudios de simulación que muestran el desempeño de nuestros métodos de estimación de densidad marginal y discriminación no paramétrica para datos funcionales. Para ello, del conjunto original de datos construimos dos muestras, una *muestra de entrenamiento* $(X_i(t), Y_i)_{i \in \mathcal{E}}$ y una *muestra de prueba* $(X_j(t), Y_j)_{j \in \mathcal{P}}$. Con la muestra de entrenamiento calcularemos el estimador de la densidad para cada grupo $(\hat{f}_e^1, \dots, \hat{f}_e^G)$ (ó $(\hat{f}_{\mathcal{X}_t}^1, \dots, \hat{f}_{\mathcal{X}_t}^G)$) usando el parámetro $\hat{k}_n^g, g = 1, \dots, G$ obtenido por validación cruzada.

Para medir el desempeño de nuestro método de clasificación, evaluamos los estimadores obtenidos con la muestra de entrenamiento en la muestra de prueba y clasificamos de acuerdo a la regla dada en (2.52), es decir, para cada j elegimos \hat{Y}_j correspondiente a $X_j(t)$. Finalmente, calculamos el *error de clasificación* como sigue

$$\text{Misclas} = \frac{1}{\#\mathcal{P}} \sum_{j \in \mathcal{P}} \mathbb{I}_{\{\hat{Y}_j \neq Y_j\}}.$$

Ejemplo 2.19. Sea $T = (0, 1]$. Consideremos el proceso estocástico estacionario $\mathcal{X}(t)$ definido por

$$\mathcal{X}(t) = \mu + \sigma e(t), \quad t \in (0, 1] \quad (2.53)$$

donde

$$e(t) = \frac{w(t)}{\sqrt{t}}, \quad \text{con } w(t) \text{ el movimiento Browniano.}$$

En una primera etapa consideramos $\mu = 0$ y $\sigma = 1$ de manera que $\mathcal{X}(t) = e(t)$ es estacionario y para cada t , $\mathcal{X}(t) \sim \mathcal{N}(0, 1)$. En la Figura 4 (a) hemos graficado 20 curvas de una muestra de tamaño 200 de $\mathcal{X}(t)$ medidas en 100 puntos igualmente espaciados en $(0, 1]$ y la Figura 4 (b) muestra la función de densidad teórica de $\mathcal{X}(t)$ y su estimador calculado con esa misma

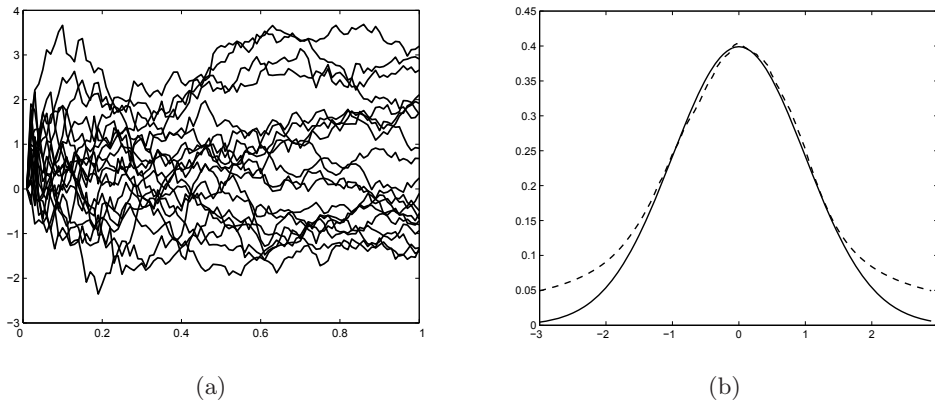


FIGURA 4. (a) 20 curvas de la muestra de tamaño 200 de $\mathcal{X}(t)$. (b) Función de densidad estimada (línea de puntos) y teórica (línea sólida) de $\mathcal{X}(t)$ para $k_n = 43.196$.

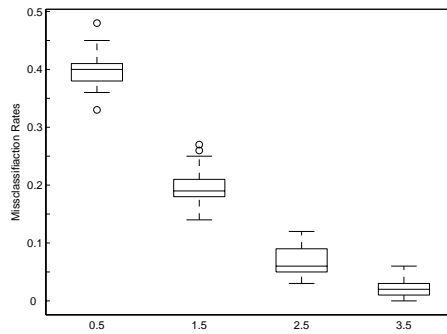


FIGURA 5. Boxplot de los errores de clasificación para 50 réplicas.

muestra. Como podemos ver en la Figura 4 (b), el estimador ajusta muy bien a la verdadera función de densidad excepto en las colas donde, debido a la naturaleza de los procesos, no tenemos suficiente información para realizar una buena estimación.

Para evaluar el desempeño de nuestro método de clasificación, en una segunda etapa consideramos dos clases bajo el modelo (2.53), ambas con $\sigma = 1$ pero una de ellas con μ constante e igual a 0, y la otra con media $\mu \neq 0$. En particular, consideraremos las clases con medias $\mu = 0.5, 1.5, 2.5, 3.5$. Generamos una muestra de entrenamiento de tamaño 200 (100 de cada clase) medidas en 150 instantes de tiempo en el intervalo $(0, 1]$ y una muestra de prueba del mismo tamaño. Con la muestra de entrenamiento calculamos los estimadores para cada clase que luego evaluamos en la muestra de prueba para obtener el error de clasificación.

Repetimos este procedimiento 50 veces para obtener 50 errores de clasificación para cada clase los cuales se muestran en la Figura 5. Notemos como el error de clasificación decrece cuando las medias se alejan; esto es debido a que cuando clasificamos dos poblaciones que son muy cercanas en media sus densidades presentan una cantidad considerable de solapamiento, haciendo que sea difícil distinguir entre grupos (ver la Figura 6).

Ejemplo 2.20. En este ejemplo mostramos el desempeño de nuestro método cuando los datos no son estacionarios. Recientemente, [Shin, 2008] propuso una extensión del análisis discriminante para procesos estocásticos (InfFLD). En particular, él publicó los resultados de un estudio de simulación donde se compara el método InfFLD con el análisis discriminante clásico multivariado (FLD), análisis discriminante penalizado (PDA) usando ambos: el penalizante “ridge” (PDA/Ridge) y una matriz penalizante para splines cúbicos suavizados (PDA/Spline), y con dos métodos de discriminación no paramétricos propuestos por [Ferraty and Vieu, 2003]: análisis componentes principales (NPCD/PCA) y regresión de mínimos cuadrados parciales multivariada (NPCD/MPLSR).

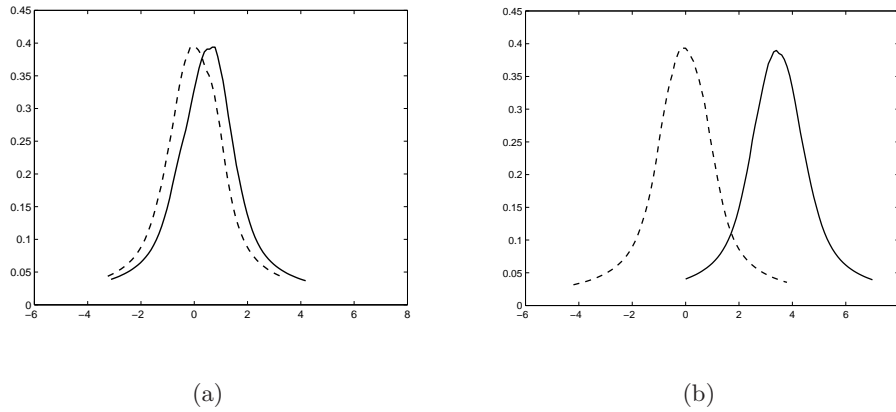


FIGURA 6. (a) Estimador de la densidad para una $\mathcal{N}(0, 1)$ (línea de puntos) y para una $\mathcal{N}(0.5, 1)$ (línea sólida). (b) Estimador de la densidad para una $\mathcal{N}(0, 1)$ (línea de puntos) y para una $\mathcal{N}(3.5, 1)$ (línea sólida).

En esas simulaciones, las poblaciones fueron generadas mediante los siguientes modelos

$$\mathcal{X}_1(t) = 3\sqrt{2}\cos(\pi t) + \sqrt{2}\cos(2\pi t) + e(t) \quad \text{y} \quad \mathcal{X}_2(t) = \sqrt{2}\cos(2\pi t) + e(t)$$

con

$$e(t) = \sum_{i=1}^{30} i^{-1/2} U_i \sqrt{2} \cos(i\pi t),$$

donde U_i son variables aleatorias iid normales standard. Para cada clase, él genera una muestra de entrenamiento de tamaño 100 (50 para cada clase) medidas en 100 instantes de tiempo en el intervalo $(0, 1]$ y una muestra de prueba de tamaño 500. La Figura 7 muestra 10 curvas para cada una de las dos clases y sus respectivas funciones media.

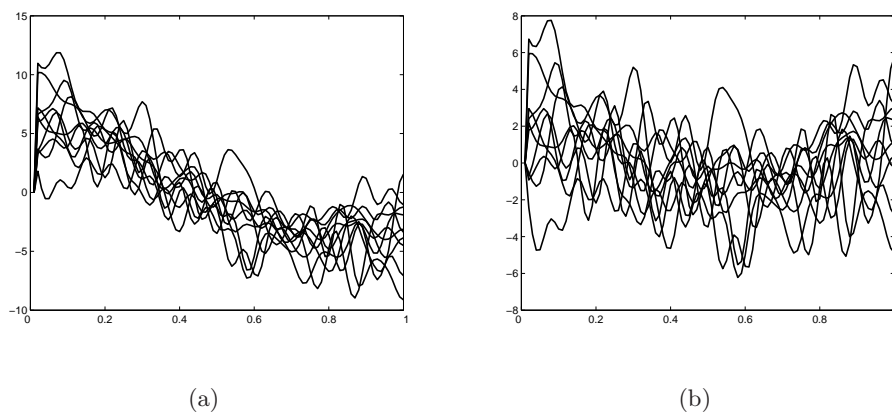


FIGURA 7. (a) 10 curvas correspondientes a una muestra de tamaño 50 de $\mathcal{X}_1(t)$. (b) 10 curvas correspondientes a una muestra de tamaño 50 de $\mathcal{X}_2(t)$.

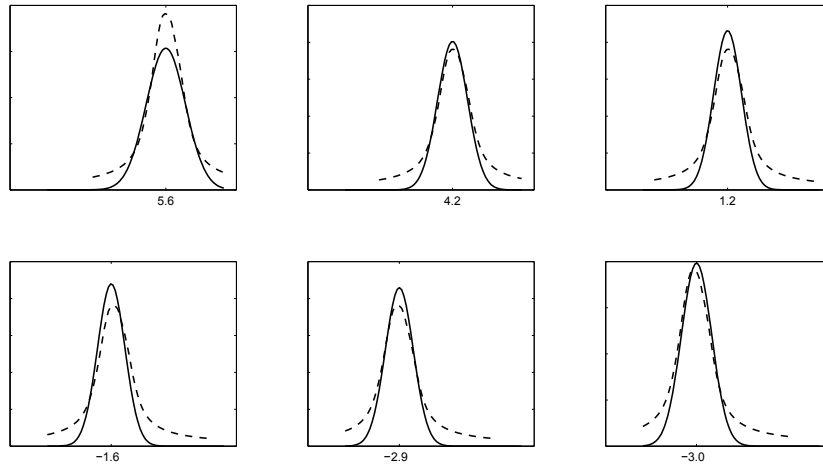


FIGURA 8. Función de densidad estimada (línea de puntos) y teórica (línea sólida) de $\mathcal{X}_1(t)$ para $t = 0.013, 0.180, 0.347, 0.513, 0.680, 0.847$.

Para nuestro propósito, corrimos el mismo experimento con los mismos tamaño de muestra. Las Figuras 8 y 9 muestran el estimador de la densidad y la densidad teórica de $\mathcal{X}_1(t)$ y $\mathcal{X}_2(t)$, respectivamente, para algunos instantes de tiempo. Las líneas de puntos corresponden a los estimadores mientras que las líneas sólidas corresponden a las densidades teóricas. Como en el caso estacionario (Ejemplo 2.19) el estimador de la densidad ajusta muy bien a la densidad teórica excepto en las colas donde no tenemos suficiente información para realizar la estimación.

Finalmente, evaluamos los estimadores en la muestra de prueba para calcular el error de clasificación. Este procedimiento fue repetido 49 veces más construyendo aleatoriamente 49 muestras de entrenamiento y 49 muestras de prueba obteniendo así 50 errores de clasificación.

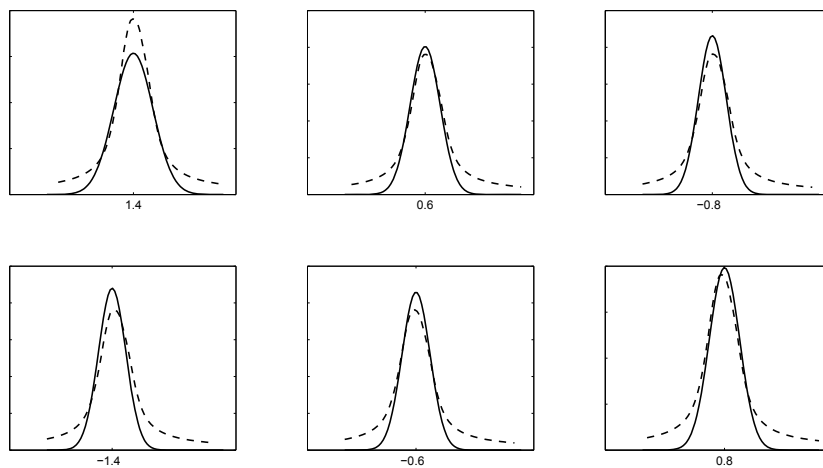


FIGURA 9. Función de densidad estimada (línea de puntos) y teórica (línea sólida) de $\mathcal{X}_2(t)$ para $t = 0.013, 0.180, 0.347, 0.513, 0.680, 0.847$.

Método	Media	Mediana	Desv. Standard
InfFLD	0.0832	0.082	0.0109
FLD	0.2086	0.2	0.0418
PDA/Ridge	0.0889	0.086	0.0181
PDA/Spline	0.0891	0.087	0.0163
NPCD/PCA	0.0906	0.086	0.0152
NPCD/MPLSR	0.0992	0.089	0.0323
NPDE	0.0889	0.0890	0.0096

CUADRO 1. Media, mediana y desviación standard del error de clasificación para el algoritmo basado en RKHS (InfFLD), método de Fisher clásico (FLD), método discriminante penalizado (PDA), método de discriminación no paramétrica de curvas (MPLSR) y método de estimación de densidad no paramétrica (NPDE).

En la Tabla 1 reproducimos los resultados dados en [Shin, 2008] y agregamos los resultados obtenidos con nuestro método al que llamamos NPDE. Como podemos ver, el método NPDE no se comporta tan bien como InfFLD pero se comporta mejor que todos los demás métodos.

Observemos que para todo t

$$e(t) \sim \mathcal{N} \left(0, 2 \sum_{i=1}^{30} \cos^2(\pi ti)/i \right).$$

Es decir, $e(t)$ no es estacionario ya que su varianza (y por lo tanto su distribución) depende del tiempo. Esto demuestra que nuestro método es robusto con respecto a la no estacionaridad.

2.5. Aplicación a datos reales

En esta sección aplicaremos nuestros resultados a un conjunto de datos reales no estacionarios. Dicho conjunto de datos corresponde a una parte del original que puede ser encontrado en (<http://www-stat.stanford.edu/ElemStatLearn>) y consiste en 400 fragmentos correspondientes a 5 fonemas medidos en 150 instantes de tiempo. La Figura 10 muestra 10 curvas de log-periodogramas para cada una de las clases.

Más precisamente, tenemos 2000 pares $(\mathcal{X}_i(t), Y_i)$ donde las X_i corresponden a log-periodogramas discretizados en 150 instantes de tiempo mientras que las Y_i indican una de

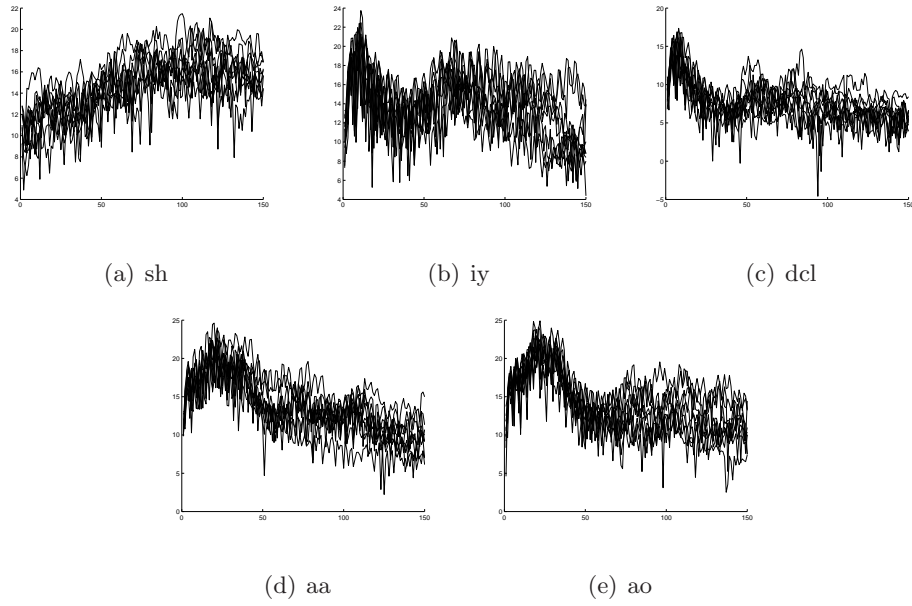


FIGURA 10. Muestra de 10 Log-Periodogramas para cada una de las 5 clases de fonemas.

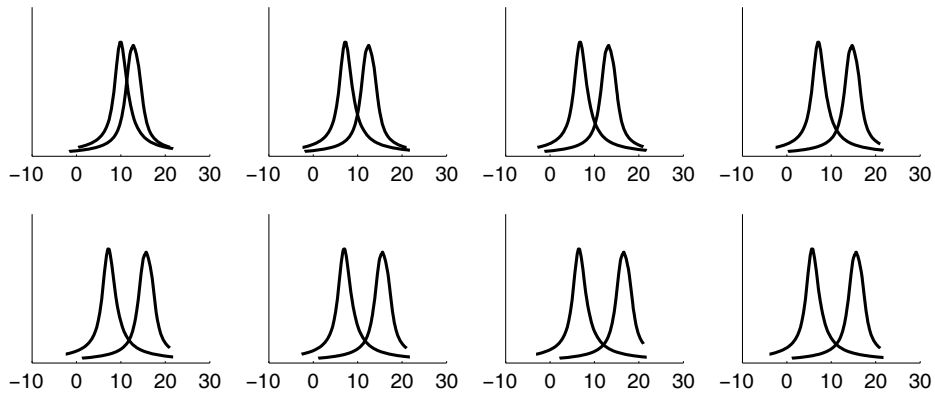


FIGURA 11. Función de densidad estimada de las 2 clases (“sh” y “dcl”) para $t = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$.

las 5 clases a la cual la curva $\mathcal{X}_i(t)$ pertenece:

$$Y_i \in \{1, 2, 3, 4, 5\} \text{ con } \begin{cases} 1 \leftrightarrow \text{“sh”} \\ 2 \leftrightarrow \text{“iy”} \\ 3 \leftrightarrow \text{“dcl”} \\ 4 \leftrightarrow \text{“aa”} \\ 5 \leftrightarrow \text{“ao”} \end{cases}$$

Para comenzar, del conjunto total de datos consideramos solo las clases “sh” y “dcl” y con ellas construimos dos muestras de tamaño 100 (50 observaciones por clase), una *muestra de entrenamiento* $(\mathcal{X}_i(t), Y_i)_{i \in \mathcal{E}}$ y una *muestra de prueba* $(X_j(t), Y_j)_{j \in \mathcal{P}}$. Con la muestra de

entrenamiento calculamos el estimador de la densidad para cada grupo $(\widehat{f}_{\mathcal{X}_t}^1, \widehat{f}_{\mathcal{X}_t}^2)$ usando el parámetro óptimo \widehat{k}_n^g , para $g = 1, 2$ obtenido por validación cruzada y luego, evaluamos los estimadores obtenidos con la muestra de entrenamiento en la muestra de prueba y clasificamos de acuerdo a la regla dada en (2.52). Finalmente, calculamos el *error de clasificación* como sigue

$$Misclas = \frac{1}{100} \sum_{j \in \mathcal{P}} \mathbb{I}_{\{\widehat{Y}_j \neq Y_j\}}.$$

La Figura 11 muestra el estimador de la densidad para cada clase para algunos instantes de tiempo, mientras que la Figura 12 muestra la distribución de los errores de clasificación.

Para continuar con este ejemplo, consideramos las cinco clases y como antes, construimos dos muestras ahora de tamaño 250 (50 observaciones por clase), una *muestra de entrenamiento* $(\mathcal{X}_i(t), Y_i)_{i \in \mathcal{E}}$ y una *muestra de prueba* $(X_j(t), Y_j)_{j \in \mathcal{P}}$. Nuevamente con la muestra de entrenamiento calculamos el estimador de la densidad para cada grupo $(\widehat{f}_{\mathcal{X}_t}^1, \dots, \widehat{f}_{\mathcal{X}_t}^5)$ usando el parámetro óptimo \widehat{k}_n^g , con $g = 1, \dots, 5$ obtenido por validación cruzada y luego evaluamos los estimadores obtenidos en la muestra de prueba para calcular el *error de clasificación*

$$Misclas = \frac{1}{250} \sum_{j \in \mathcal{P}} \mathbb{I}_{\{\widehat{Y}_j \neq Y_j\}}.$$

La Figura 13 muestra, el estimador de la densidad para cada clase para algunos instantes de tiempo mientras que la Figura 14 muestra la distribución de los errores de clasificación.

Notemos que en este ejemplo, cuando clasificamos sólo las clases “sh” y “dcl” el método funciona muy bien mientras que cuando clasificamos las cinco clases no lo hace tan bien. Esta diferencia en el comportamiento del método se debe a que en el primer caso, para la mayoría de instantes de tiempo las densidades no presentan demasiado solapamiento lo que le permite al método distinguir bien entre las dos clases (ver Figura 11). Sin embargo, cuando consideramos las cinco clases nuestro método no produce una buena clasificación debido a que para

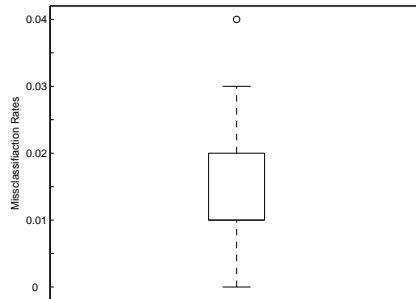


FIGURA 12. Distribución de los errores de clasificación para las clases “sh” y “dcl” .

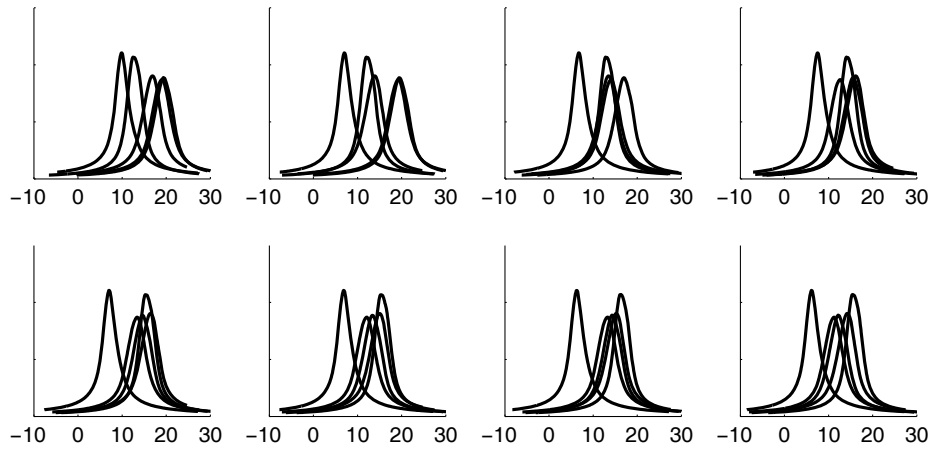


FIGURA 13. Función de densidad estimada de las 5 clases para $t = 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1$.

muchos instantes de tiempo las densidades estimadas presentan una cantidad considerable de solapamiento como se puede apreciar en la Figura 13, dificultando la distinción entre clases.

Todo el análisis computacional de las Secciones 2.4 y 2.5 fue realizado usando el lenguaje de computación técnica **Matlab** (ver [MatLab, 2009]).

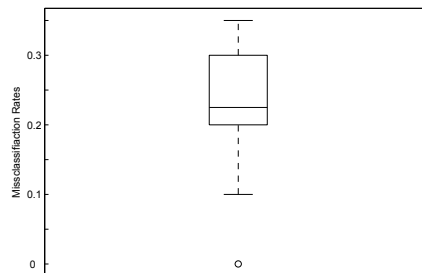


FIGURA 14. Distribución de los errores de clasificación para 5 clases.

CAPÍTULO 3

ESTIMACIÓN DE DENSIDADES PARA CAMPOS ALEATORIOS

En el Capítulo 2 hemos definido y estudiado estimadores de la densidad para procesos estocásticos definidos en \mathbb{R} con valores en \mathbb{R} . En este capítulo extenderemos esa definición a procesos estocásticos \mathbb{R} valuados pero definidos en espacios de mayores dimensiones, digamos \mathbb{R}^d . Tales procesos serán llamados **campos aleatorios** y en este contexto, el modelo bajo el cual trabajaremos será llamado modelo **espacio-temporal**. Más precisamente, desde nuestro punto de vista no paramétrico, un campo aleatorio será una superficie (hipersuperficie) aleatoria $\mathcal{X}(\mathbf{s}) : \mathbb{R}^d \rightarrow \mathbb{R}$ y un modelo espacio-temporal será una superficie (hipersuperficie) aleatoria que evoluciona en el tiempo $\{\mathcal{X}_t(\mathbf{s}), t \geq 1\}$.

Si bien en la última década ha habido un gran crecimiento en la investigación de modelos espacio-temporales en el contexto paramétrico (ver [Tang et al., 2008]) y a pesar del gran desarrollo de métodos de estimación no paramétrica de la densidad para datos funcionales (ver Capítulo 2) poco se ha hecho hasta el momento para aplicar tales métodos de estimación a campos aleatorios. En este contexto, la mayoría de trabajos se refieren a estimar la densidad marginal de campos aleatorios estrictamente estacionarios definidos en \mathbb{Z}^d (el lattice entero d -dimensional) y con valores en \mathbb{R}^N . Por ejemplo, [Tran and Yakowitz, 1993] probaron la normalidad asintótica del estimador de k -vecinos más cercanos bajo condiciones de dependencia temporal. Para estimadores de núcleo, [Tran, 1990] probó la normalidad asintótica bajo condiciones de dependencia y aplicó sus resultados a varios modelos espacio-temporales. La convergencia en L_1 de este tipo de estimadores ha sido estudiada por [Carbon et al., 1996] y luego por [Hallin et al., 2004] quienes consideraron procesos espacio-temporales con estructura lineal y no lineal y dieron condiciones suficientes para su convergencia. [Hallin et al., 2001], sin

suponer dependencia pero suponiendo linealidad, probaron la normalidad asintótica multivariada del estimador de núcleo para cada t -upla en la grilla del espacio y calcularon su matriz de covarianza límite. La consistencia uniforme bajo dependencia de este estimador fue probada por [Carbon et al., 1997].

El objetivo de este capítulo será estimar la densidad marginal de un campo aleatorio $\mathcal{X}(\mathbf{s}) : \mathbb{R}^d \rightarrow \mathbb{R}$ que verifica el modelo

$$\mathcal{X}(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s}), \quad \mathbf{s} \in \mathbf{S} \subset \mathbb{R}^d, \quad (3.1)$$

donde, \mathbf{S} es un subconjunto acotado de \mathbb{R}^d , $\mu(\mathbf{s})$ es la media poblacional y $e(\mathbf{s})$ es un campo aleatorio estacionario con media cero. En este contexto consideraremos como en el caso funcional dos casos: $\mu(\mathbf{s})$ constante con respecto al espacio, esto es, $\mu(\mathbf{s}) = \mu$ y $\mu(\mathbf{s})$ no constante. En el primer caso, $\mathcal{X}(\mathbf{s})$ será un campo aleatorio estacionario con media μ y función de densidad no conocida $f_{\mathcal{X}}$ y en el segundo, $\mathcal{X}(\mathbf{s})$ ya no será estacionario y en este caso, su función de densidad no conocida de $\mathcal{X}(\mathbf{s})$ será denotada por $f_{\mathcal{X}(\mathbf{s})} \doteq f_{\mathcal{X}_{\mathbf{S}}}$.

Este capítulo está organizado de la siguiente manera. Dado que los datos ahora tendrán una dependencia en el tiempo, en la Sección 3.1 definimos dos condiciones de dependencia e introducimos una nueva noción que será utilizada en este capítulo. En la Sección 3.2 extendemos el estimador definido en el Capítulo 2 a campos estacionarios centrados, probamos su consistencia y encontramos velocidades de convergencia. En la Sección 3.3, extendemos la definición dada en 3.2 a campos no estacionarios, probamos su consistencia y encontramos velocidades de convergencia. Finalmente, en la Sección 3.4 mediante estudios de simulación, mostramos el desempeño de nuestro método de estimación para $d = 2$ tanto para el caso estacionario como para el no estacionario. En el Apéndice 3.5 probamos algunos resultados utilizados en el capítulo.

3.1. Nociones de dependencia

Durante mucho tiempo las condiciones tipo mixing han sido el tipo de condiciones más utilizadas al momento de imponer restricciones de dependencia en sucesiones de variables aleatorias. La primera noción de dependencia mixing fue introducida por [Rosenblatt, 1956] y su definición es la siguiente.

Definición 3.1. Diremos que una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ es α -mixing si existe una sucesión no creciente de números positivos $\{\alpha(r), r \in \mathbb{N}\}$ con $\alpha(r) \xrightarrow{r \rightarrow \infty} 0$ tal que

para todo entero r

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(r),$$

donde, para $1 \leq n_1 \leq \dots \leq n_u < n_u + r = m_1 \leq \dots \leq m_v \leq \infty$, $A \in \mathcal{M}_{n_1}^{n_u}$ y $B \in \mathcal{M}_{m_1}^{m_v}$ con \mathcal{M}_a^b la σ -álgebra generada por las variables $\{X_n\}_{n=a}^b$. Si además existen $0 < \rho < 1$ y $a > 0$ tales que $\alpha(r) \leq a\rho^r$ entonces, diremos que la sucesión $\{X_n\}_{n \in \mathbb{N}}$ es **α -mixing geométrica**.

Aunque esta noción es una de las condiciones más débiles de dependencia mixing (ver [Doukhan, 1994]), [Doukhan and Louhichi, 1999] observaron que ciertos procesos que son de interés en la estadística no son α -mixing aunque presenten una sucesiva disminución de la influencia del pasado. Tal es el caso del proceso autorregresivo de orden uno (o AR(1)) dado por $\mathcal{X}_n = \theta\mathcal{X}_{n-1} + \epsilon_n, n \in \mathbb{Z}$, donde ϵ_n son independientes y $P(\epsilon_n = 1) = P(\epsilon_n = -1) = \frac{1}{2}$, $0 < |\theta| \leq \frac{1}{2}$.

Inspirados por este problema, estos autores introdujeron una noción de dependencia a la que llamaron **dependencia débil** (Def. 3.2) y utilizando el proceso AR(1) probaron que esta definición es aún más débil que la dependencia α -mixing (Lemas 3.3 y 3.4).

Definición 3.2. ([Doukhan and Louhichi, 1999]) Diremos que una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ es **$(\mathcal{G}, \alpha, \psi)$ -débilmente dependiente** si existe una clase de funciones a valores reales \mathcal{G} , una sucesión no creciente de números positivos $\{\alpha(r), r \in \mathbb{N}\}$ con $\alpha(r) \xrightarrow{r \rightarrow \infty} 0$ y una función $\psi: \mathcal{G}^2 \times \mathbb{N}^2$ tales que para toda u -upla (n_1, \dots, n_u) y para cada v -upla (m_1, \dots, m_v) con $1 \leq n_1 \leq \dots \leq n_u < n_u + r = m_1 \leq \dots \leq m_v \leq \infty$ se tiene que

$$|\text{cov}(f(X_{n_1}, \dots, X_{n_u}), g(X_{m_1}, \dots, X_{m_v}))| \leq \psi(f, g, u, v) \alpha(r)$$

para cualquier par de funciones $f, g \in \mathcal{G}$ definidas en \mathbb{R}^u y \mathbb{R}^v respectivamente.

Lema 3.3. ([Doukhan and Louhichi, 1999]) Consideremos el proceso AR(1) dado por

$$\mathcal{X}_n = \theta\mathcal{X}_{n-1} + \epsilon_n = \theta(\theta\mathcal{X}_{n-2} + \epsilon_{n-1}) + \epsilon_n = \theta^2\mathcal{X}_{n-2} + \theta\epsilon_{n-1} + \epsilon_n = \dots \doteq F(\epsilon_{n-s}, s \in \mathbb{Z})$$

donde F es una función medible definida en \mathbb{Z} y sea

$$\delta_k \doteq \sup_{n \in \mathbb{Z}} \mathbb{E} (|F(\epsilon_{n-s}, s \in \mathbb{Z}) - F(\epsilon_{n-s}, s \in \mathbb{Z}, -k < s < k)|),$$

con $\delta_k \xrightarrow{k \rightarrow \infty} 0$. Entonces, \mathcal{X}_n es $(\mathcal{G}, \alpha, \psi)$ -débilmente dependiente con \mathcal{G} el conjunto de las funciones Lipschitz acotadas, $\alpha(r) = \delta_{r/2}$ y

$$\psi(f, g, u, v) = 2(u \|g\|_\infty \text{Lip}(f) + v \|f\|_\infty \text{Lip}(g)).$$

Si bien la definición de dependencia débil fue dada en el año 1999, [Rosenblatt, 1956] ya había probado el siguiente lema, el cual hace explícita la estructura de dependencia débil de procesos α -mixing.

Lema 3.4. ([Rosenblatt, 1956]) *Si la sucesión $\{X_n\}_{n \in \mathbb{N}}$ es α -mixing entonces, es (α, L^∞, ψ) -débilmente dependiente con $\psi(f, g, u, v) = 4 \|f\|_\infty \|g\|_\infty$.*

Por lo tanto, del ejemplo del proceso AR(1) y del Lema 3.4 vemos que la dependencia α -mixing es más fuerte que la dependencia $(\mathcal{G}, \alpha, \psi)$ -débil.

Observemos que, si en la definición de dependencia $(\mathcal{G}, \alpha, \psi)$ -débil tomamos como $\mathcal{G} = \{\mathbb{I}_E : E \in \mathcal{A}\}$ la clase de las funciones indicatrices sobre conjuntos medibles y como $\psi(f, g, u, v) = \phi(u, v)$ con $\phi : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}^+$ una función mayor o igual a uno, para $A \in \mathcal{M}_{n_1}^{n_u}$ y $B \in \mathcal{M}_{m_1}^{m_v}$ tendremos que

$$|P(A \cap B) - P(A)P(B)| = |\mathbb{E}(\mathbb{I}_A \mathbb{I}_B) - \mathbb{E}(\mathbb{I}_A) \mathbb{E}(\mathbb{I}_B)| = |\text{cov}(\mathbb{I}_A, \mathbb{I}_B)| \leq \alpha(r) \phi(u, v).$$

Una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ para la cual la desigualdad anterior se verifica diremos que es una sucesión **(α, ϕ) -débilmente dependiente**. En lo que sigue, daremos la definición explícita de esta nueva noción de dependencia pues los procesos que utilizaremos en este capítulo serán (α, ϕ) -débilmente dependientes.

Definición 3.5. Diremos que una sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ es **(α, ϕ) -débilmente dependiente** si existe una sucesión no creciente de números positivos $\{\alpha(r), r \in \mathbb{N}\}$ con $\alpha(r) \xrightarrow{r \rightarrow \infty} 0$ y una función $\phi : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{R}^+$ con $\phi(u, v) \geq 1$,

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(r) \phi(u, v) \quad (3.2)$$

donde, para $1 \leq n_1 \leq \dots \leq n_u < n_u + r = m_1 \leq \dots \leq m_v \leq \infty$, $A \in \mathcal{M}_{n_1}^{n_u}$ y $B \in \mathcal{M}_{m_1}^{m_v}$ con \mathcal{M}_a^b la σ -álgebra generada por las variables $\{X_n\}_{n=a}^b$. Si además existen constantes $L_1, L_2 < \infty$ y $\mu \geq 0$ tales que

$$\sum_{j=0}^{\infty} (j+1)^k \alpha(j) \leq L_1 L_2^k (k!)^\mu, \quad \forall k \geq 0, \quad (3.3)$$

entonces, diremos que la sucesión de variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ es **geoméricamente (α, ϕ) -débilmente dependiente**.

Los siguientes dos lemas muestran que la dependencia α -mixing (α -mixing geométrica) es más fuerte que la dependencia (α, ϕ) -débil ((α, ϕ) -débil geométrica).

Lema 3.6. *Si la sucesión $\{X_n\}_{n \in \mathbb{N}}$ es α -mixing entonces es (α, ϕ) -débilmente dependiente.*

La demostración de este lema es inmediata ya que como $\phi(u, v) \geq 1$, para todo par (u, v) ,

$$|P(A \cap B) - P(A)P(B)| \leq \alpha(r) \leq \alpha(r)\phi(u, v).$$

Lema 3.7. *Si $\alpha(r) \leq a\rho^r$, $0 < \rho < 1$ y $a > 0$ entonces, $\sum_{j=0}^{\infty} (j+1)^k \alpha(j) \leq L_1 L_2^k (k!)^\mu$, $\forall k \geq 0$, con $L_1 = \frac{a}{1-\rho}$, $L_2 = \frac{1}{1-\rho}$ y $\mu = 1$.*

La demostración de este lema puede encontrarse en el Apéndice 3.5.

Por lo tanto, como consecuencia de los Lemas 3.6 y 3.7 tenemos que si una sucesión $\{X_n\}_{n \in \mathbb{N}}$ es α -mixing geométrica entonces es (α, ϕ) -débilmente dependiente geométrica para cualquier $\phi \geq 1$ con $L_1 = \frac{a}{1-\rho}$, $L_2 = \frac{1}{1-\rho}$, $\mu = 1$.

3.2. Estimando la función de densidad de un campo estacionario

En esta sección extendemos la definición estimador dado en el Capítulo 2 a un campo aleatorio estacionario de primer orden $\mathcal{X}(\mathbf{s})$ cuya función de densidad marginal es desconocida y queremos estimar. Para ello supondremos que en el modelo (3.1) la función media $\mu(\mathbf{s})$ es constante con respecto al tiempo, es decir,

$$\mathcal{X}(\mathbf{s}) = \mu + e(\mathbf{s}), \quad \mathbf{s} \in \mathbf{S},$$

donde $e(\mathbf{s})$ es un campo aleatorio estacionario de primer orden con media nula. Para este estimador presentamos resultados de consistencia y obtenemos velocidades de convergencia.

3.2.1. Preliminares y definición del estimador

Sean $S_1, S_2, \dots, S_d \subset \mathbb{R}$ intervalos finitos y sea $\mathbf{S} \doteq S_1 \times S_2 \times \dots \times S_d \in \mathbb{R}^d$ un rectángulo de medida $|\mathbf{S}|$. Consideremos el campo aleatorio $\{\mathcal{X}(\mathbf{s}, \omega) \in \mathbb{R} : \mathbf{s} \in \mathbf{S}, \omega \in \Omega\}$ definido en un espacio de probabilidad (Ω, \mathcal{A}, P) el cual es estacionario de primer orden con función de densidad marginal no conocida $f_{\mathcal{X}}$. Supongamos además que $\mathcal{X}(\mathbf{s})$ admite un tiempo local (ver Capítulo 2, Sección 2.1.1).

Notación 3.8. En lo que sigue omitiremos la variable ω en la expresión $\mathcal{X}(\mathbf{s}, \omega)$ y utilizaremos la expresión $\int_{\mathbf{S}} \cdot d\mathbf{s}$ para denotar la integración sobre el rectángulo \mathbf{S} . Más precisamente, si

$\mathbf{s} = (s_1, s_2, \dots, s_d) \in \mathbf{S}$ con $s_1 \in S_1, s_2 \in S_2, \dots, s_d \in S_d$ entonces,

$$\int_{\mathbf{S}} \cdot d\mathbf{s} \doteq \int_{S_1} \int_{S_2} \dots \int_{S_d} \cdot ds_1 ds_2 \dots ds_d. \quad (3.4)$$

Además, para denotar u integrales d -dimensionales, utilizaremos la siguiente notación.

$$\int_{\mathbf{S}^u} d\mathbf{s}^u \doteq \int_{\mathbf{S}} \dots \int_{\mathbf{S}} ds_1 \dots ds_u.$$

donde, para cada $i = 1, \dots, u$, $\int_{\mathbf{S}} ds_i$ es como en (3.4) para $\mathbf{s}_i = (s_{i1}, s_{i2}, \dots, s_{id}) \in \mathbf{S}$.

Sea $\{\mathcal{X}_t(\mathbf{s})\}_{t=1}^T$ una sucesión de campos aleatorios con la misma distribución que $\mathcal{X}(\mathbf{s})$, $I_{(x,r)} = [x-r, x+r]$ el intervalo de centro x y radio r y $\{k_T\}$, $k_T/T < |\mathbf{S}|$, una sucesión de números reales positivos que tiende a infinito.

Al igual que en el caso funcional, la variable aleatoria $h_T^{\mathcal{X}} \doteq h_T^{\mathcal{X}}(x)$ es elegida de manera tal que

$$k_T = \sum_{t=1}^T \int_{\mathbf{S}} \mathbb{I}_{I_{(x, h_T^{\mathcal{X}}(x))}}(\mathcal{X}_t(\mathbf{s})) d\mathbf{s}, \quad (3.5)$$

y el estimador de la función de densidad $f_{\mathcal{X}}$ está dado por

$$\widehat{f}_{\mathcal{X}}(x) \doteq \frac{k_T}{2T|\mathbf{S}|h_T^{\mathcal{X}}(x)}. \quad (3.6)$$

Observación 3.9. Si el proceso $\mathcal{X}(\mathbf{s})$ admite un tiempo local (ver Sección 2.1.1), entonces $\widehat{f}_{\mathcal{X}}$ está bien definida ya que $h_T^{\mathcal{X}}$ existe y es único (ver Observación 2.3).

Observación 3.10. Si $\{\mathcal{X}_t(\mathbf{s})\}_{t=1}^T$ fuera una muestra i.i.d. de $\mathcal{X}(\mathbf{s})$, una extensión directa de los resultados del Capítulo 2 a \mathbb{R}^d nos permitiría obtener la consistencia del estimador (3.6). Sin embargo, en este capítulo no supondremos independencia de los datos sino que los procesos serán (α, ϕ) -débilmente dependientes y por lo tanto necesitaremos utilizar otras desigualdades exponenciales para obtener velocidades de convergencia.

3.2.2. Hipótesis generales

Consideraremos entonces las siguientes hipótesis generales:

- H1 $\mathcal{X}(\mathbf{s})$ es un campo estacionario de primer orden con función de densidad estrictamente positiva desconocida $f_{\mathcal{X}}$ el cual admite un tiempo local;
- H2 la densidad $f_{\mathcal{X}}$ es una función Lipschitz con constante K ;

H3 $\{\mathcal{X}_t(\mathbf{s}), \mathbf{s} \in \mathbf{S}\}_{t=1}^T$ es una sucesión de campos aleatorios con la misma distribución que $\mathcal{X}(\mathbf{s})$ tal que, para cada \mathbf{s} fijo, la sucesión de variables aleatorias $\{\mathcal{X}_t(\mathbf{s})\}_{t=1}^T$ es (α, ϕ) -débilmente dependiente geométrica con ϕ alguna de las siguientes funciones:

- $\phi(u, v) = 2v$,
- $\phi(u, v) = u + v$,
- $\phi(u, v) = uv$,
- $\phi(u, v) = \rho(u + v) + (1 - \rho)uv$, para algún $\rho \in (0, 1)$;

H4 $\{k_T\}$ y $\{v_T\}$ son sucesiones de enteros positivos que tienden a infinito tales que $v_T \left(\frac{k_T}{T}\right) = o(1)$ y $\sum_{T=1}^{\infty} \exp\left(-\frac{a_1 k_T^2}{a_1 v_T^2 T + a_2 v_T^2 \left(\frac{k_T}{v_T}\right)^{(2\mu+3)/(\mu+2)}}\right) < \infty$, para cada $a_1, a_2 > 0$ y $\mu \geq 0$.

3.2.3. Resultados asintóticos

Teorema 3.11. Velocidades de convergencia. *Supongamos que H1 - H4 se cumplen. Entonces, para todo $x \in \mathbb{R}$,*

$$\lim_{T \rightarrow \infty} v_T \left(\widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right) = 0 \quad a.co.$$

Observación 3.12. Nuestras hipótesis implican que podemos elegir k_T tal que $v_T = T^\gamma$ para cualquier $\gamma < \frac{1}{4}$. Más precisamente, si $k_T = T^\beta$ y $v_T = T^\gamma$, para que las condiciones $(k_T/T) v_T = o(1)$ y $\frac{a_1 k_T^2}{a_1 v_T^2 T + a_2 v_T^2 \left(\frac{k_T}{v_T}\right)^{(2\mu+3)/(\mu+2)}} = \frac{a_1 T^{2\beta-1}}{a_1 v_T^2 T k_T^{-2} + a_2 v_T^{\frac{\mu+2}{\mu+2}} k_T^{\frac{\mu+2}{\mu+2}}} \rightarrow \infty$ se cumplan es suficiente que $\beta - 1 + \gamma < 0$ y $\beta - \gamma - \frac{1}{2} > 0$ o equivalentemente, $\beta < 1 - \gamma$ y $\beta > \gamma + \frac{1}{2}$. Luego, dado $\gamma < \frac{1}{4}$ es posible elegir β de manera que las hipótesis se verifiquen (ver Figura 1).

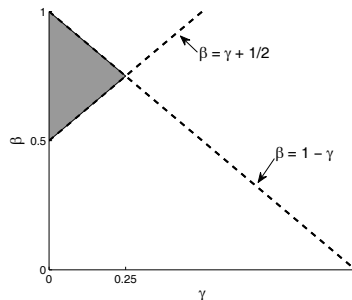


FIGURA 1. El área sombreada muestra la intersección de las regiones $\beta < 1 - \gamma$ y $\beta > \gamma + \frac{1}{2}$.

Observación 3.13. Aunque la dependencia α -mixing geométrica es más fuerte que la dependencia (α, ϕ) -débil geométrica (Lema 3.7), utilizando la desigualdad de Bernstein (Teorema A.34) puede verse que las velocidades de convergencia obtenidas en el caso α -mixing geométrico son las mismas que en este caso.

Demostración del Teorema 3.11: Por definición de convergencia casi completa necesitamos probar que para todo $\epsilon > 0$,

$$\sum_{T=1}^{\infty} P\left(v_T \left| \widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right| > \epsilon\right) < \infty.$$

Sean $\epsilon_T \doteq \frac{\epsilon}{v_T}$ y

$$C_T = C_T(x) \doteq \left\{ v_T \left| \widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right| > \epsilon \right\} = \left\{ \left| \widehat{f}_{\mathcal{X}}(x) - f_{\mathcal{X}}(x) \right| > \epsilon_T \right\}.$$

Escribimos,

$$C_T = A_T \cup B_T,$$

con

$$A_T = A_T(x) \doteq \left\{ h_T^{\mathcal{X}}(x) < \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x) + \epsilon_T)} \right\}$$

y

$$B_T = B_T(x) \doteq \begin{cases} \left\{ h_T^{\mathcal{X}}(x) > \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x) - \epsilon_T)} \right\} & \text{si } f_{\mathcal{X}}(x) > \epsilon_T \\ \emptyset & \text{si } f_{\mathcal{X}}(x) \leq \epsilon_T. \end{cases}$$

Definiendo

$$a_T = a_T(x) \doteq \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x) + \epsilon_T)} \quad \text{y} \quad b_T = b_T(x) \doteq \frac{k_T}{2T|\mathbf{S}|(f_{\mathcal{X}}(x) - \epsilon_T)}$$

tenemos que

$$A_T = \{h_T^{\mathcal{X}} < a_T\} \quad \text{y} \quad B_T \subseteq \{h_T^{\mathcal{X}} > b_T, f_{\mathcal{X}}(x) > \epsilon_T\}.$$

Además,

$$h_T^{\mathcal{X}} < a_T \Leftrightarrow \sum_{t=1}^T \underbrace{\int_{\mathbf{S}} \mathbb{I}_{I_{(x, a_T)}}(\mathcal{X}_t(\mathbf{s})) \, d\mathbf{s}}_{\doteq Y_t^{aT}} > k_T \quad \text{y} \quad h_T^{\mathcal{X}} > b_T \Leftrightarrow \sum_{t=1}^T \underbrace{\int_{\mathbf{S}} \mathbb{I}_{I_{(x, b_T)}}(\mathcal{X}_t(\mathbf{s})) \, d\mathbf{s}}_{\doteq Y_t^{bT}} < k_T,$$

de donde se sigue que,

$$P(A_T) = P\left(\sum_{t=1}^T Y_t^{aT} > k_T\right) \tag{3.7}$$

y

$$P(B_T) \leq P\left(\sum_{t=1}^T Y_t^{bT} < k_T, f_{\mathcal{X}}(x) > \epsilon_T\right).$$

Por lo tanto, dado que $P(C_T) \leq P(A_T) + P(B_T)$ será suficiente probar que

$$\sum_{T=1}^{\infty} P(A_T) = \sum_{T=1}^{\infty} P\left(\sum_{t=1}^T Y_t^{aT} > k_T\right) < \infty \tag{3.8}$$

y

$$\sum_{T=1}^{\infty} P(B_T) \leq \sum_{T=1}^{\infty} P\left(\sum_{t=1}^T Y_t^{bT} < k_T, f_{\mathcal{X}}(x) > \epsilon_T\right) < \infty. \tag{3.9}$$

La demostración de (3.9) es análoga a la de (3.8) y por lo tanto será omitida. Para probar (3.8) definimos $\bar{Y}^{a_T} \doteq Y^{a_T} - \mathbb{E}(Y^{a_T})$. Reemplazando n por T , t por s y T por \mathbf{S} en el Lema 2.6 para $c = a_T$ tenemos que $\mathbb{E}(Y_t^{a_T}) = |\mathbf{S}|p_{a_T}$ luego, en (3.7) resulta,

$$P(A_T) = P\left(\sum_{t=1}^T Y_t^{a_T} > k_T\right) = P\left(\sum_{t=1}^T \bar{Y}_t^{a_T} > k_T - T|\mathbf{S}|p_{a_T}\right).$$

Análogamente a lo hecho en la demostración del Teorema 2.7 podemos obtener la ecuación análoga a (2.23) para campos aleatorios. Más precisamente, para todo $T \geq T_1$ para algún $T_1 = T_1(x)$ y una constante $C_1 = C_1(x)$ tenemos que

$$P(A_T) \leq P\left(\sum_{t=1}^T \bar{Y}_t^{a_T} > C_1\left(\frac{k_T}{v_T}\right)\right). \quad (3.10)$$

Por lo tanto, será suficiente probar que

$$\sum_{T=1}^{\infty} P\left(\sum_{t=1}^T \bar{Y}_t^{a_T} > C_1\left(\frac{k_T}{v_T}\right)\right) < \infty.$$

Con el fin de aplicar la desigualdad de Bernstein para variables aleatorias $(\mathcal{G}, \alpha, \psi)$ -débilmente dependientes (Teorema A.33), necesitamos el siguiente lema el cuál será demostrado en el Apéndice 3.5.

Lema 3.14. *Supongamos que la hipótesis H3 se verifica. Consideremos las variables aleatorias $Y_t^c \doteq \int_{\mathbf{S}} \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_t(\mathbf{s})) d\mathbf{s}$ y las variables aleatorias centradas $\bar{Y}_t^c \doteq Y_t^c - \mathbb{E}(Y_t^c)$ donde c puede depender de x . Entonces,*

- (i) *para cada u -upla (t_1, \dots, t_u) y para cada v -upla (l_1, \dots, l_v) , $1 \leq t_1 \leq \dots \leq t_u < t_u + r = l_1 \leq \dots \leq l_v \leq T$ se tiene que*

$$|\text{cov}(\bar{Y}_{t_1}^c \dots \bar{Y}_{t_u}^c, \bar{Y}_{l_1}^c \dots \bar{Y}_{l_v}^c)| \leq (2|\mathbf{S}|)^{u+v} \phi(u, v) \alpha(r), \quad (3.11)$$

donde $2|\mathbf{S}|$ es tal que $|\bar{Y}_t^c| \leq 2|\mathbf{S}|$ para todo t , $\alpha(r) \xrightarrow{r \rightarrow \infty} 0$ con ϕ cualquiera de las funciones dadas en H3;

- (ii) *para alguna constante C se tiene que,*

$$\text{var}\left(\sum_{t=1}^T \bar{Y}_t^c\right) \leq CT.$$

En virtud de la parte (i) del Lema anterior con $c = a_T$, la sucesión $\{\bar{Y}_t^{a_T}\}_{t=1}^T$ resulta ser $(\mathcal{G}, \alpha, \psi)$ -débilmente dependiente con $\{\alpha(r), r \in \mathbb{N}\}$ dada en (3.2), $f : \mathbb{R}^u \rightarrow \mathbb{R}$ y $g : \mathbb{R}^v \rightarrow \mathbb{R}$ dadas por $f(\bar{Y}_{t_1}^{a_T}, \dots, \bar{Y}_{t_u}^{a_T}) = \bar{Y}_{t_1}^{a_T} \dots \bar{Y}_{t_u}^{a_T}$ y $g(\bar{Y}_{l_1}^{a_T}, \dots, \bar{Y}_{l_v}^{a_T}) = \bar{Y}_{l_1}^{a_T} \dots \bar{Y}_{l_v}^{a_T}$ y $\psi : \mathcal{G}^2 \times \mathbb{N}^2 \rightarrow \mathbb{R}^+$ definida por $\psi(f, g, u, v) = (2|\mathbf{S}|)^{u+v} \phi(u, v)$ con ϕ cualquiera de las 4 funciones dadas en H3.

Luego, estamos bajo las condiciones del Teorema A.33 con $K = (2|\mathbf{S}|)^2$ y $M = 2|\mathbf{S}|$ por lo que en (3.10) tenemos que

$$\begin{aligned} P(A_T) &\leq P\left(\sum_{i=1}^T \bar{Y}_t^{a_T} > C_1 \left(\frac{k_T}{v_T}\right)\right) \\ &\leq \exp\left(-\frac{\left(\frac{k_T C_1}{v_T}\right)^2 / 2}{\Sigma_T + \Delta_T^{1/(\mu+2)} \left(\frac{k_T C_1}{v_T}\right)^{(2\mu+3)/(\mu+2)}}\right), \end{aligned} \quad (3.12)$$

donde Σ_T puede ser elegido como cualquier número mayor o igual a $\sigma_T^2 \doteq \text{var}\left(\sum_{t=1}^T \bar{Y}_t^{a_T}\right)$ y

$$\Delta_T = 2(1 \vee 2|\mathbf{S}|)L_2 \left(\left(\frac{2^{4+\mu} T^2 L_1}{\Sigma_T} \right) \vee 1 \right).$$

Del Lema 3.14 (ii) para $c = a_T$ tenemos que $\sigma_T^2 \leq C_2 T$ para alguna constante C_2 luego, tomando $\Sigma_T = C_2 T$ tenemos que $\Sigma_T \geq \sigma_T^2$ y

$$\begin{aligned} \Delta_T &= 2(1 \vee 2|\mathbf{S}|)L_2 \left(\left(\frac{2^{4+\mu} T L_1}{\Sigma_T} \right) \vee 1 \right) \\ &= 2(1 \vee 2|\mathbf{S}|)L_2 \left(\left(\frac{2^{4+\mu} T L_1}{C_2 T} \right) \vee 1 \right) \\ &= 2(1 \vee 2|\mathbf{S}|)L_2 \left(\left(\frac{2^{4+\mu} L_1}{C_2} \right) \vee 1 \right) \\ &\doteq C_3. \end{aligned}$$

Por lo tanto, en el argumento de la exponencial de (3.12) tenemos que

$$\begin{aligned} \frac{\left(\frac{k_T C_1}{v_T}\right)^2 / 2}{\Sigma_T + \Delta_T^{1/(\mu+2)} \left(\frac{k_T C_1}{v_T}\right)^{(2\mu+3)/(\mu+2)}} &= \frac{\left(\frac{k_T C_1}{v_T}\right)^2 / 2}{TC_2 + C_3^{1/(\mu+2)} \left(\frac{k_T C_1}{v_T}\right)^{(2\mu+3)/(\mu+2)}} \\ &= \frac{k_T^2 C_4}{v_T^2 TC_2 + v_T^2 C_5 \left(\frac{k_T}{v_T}\right)^{(2\mu+3)/(\mu+2)}}. \end{aligned}$$

Aquí, $C_4 = C_1^2/2$ y $C_5 = C_3^{1/\mu+2} C_1^{(2\mu+3)/(\mu+2)}$. Luego, para todo $T \geq T_1(x)$ en (3.12) resulta

$$P(A_T) \leq \exp\left(-\frac{k_T^2 C_4}{v_T^2 TC_2 + v_T^2 C_5 \left(\frac{k_T}{v_T}\right)^{(2\mu+3)/(\mu+2)}}\right).$$

Finalmente, de esta desigualdad y de la hipótesis H4 tenemos que

$$\begin{aligned} \sum_{T=1}^{\infty} P(A_T) &= \sum_{T=1}^{T_1-1} P(A_T) + \sum_{T=T_1}^{\infty} P(A_T) \\ &\leq T_1 + \sum_{T=T_1}^{\infty} \exp\left(-\frac{k_T^2 C_4}{v_T^2 TC_2 + v_T^2 C_5 \left(\frac{k_T}{v_T}\right)^{(2\mu+3)/(\mu+2)}}\right) < \infty. \end{aligned}$$

□

3.3. Estimando de función de densidad de un campo no estacionario

En esta Sección extendemos el estimador dado en la Sección 3.2 a una familia particular de campos no estacionarios. Para ello, supongamos que en el modelo (3.1) la función media determinística $\mu(\mathbf{s})$ no es constante con respecto al espacio, esto es, supongamos que $\mathcal{X}(\mathbf{s})$ es dado por

$$\mathcal{X}(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s}), \quad \mathbf{s} \in \mathbf{S}$$

donde $\mu(\mathbf{s})$ es la función media y $e(\mathbf{s})$ es un campo aleatorio estacionario de primer orden con media cero y función de densidad no conocida f_e . La función de densidad de $\mathcal{X}(\mathbf{s})$ será denotada por $f_{\mathcal{X}(\mathbf{s})} \doteq f_{\mathcal{X}_{\mathbf{S}}}$.

3.3.1. Extensión del estimador

Sea $\{\mathcal{X}_t(\mathbf{s})\}_{t=1}^T$ una sucesión de campos aleatorios con la misma distribución que $\mathcal{X}(\mathbf{s})$. Igual que en el caso de datos funcionales, el estimador de la función de densidad $f_{\mathcal{X}_{\mathbf{S}}}$ es dado por

$$\widehat{f}_{\mathcal{X}_{\mathbf{S}}}(x) = \widehat{f}_u(x - \bar{\mathcal{X}}_T(\mathbf{s})), \quad (3.13)$$

con \widehat{f}_u definida como

$$\widehat{f}_u(x) \doteq \frac{k_T}{2T|\mathbf{S}|h_T^u(x)}$$

con $u = \{\mathcal{U}_{T1}, \dots, \mathcal{U}_{TT}\}$ dado por $\mathcal{U}_{Tt}(\mathbf{s}) = \mathcal{X}_t(\mathbf{s}) - \bar{\mathcal{X}}_T(\mathbf{s}) = e_t(\mathbf{s}) - \bar{e}_T(\mathbf{s})$. Aquí $\{e_1(\mathbf{s}), \dots, e_T(\mathbf{s})\}$ es una sucesión con la misma distribución que $e(\mathbf{s})$, $\bar{e}_T(\mathbf{s}) = \frac{1}{T} \sum_{t=1}^T e_t(\mathbf{s})$ y h_T^u es definida como en (3.5) reemplazando $\{e_t(\mathbf{s})\}_{t=1}^T$ por u .

Para todo \mathbf{s} fijo, las variables aleatorias $\{\mathcal{U}_{T1}(\mathbf{s}), \dots, \mathcal{U}_{TT}(\mathbf{s})\}$ con $\mathbb{E}(\mathcal{U}_{Tt}(\mathbf{s})) = 0$ son idénticamente distribuidas pero no necesariamente (α, ϕ) -débilmente dependiente. Por lo tanto, no podemos usar directamente los resultados de la Sección 3.2. Sin embargo, todavía podemos probar la convergencia completa del estimador de $f_{\mathcal{X}_{\mathbf{S}}}$ y obtener velocidades de convergencia.

3.3.2. Resultados asintóticos

Teorema 3.15. Velocidades de convergencia. *Para \mathbf{s} fijo, supongamos que H1-H4 se verifican para $e(\mathbf{s})$. Elijamos dos sucesiones $\{k_T\}$ y $\{v_T\}$ de números reales positivos que tienden*

a infinito tales que $v_T(T/k_T)|\bar{e}_T(\mathbf{s})| \xrightarrow{a.co.} 0$. Para esas sucesiones $\{k_T\}$ y $\{v_T\}$ supongamos que H5 se verifica. Entonces, para todo $x \in \mathbb{R}$,

$$\lim_{T \rightarrow \infty} v_T \left(\widehat{f}_{\mathcal{X}_{\mathbf{s}}}(x) - f_{\mathcal{X}_{\mathbf{s}}}(x) \right) = 0 \quad a.co.$$

Observación 3.16. Si para \mathbf{s} fijo la sucesión $\{e_t(\mathbf{s})\}_{t=1}^T$ es α -mixing geométrica con $|e(\mathbf{s})| < M$ para alguna constante $M > 0$, dado que en este caso $\bar{e}_T(\mathbf{s}) = o(T^{-\alpha})$ con $\alpha < 1/2$ entonces podemos elegir k_T tal que $v_T = T^\gamma$ para cualquier $\gamma < \frac{1}{4}$. Más precisamente, sean $k_T = T^\beta$ y $v_T = T^\gamma$. Para que las condiciones $(k_T/T)v_T = o(1)$ y $\frac{a_1 k_T^2}{a_1 v_T^2 T + a_2 v_T^2 \left(\frac{k_T}{v_T}\right)^{(2\mu+3)/(\mu+2)}} = \frac{a_1}{a_1 v_T^2 T k_T^{-2} + a_2 v_T^{\frac{\mu+2}{\mu+2}} k_T^{\frac{\mu+2}{\mu+2}}} \rightarrow \infty$ se cumplan es suficiente que $\beta < 1 - \gamma$ y $\beta > \gamma + \frac{1}{2}$. Además, para que $v_T(T/k_T)|\bar{e}_T(\mathbf{s})| \rightarrow 0$ a.co. se verifique, dado que $\bar{e}_T(\mathbf{s}) = o(T^{-\alpha})$ con $\alpha < 1/2$ debe ser $\beta > \gamma + \frac{1}{2}$. Por lo tanto, dado $\gamma < \frac{1}{4}$ podemos elegir β tal que las condiciones se cumplan (ver Figura 1).

La demostración del Teorema 3.15 será consecuencia inmediata del Teorema 3.11 y del siguiente lema:

Lema 3.17. *Supongamos que H1-H4 se verifican para $e(\mathbf{s})$. Elijamos dos sucesiones $\{k_T\}$ y $\{v_T\}$ de números reales positivos que tienden a infinito tales que $v_T(T/k_T)|\bar{e}_T(\mathbf{s})| \rightarrow 0$ a.co. Para esas sucesiones $\{k_T\}$ y $\{v_T\}$ supongamos que H5 se verifica. Entonces, para todo $x \in \mathbb{R}$, tenemos que*

$$\lim_{n \rightarrow \infty} v_T \left(\widehat{f}_u(x - \bar{\mathcal{X}}_T(\mathbf{s})) - \widehat{f}_e(x - \mu(\mathbf{s})) \right) = 0, \quad a.co.$$

donde \widehat{f}_e es el estimador de f_e .

La demostración de este lema es idéntica a la del Lema 2.17 en la cual no hemos usado la independencia de los datos, por lo tanto aquí será omitida.

3.4. Estudios de simulación

En esta sección presentamos algunos estudios de simulación que muestran el desempeño de nuestros métodos de estimación de la densidad marginal para campos aleatorios definidos en \mathbb{R}^2 . Más precisamente, consideremos el campo aleatorio $\mathcal{X}(\mathbf{s}) : \mathbb{R}^2 \rightarrow \mathbb{R}$ definido por

$$\mathcal{X}(\mathbf{s}) = \mu(\mathbf{s}) + e(\mathbf{s}), \quad \mathbf{s} \in \mathbf{S} = (0, 1] \times (0, 1], \quad (3.14)$$

donde $e(\mathbf{s})$ es un campo aleatorio estacionario Gaussiano con $\mathbb{E}(e(\mathbf{s})) = 0$, $\text{var}(e(\mathbf{s})) = 1$ y $\text{cov}(e(\mathbf{s}), e(\mathbf{r})) = \exp\{-d(\mathbf{s}, \mathbf{r})/0.25\}$ donde $d(\mathbf{s}, \mathbf{r})$ es la distancia del punto \mathbf{s} al punto \mathbf{r} .

Ejemplo 3.18. En este primer ejemplo consideraremos el caso estacionario. Para ello en el modelo (3.14) consideramos $\mu(\mathbf{s}) \equiv 0$ de manera que $\mathcal{X}(\mathbf{s}) = e(\mathbf{s})$ es estacionario y para cada \mathbf{s} fijo, $\mathcal{X}(\mathbf{s}) \sim \mathcal{N}(0, 1)$. En este caso tomamos una muestra de $n = 100$ campos aleatorios medido en una grilla de 100×100 puntos igualmente espaciados en el rectángulo $(0, 1] \times (0, 1]$. Dichos procesos fueron generados en el lenguaje de programación **R** (ver [R Development Core Team, 2006]) por el paquete para el análisis de datos geoestadísticos **GeoR** (ver [Ribeiro Jr. and Diggle, 2001]) y luego, todo el análisis fue realizado usando **Matlab** (ver [MatLab, 2009]).

En la Figura 2 (a) hemos graficado un elemento de una muestra de 100 campos con la misma distribución que $\mathcal{X}(\mathbf{s})$ y en la Figura 2 (b) hemos graficado la función de densidad teórica de $\mathcal{X}(\mathbf{s})$ y su estimador calculado con esa muestra. Como podemos observar, al igual que ocurría para datos funcionales, el estimador ajusta muy bien a la verdadera función de densidad excepto en las colas donde, debido a la naturaleza de los procesos, no tenemos suficiente información para realizar la estimación.

Ejemplo 3.19. En este ejemplo mostramos el desempeño de nuestro método cuando los datos no son estacionarios. Para ello, en el modelo (3.14) consideramos como función media a $\mu(\mathbf{s}) = \mu(s_1, s_2) = s_1 + s_2$. El buen comportamiento de nuestro estimador puede observarse en la Figura 3 donde hemos graficado para algunos puntos del $(0, 1] \times (0, 1]$, el estimador de la densidad (línea de puntos) y la densidad teórica (línea sólida) de $\mathcal{X}(\mathbf{s})$.

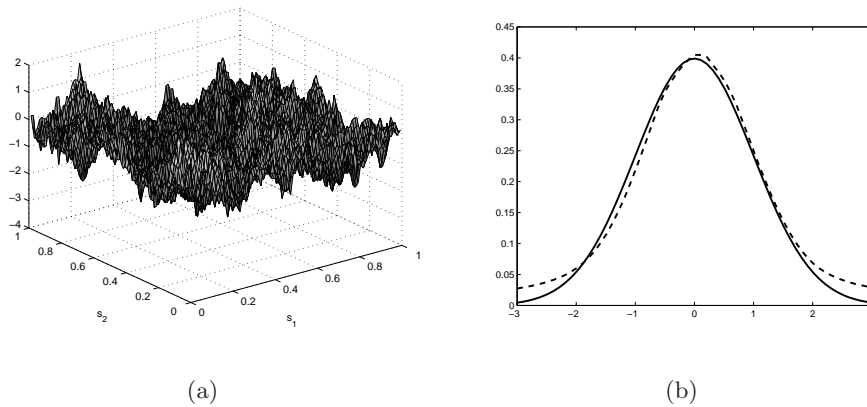


FIGURA 2. (a) 1 elemento de la muestra de tamaño 100 de $\mathcal{X}(\mathbf{s})$. (b) Función de densidad estimada (línea de puntos) y teórica (línea sólida) de $\mathcal{X}(\mathbf{s})$ para $k_n = 10$.

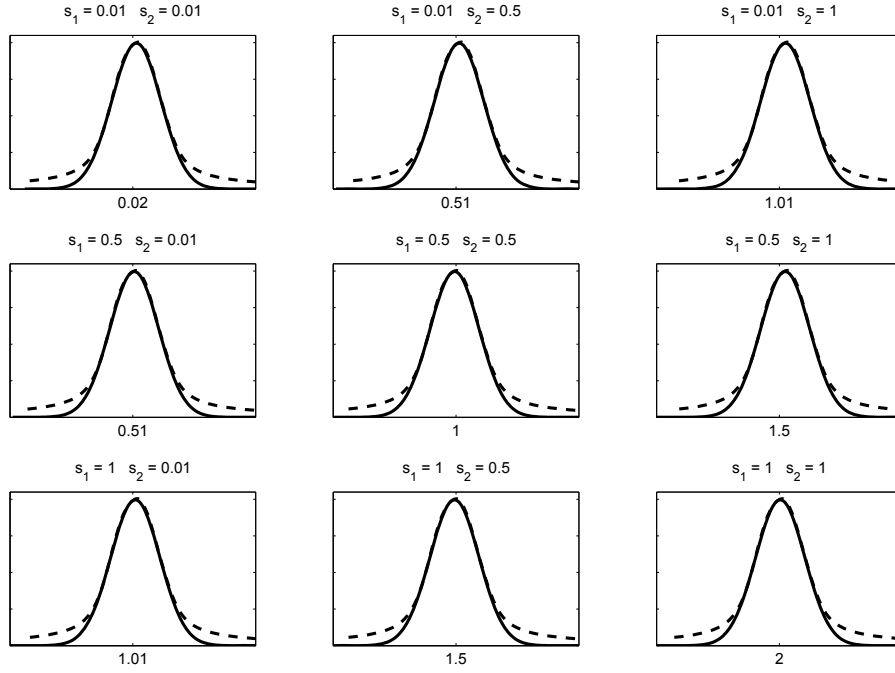


FIGURA 3. Función de densidad estimada (línea de puntos) y teórica (línea sólida) de $\mathcal{X}(\mathbf{s})$ para $\mu(s_1, s_2) = s_1 + s_2$ y $k_n = 15$.

3.5. Apéndice

Demostración del Lema 3.7: Supongamos que $\alpha(r) \leq a\rho^r$ para algunos $0 < \rho < 1$ y $a > 0$ y si $f^{(k)}$ indica derivada k -ésima de f entonces,

$$\begin{aligned}
 \sum_{j=0}^{\infty} (j+1)^k \alpha(j) &\leq a \sum_{j=0}^{\infty} (j+1)^k \rho^j \\
 &\leq a \sum_{j=0}^{\infty} (j+k)(j+k-1)\dots(j+1)\rho^j && \text{(para } k \geq 1) \\
 &= a \left(\sum_{j=0}^{\infty} \rho^{j+k} \right)^{(k)} \\
 &= a \left(\rho^k \sum_{j=0}^{\infty} \rho^j \right)^{(k)} \\
 &= a \left(\rho^k \frac{1}{1-\rho} \right)^{(k)}.
 \end{aligned} \tag{3.15}$$

Ahora, dado que

$$(fg)^{(k)} = \sum_{j=0}^k \binom{k}{j} f^{(k-j)} g^{(j)}, \quad (\rho^k)^{(j)} = \frac{k!}{(k-j)!} \rho^{k-j} \quad \text{y} \quad \left(\frac{1}{1-\rho} \right)^{(j)} = \frac{j!}{(1-\rho)^{j+1}}$$

resulta,

$$\begin{aligned}
a \left(\rho^k \frac{1}{1-\rho} \right)^{(k)} &= a \sum_{j=0}^k \binom{k}{j} (\rho^k)^{(k-j)} \left(\frac{1}{1-\rho} \right)^{(j)} \\
&= a \sum_{j=0}^k \binom{k}{j} \frac{k!}{j!} \rho^j \frac{j!}{(1-\rho)^{j+1}} \\
&= a \frac{k!}{1-\rho} \sum_{j=0}^k \binom{k}{j} \left(\frac{\rho}{1-\rho} \right)^j 1^{k-j} \\
&= a \frac{k!}{1-\rho} \left(\frac{\rho}{1-\rho} + 1 \right)^k && \text{(Teo. del Binomio)} \\
&= \frac{a}{1-\rho} \left(\frac{1}{1-\rho} \right)^k k!.
\end{aligned}$$

Con esta igualdad en (3.15) tenemos que

$$\sum_{j=0}^{\infty} (j+1)^k \alpha(j) \leq \frac{a}{1-\rho} \left(\frac{1}{1-\rho} \right)^k k!$$

y por tanto, tomando $L_1 = \frac{a}{1-\rho}$, $L_2 = \frac{1}{1-\rho}$ y $\mu = 1$ tenemos el resultado. □

Demostración del Lema 3.14: Para $1 \leq t_1 \leq \dots \leq t_u < t_u + r = l_1 \leq \dots \leq l_v \leq T$, consideremos la u -upla (t_1, \dots, t_u) y la v -upla (l_1, \dots, l_v) . Dado que por el Lema 2.6 $\mathbb{E}(Y^c) = p_c |S|$ entonces,

$$\prod_{k=1}^u \bar{Y}_{t_k}^c = \prod_{k=1}^u (Y_{t_k}^c - p_c |S|) \doteq \sum_{k=0}^u \binom{u}{k} (-p_c |S|)^{u-k} \prod_{i=1}^k Y_{t_i}^c.$$

Por lo tanto,

$$\begin{aligned}
\text{cov}(\bar{Y}_{t_1}^c \dots \bar{Y}_{t_u}^c, \bar{Y}_{l_1}^c \dots \bar{Y}_{l_v}^c) &= \text{cov} \left(\prod_{k=1}^u \bar{Y}_{t_k}^c, \prod_{m=1}^v \bar{Y}_{l_m}^c \right) \\
&= \text{cov} \left(\sum_{k=1}^u \binom{u}{k} (-p_c |S|)^{u-k} \prod_{i=1}^k Y_{t_i}^c, \sum_{m=1}^v \binom{v}{m} (-p_c |S|)^{v-m} \prod_{j=1}^m Y_{l_j}^c \right) \\
&= \sum_{k=1}^u \binom{u}{k} \sum_{m=1}^v \binom{v}{m} (-p_c |S|)^{u+v-(k+m)} \text{cov} \left(\prod_{i=1}^k Y_{t_i}^c, \prod_{j=1}^m Y_{l_j}^c \right).
\end{aligned}$$

Tomando valor absoluto y teniendo en cuenta que $|p_c| \leq 1$ tenemos que

$$|\text{cov}(\bar{Y}_{t_1}^c \dots \bar{Y}_{t_u}^c, \bar{Y}_{l_1}^c \dots \bar{Y}_{l_v}^c)| \leq \sum_{k=1}^u \binom{u}{k} \sum_{m=1}^v \binom{v}{m} |S|^{u+v-(k+m)} \left| \text{cov} \left(\prod_{i=1}^k Y_{t_i}^c, \prod_{j=1}^m Y_{l_j}^c \right) \right| \quad (3.16)$$

donde

$$\left| \text{cov} \left(\prod_{i=1}^k Y_{t_i}^c, \prod_{j=1}^m Y_{l_j}^c \right) \right| = \left| \mathbb{E} \left(\prod_{i=1}^k Y_{t_i}^c \prod_{j=1}^m Y_{l_j}^c \right) - \mathbb{E} \left(\prod_{i=1}^k Y_{t_i}^c \right) \mathbb{E} \left(\prod_{j=1}^m Y_{l_j}^c \right) \right|. \quad (3.17)$$

En lo que sigue consideraremos los siguientes eventos,

$$S_{t_i}(\mathbf{s}) \doteq \{ \omega \in \Omega : \mathcal{X}_{t_i}(\mathbf{s}, \omega) \in I_{(x,c)} \}, \quad i = 1, \dots, u$$

y

$$S_{l_j}(\mathbf{r}) \doteq \{ \omega \in \Omega : \mathcal{X}_{l_j}(\mathbf{r}, \omega) \in I_{(x,c)} \}, \quad j = 1, \dots, v.$$

Luego,

$$\begin{aligned} \mathbb{E} \left(\prod_{i=1}^k Y_{t_i}^c \prod_{j=1}^m Y_{l_j}^c \right) &= \mathbb{E} \left(\prod_{i=1}^k \left(\int_{\mathbf{s}} \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_i}(\mathbf{s})) d\mathbf{s} \right) \prod_{j=1}^m \left(\int_{\mathbf{r}} \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_j}(\mathbf{r})) d\mathbf{r} \right) \right) \\ &= \mathbb{E} \left(\left(\int_{\mathbf{s}^u} \prod_{i=1}^k \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_i}(\mathbf{s})) d\mathbf{s}^u \right) \left(\int_{\mathbf{r}^v} \prod_{j=1}^m \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_j}(\mathbf{r})) d\mathbf{r}^v \right) \right) \\ &= \mathbb{E} \left(\int_{\mathbf{s}^u} \int_{\mathbf{r}^v} \prod_{i=1}^k \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_i}(\mathbf{s})) \prod_{j=1}^m \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_j}(\mathbf{s})) d\mathbf{s}^u d\mathbf{r}^v \right) \\ &= \int_{\mathbf{s}^u} \int_{\mathbf{r}^v} \mathbb{E} \left(\prod_{i=1}^k \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_i}(\mathbf{s})) \prod_{j=1}^m \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_j}(\mathbf{s})) \right) d\mathbf{s}^u d\mathbf{r}^v. \end{aligned} \quad (\text{Teo. de Tonelli})$$

Dado que,

$$\begin{aligned} \mathbb{E} \left(\prod_{i=1}^k \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_i}(\mathbf{s})) \prod_{j=1}^m \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_j}(\mathbf{s})) \right) &= \\ &= \mathbb{E} \left(\mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_1}(\mathbf{s})) \dots \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_k}(\mathbf{s})) \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_1}(\mathbf{s})) \dots \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_m}(\mathbf{s})) \right) \\ &= P(\mathcal{X}_{t_1}(\mathbf{s}) \in I_{(x,c)} \wedge \dots \wedge \mathcal{X}_{t_k}(\mathbf{s}) \in I_{(x,c)} \wedge \mathcal{X}_{l_1}(\mathbf{s}) \in I_{(x,c)} \wedge \dots \wedge \mathcal{X}_{l_m}(\mathbf{s}) \in I_{(x,c)}) \\ &= P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \cap \bigcap_{j=1}^m S_{l_j}(\mathbf{s}) \right) \end{aligned}$$

entonces,

$$\begin{aligned} \mathbb{E} \left(\prod_{i=1}^k Y_{t_i}^c \prod_{j=1}^m Y_{l_j}^c \right) &= \int_{\mathbf{s}^u} \int_{\mathbf{r}^v} \mathbb{E} \left(\prod_{i=1}^k \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{t_i}(\mathbf{s})) \prod_{j=1}^m \mathbb{I}_{I_{(x,c)}}(\mathcal{X}_{l_j}(\mathbf{s})) \right) d\mathbf{s}^u d\mathbf{r}^v \\ &= \int_{\mathbf{s}^u} \int_{\mathbf{r}^v} P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \cap \bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) d\mathbf{s}^u d\mathbf{r}^v. \end{aligned} \quad (3.18)$$

Por otro lado,

$$\begin{aligned}
\mathbb{E} \left(\prod_{i=1}^k Y_{t_i}^c \right) \mathbb{E} \left(\prod_{j=1}^m Y_{l_j}^c \right) &= \mathbb{E} \left(\prod_{i=1}^k \int_{\mathbf{S}} \mathbb{I}_{I(x,c)}(\mathcal{X}_{t_i}(\mathbf{s})) d\mathbf{s} \right) \mathbb{E} \left(\prod_{j=1}^m \int_{\mathbf{S}} \mathbb{I}_{I(x,c)}(\mathcal{X}_{l_j}(\mathbf{r})) d\mathbf{r} \right) \\
&= \mathbb{E} \left(\int_{\mathbf{S}^u} \prod_{i=1}^k \mathbb{I}_{I(x,c)}(\mathcal{X}_{t_i}(\mathbf{s})) d\mathbf{s}^u \right) \mathbb{E} \left(\int_{\mathbf{S}^v} \prod_{j=1}^m \mathbb{I}_{I(x,c)}(\mathcal{X}_{l_j}(\mathbf{r})) d\mathbf{r}^v \right) \quad (3.19) \\
&= \left(\int_{\mathbf{S}^u} \mathbb{E} \left(\prod_{i=1}^k \mathbb{I}_{I(x,c)}(\mathcal{X}_{t_i}(\mathbf{s})) \right) d\mathbf{s}^u \right) \left(\int_{\mathbf{S}^v} \mathbb{E} \left(\prod_{j=1}^m \mathbb{I}_{I(x,c)}(\mathcal{X}_{l_j}(\mathbf{r})) \right) d\mathbf{r}^v \right) \\
&= \int_{\mathbf{S}^u} \int_{\mathbf{S}^v} \mathbb{E} \left(\prod_{i=1}^k \mathbb{I}_{I(x,c)}(\mathcal{X}_{t_i}(\mathbf{s})) \right) \mathbb{E} \left(\prod_{j=1}^m \mathbb{I}_{I(x,c)}(\mathcal{X}_{l_j}(\mathbf{r})) \right) d\mathbf{s}^u d\mathbf{r}^v \quad (\text{Tonelli}) \\
&= \int_{\mathbf{S}^u} \int_{\mathbf{S}^v} P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \right) P \left(\bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) d\mathbf{s}^u d\mathbf{r}^v.
\end{aligned}$$

Por lo tanto, con (3.18) y (3.19) en (3.17) tenemos que

$$\begin{aligned}
\left| \text{cov} \left(\prod_{i=1}^k Y_{t_i}^c, \prod_{j=1}^m Y_{l_j}^c \right) \right| &= \left| \mathbb{E} \left(\prod_{i=1}^k Y_{t_i}^c \prod_{j=1}^m Y_{l_j}^c \right) - \mathbb{E} \left(\prod_{i=1}^k Y_{t_i}^c \right) \mathbb{E} \left(\prod_{j=1}^m Y_{l_j}^c \right) \right| \\
&= \left| \int_{\mathbf{S}^u} \int_{\mathbf{S}^v} P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \cap \bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) d\mathbf{s}^u d\mathbf{r}^v - \int_{\mathbf{S}^u} \int_{\mathbf{S}^v} P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \right) P \left(\bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) d\mathbf{s}^u d\mathbf{r}^v \right| \\
&\leq \int_{\mathbf{S}^u} \int_{\mathbf{S}^v} \left| P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \cap \bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) - P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \right) P \left(\bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) \right| d\mathbf{s}^u d\mathbf{r}^v.
\end{aligned}$$

Observemos que para cada \mathbf{s} fijo, dado que $k \leq u$ y $m \leq v$,

$$\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \in \mathcal{M}_{t_1}^{t_u} \quad \text{y} \quad \bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \in \mathcal{M}_{l_1}^{l_v},$$

con $1 \leq t_1 \leq \dots \leq t_u < t_u + r = l_1 \leq \dots \leq l_v \leq T$. Luego, de la hipótesis H4 se tiene que

$$\begin{aligned}
\left| \text{cov} \left(\prod_{i=1}^k Y_{t_i}^c, \prod_{j=1}^m Y_{l_j}^c \right) \right| &\leq \int_{\mathbf{S}^u} \int_{\mathbf{S}^v} \left| P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \cap \bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) - P \left(\bigcap_{i=1}^k S_{t_i}(\mathbf{s}) \right) P \left(\bigcap_{j=1}^m S_{l_j}(\mathbf{r}) \right) \right| d\mathbf{s}^u d\mathbf{r}^v \\
&\leq \int_{\mathbf{S}^u} \int_{\mathbf{S}^v} \alpha(r) \phi(u, v) d\mathbf{s}^u d\mathbf{r}^v \quad (3.20) \\
&= |\mathbf{S}|^{u+v} \alpha(r) \phi(u, v).
\end{aligned}$$

Luego, con esta desigualdad en (3.16) resulta

$$\left| \text{cov} \left(\prod_{i=1}^k \bar{Y}_{t_i}^c, \prod_{j=1}^m \bar{Y}_{l_j}^c \right) \right| \leq \sum_{k=1}^u \binom{u}{k} \sum_{m=1}^v \binom{v}{m} |\mathbf{S}|^{u+v-(k+m)} \left| \text{cov} \left(\prod_{i=1}^k Y_{t_i}^c, \prod_{j=1}^m Y_{l_j}^c \right) \right|$$

$$\begin{aligned}
&\leq \sum_{k=1}^u \binom{u}{k} \sum_{m=1}^v \binom{v}{m} |\mathbf{S}|^{u+v-(k+m)} |\mathbf{S}|^{k+m} \alpha(l_1 - t_i) \phi(u, v) \quad (\text{de (3.20)}) \\
&= |\mathbf{S}|^{u+v} \sum_{k=1}^u \binom{u}{k} \sum_{m=1}^v \binom{v}{m} \alpha(l_1 - t_i) \phi(u, v) \\
&\leq |\mathbf{S}|^{u+v} \sum_{k=1}^u \binom{u}{k} \sum_{m=1}^v \binom{v}{m} \alpha(r) \phi(u, v) \quad (t_i \leq t_u \Rightarrow l_1 - t_i > r \text{ y } \alpha \searrow 0) \\
&= |\mathbf{S}|^{u+v} (2^u - 1)(2^v - 1) \alpha(r) \phi(u, v) \quad (\text{Cor. A.13}) \\
&\leq |\mathbf{S}|^{u+v} 2^u 2^v \alpha(r) \phi(u, v) \\
&= (2|\mathbf{S}|)^{u+v} \alpha(r) \phi(u, v).
\end{aligned}$$

Probemos ahora la segunda parte del Lema.

$$\begin{aligned}
\text{var} \left(\sum_{t=1}^T \bar{Y}_t^c \right) &\leq \sum_{t=1}^T \text{var} (Y_t^c) + 2 \sum_{t=1}^{T-1} \sum_{l=t+1}^T |\text{cov} (\bar{Y}_t^c, \bar{Y}_l^c)| \\
&\leq \sum_{t=1}^T |\mathbf{S}|^2 p_c + 2^3 \sum_{t=1}^{T-1} \sum_{l=t+1}^T |\mathbf{S}|^2 \alpha(l-t) \phi(1, 1) \quad (\text{Lema 2.6 y (3.11)}) \\
&= T |\mathbf{S}|^2 p_c + 2^3 |\mathbf{S}|^2 \sum_{t=1}^{T-1} \sum_{l=t+1}^T \alpha(l-t) \phi(1, 1) \quad (3.21) \\
&\leq T |\mathbf{S}|^2 + 2^3 |\mathbf{S}|^2 \phi(1, 1) \sum_{t=1}^{T-1} \sum_{l=t+1}^T \alpha(l-t). \quad (p_c \leq 1)
\end{aligned}$$

Analícemos la última suma. Llamando $s = l - t$ tenemos que

$$\begin{aligned}
\sum_{t=1}^{T-1} \sum_{l=t+1}^T \alpha(l-t) &= \sum_{t=1}^{T-1} \sum_{s=1}^{T-t} \alpha(s) \\
&= \sum_{s=1}^{T-1} \sum_{t=1}^{T-s} \alpha(s) \quad (s \leq T-t \Rightarrow t \leq T-s) \\
&\leq \sum_{s=1}^{T-1} \alpha(s) (T-1) \quad (1 \leq s \leq T-1 \Rightarrow 1 \leq T-s \leq T-1) \\
&\leq T \sum_{s=0}^{\infty} \alpha(s) \quad (\alpha > 0) \\
&\leq TL_1. \quad (\text{de (3.3) para } k=0)
\end{aligned}$$

Con esta desigualdad en (3.21) tenemos que

$$\text{var} \left(\sum_{t=1}^T \bar{Y}_t^c \right) \leq T |\mathbf{S}|^2 + 2^3 |\mathbf{S}|^2 \phi(1, 1) \sum_{t=1}^{T-1} \sum_{l=t+1}^T \alpha(l-t)$$

$$\leq T|\mathbf{S}|^2 + 2^3|\mathbf{S}|^2\phi(1,1)TL_1$$

$$= TC.$$

$$(C \doteq |\mathbf{S}|^2 (1 + 2^3\phi(1,1)L_1))$$

□

REGRESIÓN PARA DATOS FUNCIONALES

En el contexto de datos funcionales, el problema de estimar la función de regresión de manera paramétrica usando modelos lineales, ha sido de cierta manera popularizado por Ramsay y Silverman ([**Ramsay and Silverman, 1997**]). Desde entonces, muchos autores han trabajado en este tema, entre ellos podemos mencionar a aquellos que han obtenido resultados más recientes como por ejemplo, [**Cuevas et al., 2002**], [**Cardot et al., 2003**], [**Cardot et al., 2004**], [**Hall and Horowitz, 2007**], [**Preda and Saporta, 2005a**], [**Preda and Saporta, 2005b**], [**Baïllo and Grané, 2009**] y [**Preda et al., 2007**].

Dentro del marco no paramétrico, existen varios resultados sobre consistencia de estimadores de la función de regresión aunque la mayoría son para estimadores concretos (estimador de k -vecinos más cercanos y estimador de núcleo) y en algunos casos sólo para el problema de clasificación. Para el estimador de núcleo, en el clásico libro *Nonparametric functional data analysis*, sus autores [**Ferraty and Vieu, 2006**] obtienen resultados de consistencia casi completa bajo hipótesis de continuidad de la función de regresión η , sobre la distribución subyacente de los procesos (*small ball probability*), regularidad del núcleo y acotación de los momentos de la respuesta. Además, si la función de regresión es Lipchitz, obtienen velocidades de convergencia.

Para estimadores de k -vecinos más cercanos, [**Biau et al., 2005**] probaron la consistencia débil de la respectiva regla de clasificación en espacios de Hilbert separables cuando la respuesta es binaria (η acotada). Para ello, utilizaron el método de filtrado para reducir la dimensión del espacio a \mathbb{R}^d , y así poder utilizar el clásico Teorema de consistencia de Stone finito dimensional (ver [**Stone, 1977**]). Aunque ellos afirman que han extendido este importante teorema al contexto infinito dimensional, sólo han utilizado su versión original en \mathbb{R}^d . Por otro lado, para espacios métricos separables, [**Cèrou and Guyader, 2006**] probaron la consistencia débil de

esa regla de clasificación bajo cierta condición de regularidad de la función de regresión η con respecto a μ , la medida de probabilidad de \mathcal{X} , conocida como la *condición de Besicovitch*: para todo $\epsilon > 0$,

$$\lim_{\delta \rightarrow 0} \mu \left(\mathcal{X} \in \mathcal{F} : \frac{1}{\mu(\mathcal{B}(\mathcal{X}, \delta))} \int_{\mathcal{B}(\mathcal{X}, \delta)} |\eta(z) - \eta(\mathcal{X})| d\mu(z) > \epsilon \right) = 0.$$

[**Abraham et al., 2006**] mostraron que el estimador de ventana móvil (suma de indicadoras sobre bandas que varían con n) no es universalmente fuertemente consistente para espacios métricos en general y dieron condiciones suficientes sobre el espacio y la función de regresión para asegurar la consistencia fuerte del estimador.

Recientemente, [**Biau et al., 2010**] presentaron contraejemplos que muestran que el estimador k -NN no es consistente para espacios generales \mathcal{F} . Sin embargo, demostraron su consistencia en media cuadrática y encontraron tasas de convergencia en espacios de Banach separables, usando la compleja teoría de inclusiones compactas.

En este capítulo presentamos un resultado de consistencia para una familia muy general de estimadores de regresión que incluye al estimador de k -vecinos más cercanos y al de núcleo. Luego, suponiendo la separabilidad del espacio, utilizamos el resultado para probar la consistencia en media cuadrática de estos estimadores.

4.1. Preliminares

Sea (\mathcal{F}, d) un espacio métrico. Consideremos el par $(\mathcal{X}, Y) \in \mathcal{F} \times \mathbb{R}$ (con \mathcal{X} un proceso estocástico) que verifica el siguiente modelo de regresión,

$$Y = \eta(\mathcal{X}) + e, \tag{4.1}$$

donde el **error** es independiente de \mathcal{X} , $\mathbb{E}(e) = 0$ y $\mathbb{E}(e^2) = \sigma^2 < \infty$. En estas condiciones, $\eta(\mathcal{X}) = \mathbb{E}(Y|\mathcal{X}) : \mathcal{F} \rightarrow \mathbb{R}$ es el **operador de regresión**.

Dada una sucesión de pares aleatorios $\{(\mathcal{X}_i, Y_i)\}_{i=1}^n$ i.i.d. con la misma distribución que la del par (\mathcal{X}, Y) , estimaremos el operador de regresión η mediante

$$\hat{\eta}(\mathcal{X}) = \sum_{i=1}^n Y_i W_{ni}(\mathcal{X}) \tag{4.2}$$

donde, para cada i los pesos $W_{ni}(\mathcal{X}) = W_{ni}(\mathcal{X}, \mathcal{X}_1, \dots, \mathcal{X}_n)$ son no negativos y $\sum_{i=1}^n W_{ni}(\mathcal{X}) = 1$. Una sucesión de pesos $W_n = \{W_{ni}\}_{i=1}^n$ cuyos miembros satisfagan esta última condición será llamada sucesión de **pesos normales**. Además, una sucesión de pesos no negativos y normales,

será llamada sucesión de **pesos de probabilidad** (estas definiciones fueron introducidas por [Stone, 1977]).

Notación 4.1.

- Por simplicidad de notación llamaremos \mathfrak{D}_n a la sucesión $\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$.
- Los símbolos \mathbb{E} , var o P sin subíndice, indicarán que estamos calculando la esperanza, varianza o probabilidad sobre $\mathcal{X}, \mathcal{X}_1, \dots, \mathcal{X}_n$. En caso contrario el subíndice indicará con respecto a que elemento aleatorio estamos calculando dichas medidas.
- Utilizaremos la notación simplificada $p_r(\mathcal{X})$ para referirnos a la medida de probabilidad de la bola $\mathcal{B}(\mathcal{X}, r)$. Esto es, $p_r(\mathcal{X}) = \mu(\mathcal{B}(\mathcal{X}, r))$ donde μ es la medida de probabilidad de \mathcal{X} .

4.2. Extensión del Teorema de Stone

En esta sección extenderemos el resultado de [Stone, 1977] al contexto infinito dimensional.

Teorema 4.2. *Sea (\mathcal{F}, d) un espacio métrico. Sea $\{(\mathcal{X}_i, Y_i)\}_{i=1}^n$ una muestra aleatoria independiente e idénticamente distribuída como el par $(\mathcal{X}, Y) \in \mathcal{F} \times \mathbb{R}$ que satisface el modelo (4.1) y supongamos que $W_n(\mathcal{X}) = \{W_{ni}(\mathcal{X})\}_{i=1}^n$ es una sucesión de pesos de probabilidad que satisface las siguientes hipótesis:*

(i) *para todo $a > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) = 0;$$

(ii)

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\max_{1 \leq i \leq n} W_{ni}(\mathcal{X}) \right) = 0;$$

(iii) *existe una constante $c > 0$ tal que, para toda función medible no negativa f que satisface*

$\mathbb{E}(f(\mathcal{X})) < \infty$ *se tiene que*

$$\mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) f(\mathcal{X}_i) \right) \leq c \mathbb{E}_{\mathcal{X}}(f(\mathcal{X})).$$

Si además la función de regresión η satisface la siguiente hipótesis:

(iv) *η puede aproximarse en $L^2(\mu)$ por funciones uniformemente continuas y acotadas.*

Entonces, el estimador dado en (4.2) es consistente en media cuadrática. Es decir,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left((\hat{\eta}(\mathcal{X}) - \eta(\mathcal{X}))^2 \right) = 0.$$

Corolario 4.3. Sean W_n una sucesión de pesos de probabilidad que satisface las hipótesis (i)-(iii) del Teorema 4.2 y η una función de regresión que verifica la hipótesis (iv) del mismo teorema. Si U_n es una sucesión de pesos normales tal que, para cada $n \geq 1$, $|U_n| \leq MW_n$ para alguna constante $M \geq 1$. Entonces, el estimador definido por

$$\hat{\eta}(\mathcal{X}) = \sum_{i=1}^n Y_i U_{ni}(\mathcal{X})$$

es consistente en media cuadrática.

Aquí, $|U_n| \leq MW_n$ significa que, para cada $n \geq 1$, $|U_{ni}| \leq MW_{ni}$ para todo $i = 1, \dots, n$.

Observación 4.4. (Algunos comentarios sobre las hipótesis)

- La condición (i) requiere que el peso total fuera de toda “bola” centrada en \mathcal{X} y de radio fijo tienda a cero. En otras palabras, requiere que la influencia en el estimador de los elementos \mathcal{X}_i que están muy alejados de \mathcal{X} sea pequeña.
- La condición (ii) requiere que ningún elemento \mathcal{X}_i tenga demasiada contribución en el estimador.
- Observemos que las hipótesis (i)-(iii) son las mismas que las del Teorema 1.4 ([Stone, 1977]).
- La hipótesis (ii) puede debilitarse pidiendo que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{i=1}^n W_{ni}^2(\mathcal{X}) \right) = 0.$$

- Observemos que si bien las condiciones (i)-(iii) son muy generales, no pedimos condiciones sobre la distribución de \mathcal{X} (por ejemplo, del tipo *small ball probability*). La condición (iv) requiere simplemente cierta regularidad de la función de regresión η .
- Si μ es una medida de Borel y el espacio \mathcal{F} es localmente compacto, la hipótesis (iv) es siempre cierta.

Demostración. Consideremos la función auxiliar $\eta_n(x) = \sum_{i=1}^n \eta(\mathcal{X}_i) W_{ni}(x)$. En base a la desigualdad $(a + b)^2 \leq 2(a^2 + b^2)$ tenemos que

$$\begin{aligned} \mathbb{E} ((\hat{\eta}(\mathcal{X}) - \eta(\mathcal{X}))^2) &= \mathbb{E} \left((\hat{\eta}(\mathcal{X}) - \eta_n(\mathcal{X}) + \eta_n(\mathcal{X}) - \eta(\mathcal{X}))^2 \right) \\ &\leq 2\mathbb{E} (\eta(\mathcal{X}) - \eta_n(\mathcal{X}))^2 + 2\mathbb{E} ((\eta_n(\mathcal{X}) - \hat{\eta}(\mathcal{X}))^2) \\ &\doteq I + II. \end{aligned}$$

Por lo tanto, será suficiente probar que I y II tienden a cero. Comencemos con I , para ello escribimos

$$\eta(\mathcal{X}) - \eta_n(\mathcal{X}) = \eta(\mathcal{X}) - \sum_{i=1}^n \eta(\mathcal{X}_i) W_{ni}(\mathcal{X}) = \sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta(\mathcal{X}) - \eta(\mathcal{X}_i)).$$

Dado que los pesos son no negativos y $\phi(x) = x^2$ es convexa, de la desigualdad de Jensen tenemos que

$$(\eta(\mathcal{X}) - \eta_n(\mathcal{X}))^2 = \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta(\mathcal{X}) - \eta(\mathcal{X}_i)) \right)^2 \leq \sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta(\mathcal{X}) - \eta(\mathcal{X}_i))^2$$

de donde se sigue que,

$$I = \mathbb{E} ((\eta(\mathcal{X}) - \eta_n(\mathcal{X}))^2) \leq \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta(\mathcal{X}) - \eta(\mathcal{X}_i))^2 \right). \quad (4.3)$$

Por la hipótesis (iv), para todo $\epsilon > 0$ existe η^* uniformemente continua y acotada tal que

$$\int (\eta(x) - \eta^*(x))^2 d\mu(x) = \mathbb{E} ((\eta(\mathcal{X}) - \eta^*(\mathcal{X}))^2) < \epsilon. \quad (4.4)$$

Para esta elección de η^* y usando la desigualdad $(a + b + c)^2 \leq 3(a^2 + b^2 + c^2)$ podemos escribir

$$\begin{aligned} I &= \mathbb{E} ((\eta(\mathcal{X}) - \eta_n(\mathcal{X}))^2) \\ &\leq \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta(\mathcal{X}) - \eta(\mathcal{X}_i))^2 \right) \\ &\leq 3\mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta(\mathcal{X}) - \eta^*(\mathcal{X}))^2 \right) \\ &\quad + 3\mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta^*(\mathcal{X}) - \eta^*(\mathcal{X}_i))^2 \right) \\ &\quad + 3\mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta^*(\mathcal{X}_i) - \eta(\mathcal{X}_i))^2 \right) \\ &\leq 3\epsilon + 3\mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta^*(\mathcal{X}) - \eta^*(\mathcal{X}_i))^2 \right) + 3c\epsilon, \end{aligned} \quad (4.5)$$

donde, en el primer término hemos usado (4.4) y en el tercer término hemos usado la hipótesis (iii) y nuevamente (4.4). Resta ver que ocurre con el segundo término. Dado que η^* es uniformemente continua para todo $\epsilon > 0$ existe $a > 0$ tal que si $d(\mathcal{X}, \mathcal{X}_i) \leq a$,

$$(\eta^*(\mathcal{X}) - \eta^*(\mathcal{X}_i))^2 < \epsilon$$

para todo $\mathcal{X}, \mathcal{X}_i$. Para ese a escribimos,

$$\begin{aligned}
\mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta^*(\mathcal{X}) - \eta^*(\mathcal{X}_i))^2 \right) &= \\
&= \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta^*(\mathcal{X}) - \eta^*(\mathcal{X}_i))^2 \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) \\
&\quad + \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta^*(\mathcal{X}) - \eta^*(\mathcal{X}_i))^2 \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq a\}} \right) \tag{4.6} \\
&\leq 2C^2 \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) + \epsilon \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) = 1 \text{ y } \eta^* \leq C \right) \\
&\leq (2C^2 + 1)\epsilon,
\end{aligned}$$

donde, en la última desigualdad hemos usado que por la hipótesis (i) existe N_1 tal que si $n \geq N_1$,

$$\mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) \leq \epsilon.$$

Reemplazando la desigualdad (4.6) en (4.5) resulta

$$I \leq 3\epsilon + 3(2C^2 + 1)\epsilon + 3c\epsilon.$$

Continuemos con *II*. Para ello escribimos,

$$\begin{aligned}
II &= \mathbb{E} \left((\eta_n(\mathcal{X}) - \hat{\eta}(\mathcal{X}))^2 \right) \\
&= \mathbb{E} \left(\left(\sum_{i=1}^n W_{ni}(\mathcal{X}) (\eta(\mathcal{X}_i) - Y_i) \right)^2 \right) \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} (W_{ni}(\mathcal{X}) W_{nj}(\mathcal{X}) e_i e_j) \tag{(\eta(\mathcal{X}_i) - Y_i = e_i)} \\
&= \sum_{i=1}^n \sum_{j=1}^n \mathbb{E} (W_{ni}(\mathcal{X}) W_{nj}(\mathcal{X})) \mathbb{E} (e_i e_j) \tag{(por hipótesis $e \perp \mathcal{X}$)} \\
&= \sum_{i=1}^n \mathbb{E} (W_{ni}^2(\mathcal{X})) \mathbb{E} (e_i^2) \tag{(e_i \perp e_j, \forall i \neq j \text{ y } \mathbb{E} (e_i) = 0, \forall i)} \\
&\leq \sigma^2 \mathbb{E} \left(\max_{1 \leq j \leq n} W_{nj}(\mathcal{X}) \sum_{i=1}^n W_{ni}(\mathcal{X}) \right) \\
&= \sigma^2 \mathbb{E} \left(\max_{1 \leq j \leq n} W_{nj}(\mathcal{X}) \right) \tag{\left(\sum_{i=1}^n W_{ni}(\mathcal{X}) = 1 \right)} \\
&< \sigma^2 \epsilon.
\end{aligned}$$

En la última desigualdad hemos usado que por la hipótesis (ii), para todo $\epsilon > 0$ existe N_2 tal que si $n \geq N_2$,

$$\mathbb{E} \left(\max_{1 \leq j \leq n} W_{nj}(\mathcal{X}) \right) < \epsilon.$$

Por lo tanto, tomando $n \geq \max\{N_1, N_2\}$ el teorema queda demostrado. \square

El siguiente resultado es una nueva versión del Teorema 4.2 donde hemos eliminado la hipótesis (iii) y hemos reemplazado la hipótesis (iv) por una condición más fuerte sobre la función de regresión η .

Teorema 4.5. *Sea (\mathcal{F}, d) un espacio métrico. Sea $\{(\mathcal{X}_i, Y_i)\}_{i=1}^n$ una muestra aleatoria independiente e idénticamente distribuída como el par $(\mathcal{X}, Y) \in \mathcal{F} \times \mathbb{R}$ que satisface el modelo (4.1) y supongamos que $W_n(\mathcal{X}) = \{W_{ni}(\mathcal{X})\}_{i=1}^n$ es una sucesión de pesos de probabilidad que satisface las siguientes hipótesis:*

(i) *Para todo $a > 0$,*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) = 0;$$

(ii)

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\max_{1 \leq i \leq n} W_{ni}(\mathcal{X}) \right) = 0;$$

(iii') *la función η es uniformemente continua y acotada.*

Entonces, el estimador dado en (4.2) es consistente en media cuadrática. Es decir,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left((\hat{\eta}(\mathcal{X}) - \eta(\mathcal{X}))^2 \right) = 0.$$

Corolario 4.6. *Sea W_n una sucesión de pesos de probabilidad que satisface las hipótesis (i) y (ii) del Teorema 4.5 y sea η una función de regresión que verifica la hipótesis (iii') del mismo teorema. Si U_n es una sucesión de pesos normales tal que, para cada $n \geq 1$, $|U_n| \leq MW_n$ para alguna constante $M \geq 1$. Entonces, el estimador definido por*

$$\hat{\eta}(\mathcal{X}) = \sum_{i=1}^n Y_i U_{ni}(\mathcal{X})$$

es consistente en media cuadrática.

Aquí nuevamente $|U_n| \leq MW_n$ significa que para cada $n \geq 1$, $|U_{ni}| \leq MW_{ni}$ para todo $i = 1, \dots, n$.

Demostración. La demostración es idéntica a la del Teorema 4.2 con la diferencia que, en (4.5) en lugar de usar (iii) y (iv) debemos usar (iii'). Luego, por esta hipótesis tenemos que para todo $\epsilon > 0$ existe $a > 0$ tal que si $d(\mathcal{X}, \mathcal{X}_i) \leq a$,

$$(\eta(\mathcal{X}) - \eta(\mathcal{X}_i))^2 < \epsilon.$$

Para ese a , usando que η es acotada y haciendo el mismo razonamiento que en (4.6) para η en lugar de η^* , en (4.5) tenemos que, para todo $\epsilon > 0$ existe N_1 tal que si $n \geq N_1$,

$$I \leq 2C^2\epsilon + \epsilon.$$

El resto de la prueba es idéntica por lo que el teorema queda demostrado. \square

4.3. Consistencia del estimador de k -vecinos más cercanos (k -NN)

En esta sección, utilizaremos el Teorema 4.2 para probar la consistencia en media cuadrática del estimador de k -vecinos más cercanos para datos funcionales. Para modelos donde la covariable es una curva, la definición del k -vecino más cercano de \mathcal{X} (y por lo tanto el respectivo estimador), son una extensión directa de las definiciones dadas en la Sección 1.1.3. En efecto, para calcular el k -vecino más cercano de \mathcal{X} seguimos los siguientes pasos:

- calculamos las distancias de cada \mathcal{X}_i a \mathcal{X} ,

$$D_i = d(\mathcal{X}, \mathcal{X}_i), \quad i = 1, \dots, n;$$

- calculamos los estadísticos de orden para esas distancias,

$$D_{(1)} \leq D_{(2)} \leq \dots \leq D_{(k)} \leq \dots \leq D_{(n)};$$

- las observaciones correspondientes a esos estadísticos de orden son los **vecinos más cercanos** de \mathcal{X} ,

$$\mathcal{X}_{(1)}, \mathcal{X}_{(2)}, \dots, \mathcal{X}_{(k)}, \dots, \mathcal{X}_{(n)}.$$

Así por ejemplo, $\mathcal{X}_{(1)}$ es el 1-vecino más cercano de \mathcal{X} , $\mathcal{X}_{(2)}$ es el 2-vecino más cercano de \mathcal{X} y $\mathcal{X}_{(k)}$ es el k -vecino más cercano de \mathcal{X} . Dado que k dependerá del tamaño de la muestra lo llamaremos k_n aunque seguiremos utilizando la notación k -vecinos más cercanos (k -NN) en lugar de k_n -vecinos más cercanos (k_n -NN). En caso de haber **empates**, es decir, en caso de que $D_{(i)} = D_{(j)}$ para $i \neq j$, el orden de $\mathcal{X}_{(i)}$ y $\mathcal{X}_{(j)}$ se asigna al azar entre ellos, como en el caso finito dimensional. En lo que sigue, por simplicidad de notación supondremos que no hay empates con probabilidad 1.

Definimos el estimador de k -vecinos más cercanos como

$$\hat{\eta}_n(\mathcal{X}) \doteq \sum_{i=1}^n W_{ni}(\mathcal{X}) Y_i, \quad (4.7)$$

donde los pesos están definidos como $W_{ni} = \frac{1}{k_n}$ si \mathcal{X}_i está entre los k -vecinos más cercanos a \mathcal{X} entre $\mathcal{X}_1, \dots, \mathcal{X}_n$, y $W_{ni} = 0$ sino. Más precisamente,

$$W_{ni}(\mathcal{X}) = \begin{cases} \frac{1}{k_n} & \text{si } d(\mathcal{X}, \mathcal{X}_i) \leq d(\mathcal{X}_{(k_n)}, \mathcal{X}), \\ 0 & \text{en otro caso.} \end{cases} \quad (4.8)$$

Teorema 4.7. *Supongamos que (\mathcal{F}, d) es un espacio métrico separable y que la función de regresión η es uniformemente continua y acotada. Sea k_n una sucesión de números reales positivos que tiende a infinito tal que $k_n/n = o(1)$. Entonces, el estimador de k -vecinos más cercanos (4.7) con pesos dados por (4.8) satisface*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left((\eta(\mathcal{X}) - \hat{\eta}_n(\mathcal{X}))^2 \right) = 0.$$

Demostración. Probaremos que los pesos (4.8) satisfacen las hipótesis (i) y (ii) del Teorema 4.5. La hipótesis (ii) se verifica de manera inmediata ya que por hipótesis $k_n \rightarrow \infty$ y por lo tanto,

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\max_{1 \leq i \leq n} W_{ni}(\mathcal{X}) \right) = \lim_{n \rightarrow \infty} \mathbb{E} \left(\max_{1 \leq i \leq n} \frac{1}{k_n} \right) = \lim_{n \rightarrow \infty} \frac{1}{k_n} = 0.$$

Para probar la hipótesis (i) consideremos $a > 0$ arbitrario y escribimos

$$\begin{aligned} \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) &= \mathbb{E} \left(\sum_{i=1}^n \frac{1}{k_n} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq d(\mathcal{X}, \mathcal{X}_{(k_n)})\}} \right) \\ &\leq \mathbb{E} \left(\frac{1}{k_n} \sum_{i=1}^{k_n} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a\}} \right) \\ &= P(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a). \end{aligned} \quad (4.9)$$

El siguiente lema será demostrado en el Apéndice 4.5.

Lema 4.8. *Sea $\{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ una muestra aleatoria de \mathcal{X} . Supongamos que el espacio (\mathcal{F}, d) es separable y que \mathcal{X} es independiente de los datos. Entonces $\forall a > 0$, $P(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a) \rightarrow 0$ siempre que $k_n/n \rightarrow 0$.*

Luego, aplicando este lema en (4.9) tenemos que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{i=1}^n W_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) = 0,$$

por lo que la parte (i) queda demostrada. \square

4.4. Consistencia del estimador de núcleos

En el contexto infinito dimensional, este estimador se define de la misma manera que el estimador de Nadaraya-Watson finito-dimensional presentado en la Sección 1.2.2 con la única diferencia que la norma euclidea utilizada en \mathbb{R} o \mathbb{R}^d es reemplazada por la métrica d definida en \mathcal{F} . Más precisamente, definimos el estimador de núcleos para datos funcionales como

$$\widehat{\eta}_n(\mathcal{X}) \doteq \sum_{i=1}^n W_{ni}(\mathcal{X}) Y_i \quad (4.10)$$

donde los pesos son de la forma

$$W_{ni}(\mathcal{X}) = \begin{cases} \frac{K\left(\frac{d(\mathcal{X}, \mathcal{X}_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{d(\mathcal{X}, \mathcal{X}_i)}{h_n}\right)} & \text{si } \sum_{i=1}^n K\left(\frac{d(\mathcal{X}, \mathcal{X}_i)}{h_n}\right) \neq 0, \\ 0 & \text{si } \sum_{i=1}^n K\left(\frac{d(\mathcal{X}, \mathcal{X}_i)}{h_n}\right) = 0, \end{cases} \quad (4.11)$$

con h_n una sucesión que tiende a cero y K una función núcleo **regular** lo que significa que existen constantes $0 < c_1 < c_2 < \infty$ tales que $c_1 \mathbb{I}_{[0,1]}(u) \leq K(u) \leq c_2 \mathbb{I}_{[0,1]}(u)$.

En lo que sigue, probaremos la convergencia en media cuadrática del estimador de núcleos con ventana variable $h_n(x)$.

Teorema 4.9. *Supongamos que (\mathcal{F}, d) es un espacio métrico separable y que la función de regresión η es uniformemente continua y acotada. Para cada $x \in \text{sop}(\mu)$ fijo, supongamos que existe una sucesión de números reales positivos $h_n = h_n(x) \rightarrow 0$ tal que $\frac{nh_n(x)}{\log n} \rightarrow \infty$ y sea K un núcleo regular. Entonces, el estimador de núcleos (4.10) con pesos dados por (4.11) es consistente en media cuadrática. Es decir,*

$$\lim_{n \rightarrow \infty} \mathbb{E} \left((\eta(\mathcal{X}) - \widehat{\eta}_n(\mathcal{X}))^2 \right) = 0.$$

Observación 4.10. Si la medida μ es tal que la medida de la frontera de las bolas es cero, es decir, $\mu(\partial \mathcal{B}(x, d)) = 0$ para todo $x \in \text{sop}(\mu)$ y para todo $d > 0$, la sucesión $h_n = h_n(x)$ del Teorema 4.9 existe. Esto será demostrado en el Lema 4.12 del Apéndice 4.5.

Demostración. Sea $Z_n(\mathcal{X}, \mathcal{X}_1, \dots, \mathcal{X}_n) = Z_n(\mathcal{X}) \doteq \sum_{j=1}^n \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_j) \leq h_n\}}$. Consideremos los pesos U_{ni} definidos por

$$U_{ni}(\mathcal{X}, \mathcal{X}_1, \dots, \mathcal{X}_n) = U_{ni}(\mathcal{X}) \doteq \begin{cases} 0 & \text{si } Z_n(\mathcal{X}) = 0, \\ \frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} & \text{si } Z_n(\mathcal{X}) \neq 0. \end{cases}$$

Dado que K es regular existen constantes $0 < c_1 < c_2 < \infty$ tales que, para cada i ,

$$c_1 \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \leq K \left(\frac{d(\mathcal{X}, \mathcal{X}_i)}{h_n} \right) \leq c_2 \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}}.$$

De esta desigualdad se sigue que $Z_n(\mathcal{X}) \neq 0$ si y sólo si $\sum_{i=1}^n K \left(\frac{d(\mathcal{X}, \mathcal{X}_i)}{h_n} \right) \neq 0$ y en este caso

$$W_{ni}(\mathcal{X}) = \frac{K \left(\frac{d(\mathcal{X}, \mathcal{X}_j)}{h_n} \right)}{\sum_{j=1}^n K \left(\frac{d(\mathcal{X}, \mathcal{X}_j)}{h_n} \right)} \leq \frac{c_2}{c_1} \frac{\mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}}}{\sum_{j=1}^n \mathbb{I}_{\{d(\mathcal{X}_j, \mathcal{X}) \leq h_n\}}} = \frac{c_2}{c_1} \frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}}$$

de donde se sigue que

$$W_{ni} \leq \frac{c_2}{c_1} U_{ni}.$$

Por el Corolario 4.6 será suficiente probar que estos pesos satisfacen las hipótesis (i) y (ii) del Teorema 4.5. Comencemos verificando la hipótesis (ii), para ello sea

$$Z_n^*(\mathcal{X}, \mathcal{X}_1, \dots, \mathcal{X}_n) = Z_n^*(\mathcal{X}) \doteq \begin{cases} 0 & \text{si } Z_n(\mathcal{X}) = 0, \\ \frac{1}{Z_n(\mathcal{X})} & \text{si } Z_n(\mathcal{X}) \neq 0. \end{cases}$$

Dado que $\max_{1 \leq i \leq n} U_{ni}(\mathcal{X}) \leq Z_n^*(\mathcal{X})$ tenemos que

$$\mathbb{E} \left(\max_{1 \leq i \leq n} U_{ni}(\mathcal{X}) \right) \leq \mathbb{E} (Z_n^*(\mathcal{X})) = \mathbb{E} \left(\frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \right). \quad (4.12)$$

Luego será suficiente probar que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \right) = 0.$$

Dado que (\mathcal{F}, d) es separable, el Lema A.18 nos asegura que $P(\mathcal{X} \in \text{sop}(\mu)) = 1$ luego,

$$\begin{aligned} & \mathbb{E} \left(\frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \right) = \\ &= \mathbb{E} \left(\frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \mid \mathcal{X} \in \text{sop}(\mu) \right) P(\mathcal{X} \in \text{sop}(\mu)) \\ &+ \mathbb{E} \left(\frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \mid \mathcal{X} \notin \text{sop}(\mu) \right) P(\mathcal{X} \notin \text{sop}(\mu)) \quad (\text{Cor. A.21}) \\ &= \mathbb{E} \left(\frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \mid \mathcal{X} \in \text{sop}(\mu) \right) \quad (4.13) \\ &= \mathbb{E} \left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \right) \quad (\text{Obs. A.19}) \\ &= \mathbb{E}_{\mathcal{X}} \left(\mathbb{E}_{\mathcal{D}_n | \mathcal{X}} \left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \mid \mathcal{X} \in \text{sop}(\mu) \right) \right). \quad (\text{Prop. A.23}) \end{aligned}$$

Sea $x \in \text{sop}(\mu)$ fijo. Escribimos

$$\frac{1}{Z_n(x)} \mathbb{I}_{\{Z_n(x) \neq 0\}} = \frac{1/n p_{h_n}(x)}{Z_n(x)/n p_{h_n}(x)} \mathbb{I}_{\{Z_n(x) \neq 0\}}. \quad (4.14)$$

El siguiente lema que será demostrado en el Apéndice 4.5, nos asegura que el denominador de esta expresión tiende a uno.

Lema 4.11. *Para cada $x \in \text{sop}(\mu)$ fijo, si $Z_n(x) \neq 0$ entonces*

$$\frac{Z_n(x)}{np_{h_n}(x)} \xrightarrow{\text{a.co.}} 1 \quad \text{siempre que} \quad \frac{np_{h_n}(x)}{\log n} \rightarrow \infty.$$

Por el Lema 4.11, para todo $\epsilon > 0$ existe $C_1 > 0$ tal que si $\frac{np_{h_n}(x)}{\log n} > C_1$,

$$\left| \frac{Z_n(x)}{np_{h_n}(x)} - 1 \right| < \epsilon.$$

Para ese C_1 , existe $N_1 = N_1(x)$ tal que si $n \geq N_1$,

$$\frac{np_{h_n}(x)}{\log n} > C_1.$$

Entonces, en (4.14) tenemos que $\forall \epsilon > 0$ existe $N_1 = N_1(x)$ tal que si $n \geq N_1$,

$$\frac{1}{Z_n(x)} \mathbb{I}_{\{Z_n(x) \neq 0\}} \leq \frac{1}{C_1(1 - \epsilon) \log n}$$

para alguna constante positiva C_1 . Por lo tanto,

$$\lim_{n \rightarrow \infty} \frac{1}{Z_n(x)} \mathbb{I}_{\{Z_n(x) \neq 0\}} = 0.$$

Además,

$$\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \leq \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}}$$

con $\int \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} d\mu = \int_{\text{sop}(\mu)} 1 d\mu = 1$ por lo que podemos aplicar el Teorema de la convergencia dominada y concluir que

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{X}} \left(\mathbb{E}_{\mathcal{D}_n} \left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \middle| \mathcal{X} \in \text{sop}(\mu) \right) \right) = 0. \quad (4.15)$$

Luego, de (4.12), (4.13) y (4.15) tenemos que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\max_{1 \leq i \leq n} U_{ni}(\mathcal{X}) \right) = 0.$$

Para probar que los pesos $U_{ni} = \frac{1}{Z_n(\mathcal{X})} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}}$ satisfacen la hipótesis (i) consideremos $a > 0$ arbitrario y escribimos,

$$\begin{aligned}
& \mathbb{E} \left(\sum_{i=1}^n U_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) = \\
& = \mathbb{E} \left(\sum_{i=1}^n \frac{1}{Z_n} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) \tag{4.16} \\
& = \mathbb{E} \left(\frac{1}{Z_n} \sum_{i=1}^n \mathbb{I}_{\{a < d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \right) \\
& = \mathbb{E}_{\mathcal{X}} \left(\mathbb{E}_{\mathfrak{D}_n | \mathcal{X}} \left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n} \sum_{i=1}^n \mathbb{I}_{\{a < d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \mid \mathcal{X} \in \text{sop}(\mu) \right) \right). \quad (\text{Prop. A.23})
\end{aligned}$$

Sea $x \in \text{sop}(\mu)$ fijo. Dado que $h_n \rightarrow 0$, para todo $a > 0$ existe $N_2 = N_2(a, x)$ tal que si $n \geq N_2$ entonces $h_n < a$. Es decir, para todo $n \geq N_2$,

$$\sum_{i=1}^n \mathbb{I}_{\{a < d(\mathcal{X}_i, x) \leq h_n\}} = 0$$

luego,

$$\lim_{n \rightarrow \infty} \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n} \sum_{i=1}^n \mathbb{I}_{\{a < d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} = 0.$$

Además, dado que $Z_n(\mathcal{X}) = \sum_{i=1}^n \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}}$,

$$\frac{1}{Z_n} \sum_{i=1}^n \mathbb{I}_{\{a < d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \leq 1$$

de donde se sigue que,

$$\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n} \sum_{i=1}^n \mathbb{I}_{\{a < d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \leq \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}}$$

con $\int \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} d\mu = \int_{\text{sop}(\mu)} 1 d\mu = 1$ luego, del Teorema de la convergencia dominada tenemos que

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathcal{X}} \left(\mathbb{E}_{\mathfrak{D}_n | \mathcal{X}} \left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \frac{1}{Z_n} \sum_{i=1}^n \mathbb{I}_{\{a < d(\mathcal{X}, \mathcal{X}_i) \leq h_n\}} \mathbb{I}_{\{Z_n(\mathcal{X}) \neq 0\}} \mid \mathcal{X} \in \text{sop}(\mu) \right) \right) = 0.$$

Finalmente, con esta igualdad en (4.16) tenemos que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left(\sum_{i=1}^n U_{ni}(\mathcal{X}) \mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_i) > a\}} \right) = 0.$$

Por lo tanto, el estimador de núcleo (4.10) con pesos dados por (4.11) es consistente en media cuadrática siempre que la función η sea uniformemente continua y acotada. \square

4.5. Apéndice

Demostración del Lema 4.8: Fijemos $x \in \text{sop}(\mu)$. Probaremos que $d(x, \mathcal{X}_{(k_n)}(x)) \xrightarrow{a.s.} 0$. Por la proposición (A.25), esto es equivalente a probar que para todo $a > 0$,

$$P_{\mathfrak{D}_n} \left(d(x, \mathcal{X}_{(k_n)}(x)) < a, \forall n \geq N(x) \right) \xrightarrow{N(x) \rightarrow \infty} 1. \quad (4.17)$$

Sea $a > 0$, por definición de soporte tenemos que $p_a(x) \doteq \mu(\mathcal{B}(x, a)) > 0$. Observemos que $d(x, \mathcal{X}_{(k_n)}(x)) < a$ si y sólo si

$$\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} > \frac{k_n}{n}. \quad (4.18)$$

Luego, (4.17) es equivalente a probar

$$P_{\mathfrak{D}_n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} > \frac{k_n}{n}, \forall n \geq N(x) \right) \xrightarrow{N \rightarrow \infty} 1. \quad (4.19)$$

Observemos que el lado derecho de (4.18) tiende a 0 por hipótesis, por lo que existe N_1 tal que si $n \geq N_1$,

$$\frac{k_n}{n} < \frac{p_a(x)}{2}. \quad (4.20)$$

Por la Ley fuerte de los grandes números, el lado izquierdo de la ecuación (4.18) converge casi seguramente a

$$\mathbb{E}_{\mathfrak{D}_n} \left(\mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} \right) = P_{\mathfrak{D}_n} (\mathcal{X}_i \in \mathcal{B}(x, a)) = p_a(x) > 0$$

Es decir, para todo $\delta > 0$

$$P_{\mathfrak{D}_n} \left(\left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} - p_a(x) \right| < \delta, \forall n \geq N \right) \xrightarrow{N \rightarrow \infty} 1.$$

En particular, para $\delta = \frac{p_a(x)}{2}$ se tiene que

$$P_{\mathfrak{D}_n} \left(\left| \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} - p_a(x) \right| < \frac{p_a(x)}{2}, \forall n \geq N \right) \xrightarrow{N \rightarrow \infty} 1$$

de donde se sigue que

$$P_{\mathfrak{D}_n} \left(\frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} > \frac{p_a(x)}{2}, \forall n \geq N \right) \xrightarrow{N \rightarrow \infty} 1.$$

Por otro lado, de (4.20) tenemos que

$$\left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} > \frac{p_a(x)}{2}, \forall n \geq N \right\} \subset \left\{ \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{\mathcal{X}_i: \mathcal{X}_i \in \mathcal{B}(x, a)\}} > \frac{k_n}{n}, \forall n \geq N \right\}$$

de donde se sigue (4.19).

Consideremos ahora \mathcal{X} aleatorio. Dado que \mathcal{F} es separable, el Lema A.18 nos asegura que $P(\mathcal{X} \in \text{sop}(\mu)) = 1$. Luego,

$$\begin{aligned}
P(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a) &= \\
&= P\left(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a \mid \mathcal{X} \in \text{sop}(\mu)\right) P(\mathcal{X} \in \text{sop}(\mu)) \\
&\quad + P\left(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a \mid \mathcal{X} \notin \text{sop}(\mu)\right) P(\mathcal{X} \notin \text{sop}(\mu)) \text{Teo. de la prob. total} \\
&= P\left(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a \mid \mathcal{X} \in \text{sop}(\mu)\right) \\
&= P(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a, \mathcal{X} \in \text{sop}(\mu)) P(\mathcal{X} \in \text{sop}(\mu)) \quad (\text{Def. Prob. Cond.}) \\
&= \mathbb{E}\left(\mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a\}} \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}}\right) \\
&= \mathbb{E}_{\mathcal{X}}\left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} \mathbb{E}_{\mathfrak{D}_n | \mathcal{X}}\left(\mathbb{I}_{\{d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a\}} \mid \mathcal{X} \in \text{sop}(\mu)\right)\right) \quad (\text{Prop. A.23}) \\
&= \mathbb{E}_{\mathcal{X}}\left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} P_{\mathfrak{D}_n | \mathcal{X}}\left(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a \mid \mathcal{X} \in \text{sop}(\mu)\right)\right).
\end{aligned}$$

Por el caso no aleatorio resulta que para $\mathcal{X} \in \text{sop}(\mu)$, $P(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a) \rightarrow 0$ en probabilidad y además $\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} P_{\mathfrak{D}_n}\left(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a \mid \mathcal{X} \in \text{sop}(\mu)\right) \leq \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}}$ con $\int \mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} d\mu = \int_{\text{sop}(\mu)} 1 d\mu = \mu(\text{sop}(\mu)) = 1$ luego, podemos aplicar el Teorema de la convergencia dominada y concluir que

$$P(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a) = \mathbb{E}_{\mathcal{X}}\left(\mathbb{I}_{\{\mathcal{X} \in \text{sop}(\mu)\}} P_{\mathfrak{D}_n | \mathcal{X}}\left(d(\mathcal{X}, \mathcal{X}_{(k_n)}) > a \mid \mathcal{X} \in \text{sop}(\mu)\right)\right) \rightarrow 0.$$

□

Demostración del Lema 4.11: Sea $x \in \text{sop}(\mu)$ fijo tal que $Z_n(x) \neq 0$. Debemos ver que para todo $\epsilon > 0$,

$$\sum_{n=1}^{\infty} P_{\mathfrak{D}_n}\left(\left|\frac{Z_n(x)}{np_{h_n}(x)} - 1\right| > \epsilon\right) < \infty.$$

Para ello observemos que

$$P_{\mathfrak{D}_n}\left(\left|\frac{Z_n(x)}{np_{h_n}(x)} - 1\right| > \epsilon\right) = P_{\mathfrak{D}_n}\left(|Z_n(x) - np_{h_n}(x)| > \epsilon np_{h_n}(x)\right). \quad (4.21)$$

Dado que $Z_n(x) = \sum_{i=1}^n \mathbb{I}_{\{d(\mathcal{X}_i, x) \leq h_n\}}$ si definimos las variables aleatorias $Y_{ni}(x) \doteq \mathbb{I}_{\{d(\mathcal{X}_i, x) \leq h_n\}}$ las cuales son i.i.d. con esperanza dada por

$$\mathbb{E}_{\mathfrak{D}_n}(Y_{ni}(x)) = \mathbb{E}_{\mathfrak{D}_n}\left(\mathbb{I}_{\{d(\mathcal{X}_i, x) \leq h_n\}}\right) = \mu(\mathcal{B}(x, h_n)) = p_{h_n}(x)$$

tenemos que $Z_n(x) = \sum_{i=1}^n Y_{ni}(x)$ y además,

$$\mathbb{E}_{\mathfrak{D}_n}(Z_n(x)) = \mathbb{E}_{\mathfrak{D}_n}\left(\sum_{i=1}^n Y_{ni}(x)\right) = \sum_{i=1}^n \mathbb{E}_{\mathfrak{D}_n}(Y_{ni}(x)) = np_{h_n}(x).$$

Luego, en (4.21) tenemos que

$$\begin{aligned} P_{\mathfrak{D}_n} (|Z_n(x) - np_{h_n}(x)| > \epsilon np_{h_n}(x)) &= P_{\mathfrak{D}_n} \left(\left| \sum_{i=1}^n (Y_{ni}(x) - \mathbb{E}_{\mathfrak{D}_n}(Y_{ni}(x))) \right| > \epsilon p_{h_n}(x)n \right) \\ &\doteq P_{\mathfrak{D}_n} \left(\left| \sum_{i=1}^n \bar{Y}_{ni}(x) \right| > \epsilon p_{h_n}(x)n \right), \end{aligned} \quad (4.22)$$

con $\bar{Y}_{ni}(x) \doteq Y_{ni}(x) - \mathbb{E}_{\mathfrak{D}_n}(Y_{ni}(x))$. Con el fin de aplicar la desigualdad de Bernstein (A.32) observemos que $|\bar{Y}_{ni}(x)| \leq 2$. Además, por la definición de $Y_{ni}(x)$,

$$\mathbb{E}_{\mathfrak{D}_n}(Y_{ni}^2(x)) = \mathbb{E}_{\mathfrak{D}_n}(Y_{ni}(x)) = p_{h_n}(x)$$

entonces,

$$\text{var}_{\mathfrak{D}_n}(\bar{Y}_{ni}(x)) = \text{var}_{\mathfrak{D}_n}(Y_{ni}(x)) = p_{h_n}(x) - p_{h_n}^2(x) = p_{h_n}(x)(1 - p_{h_n}(x)).$$

Por lo tanto, aplicando el Teorema (A.32) en (4.22) con $\epsilon = \epsilon p_{h_n}(x)$, $\sigma^2 = p_{h_n}(x)(1 - p_{h_n}(x))$ y $M = 2$ resulta,

$$\begin{aligned} P_{\mathfrak{D}_n} \left(\left| \sum_{i=1}^n \bar{Y}_{ni}(x) \right| > \epsilon p_{h_n}(x)n \right) &\leq 2 \exp \left\{ -\frac{(\epsilon p_{h_n}(x))^2 n}{2p_{h_n}(x)(1 - p_{h_n}(x)) + 4\epsilon p_{h_n}(x)} \right\} \\ &= 2 \exp \left\{ -\frac{\epsilon^2 p_{h_n}(x)n}{2(1 - p_{h_n}(x)) + 4\epsilon} \right\} \\ &\leq 2 \exp \{-C_1 p_{h_n}(x)n\}. \end{aligned} \quad (4.23)$$

$(C_1 \doteq \frac{\epsilon^2}{2 + 4\epsilon})$

Por lo tanto, de (4.21), (4.22) y (4.23) tenemos que

$$\sum_{n=1}^{\infty} P_{\mathfrak{D}_n} \left(\left| \frac{Z_n(x)}{np_{h_n}(x)} - 1 \right| > \epsilon \right) \leq \sum_{n=1}^{\infty} \exp \{-C_1 p_{h_n}(x)n\} = \sum_{n=1}^{\infty} n^{-C_1 p_{h_n}(x)n/\log n}. \quad (4.24)$$

Dado que por hipótesis $\frac{p_{h_n}(x)n}{\log n} \rightarrow \infty$, existe $N_3 = N_3(x)$ tal que si $n \geq N_3$,

$$\frac{p_{h_n}(x)n}{\log n} \geq \frac{2}{C_1}.$$

Luego, en (4.24) tenemos que

$$\sum_{n=1}^{\infty} P_{\mathfrak{D}_n} \left(\left| \frac{Z_n(x)}{np_{h_n}(x)} - 1 \right| > \epsilon \right) \leq N_3 + \sum_{n=N_3}^{\infty} n^{-C_1 p_{h_n}(x)n/\log n} < \infty,$$

y por lo tanto, el lema queda demostrado. □

Lema 4.12. Si μ es tal que $\mu(\partial\mathcal{B}(x, d)) = 0$ para todo $x \in \text{sop}(\mu)$ y para todo $d > 0$ entonces, existe una sucesión de números reales positivos $h_n = h_n(x) \rightarrow 0$ tal que

$$\frac{np_{h_n}(x)}{\log n} \rightarrow \infty.$$

Demostración del Lema 4.12: Sean $x \in \text{sop}(\mu)$ fijo. Consideremos la función $M : (0, 1) \rightarrow (0, p_1(x))$ definida por

$$M(h) \doteq p_h(x) = \mu(\mathcal{B}(x, h)).$$

Probaremos que M es continua. En efecto, sea $\overline{\mathcal{B}}(x, h)$ la bola cerrada de centro x y radio h , dado que por hipótesis $\mu(\partial\mathcal{B}(x, h)) = 0$ entonces $\mu(\mathcal{B}(x, h)) = \mu(\overline{\mathcal{B}}(x, h))$. Sea $d > 0$, tal que $d \rightarrow 0^+$, luego por la continuidad de la medida tenemos que $\mu(\overline{\mathcal{B}}(x, h+d)) \searrow \mu(\overline{\mathcal{B}}(x, h))$ de donde se sigue que

$$\lim_{d \rightarrow 0^+} (M(h+d) - M(h)) = \lim_{d \rightarrow 0^+} (\mu(\overline{\mathcal{B}}(x, h+d)) - \mu(\overline{\mathcal{B}}(x, h))) = 0.$$

Análogamente, si $d < 0$ con $d \rightarrow 0^-$ entonces, $\mu(\mathcal{B}(x, h-d)) \nearrow \mu(\mathcal{B}(x, h))$ de donde se sigue que

$$\lim_{d \rightarrow 0^-} (M(h-d) - M(h)) = \lim_{d \rightarrow 0^-} (\mu(\mathcal{B}(x, h-d)) - \mu(\mathcal{B}(x, h))) = 0.$$

Para probar el lema debemos encontrar una sucesión $h_n = h_n(x) \rightarrow 0$ tal que

$$\frac{nM(h_n)}{\log n} \rightarrow \infty.$$

Por definición de M tenemos que si $h_1 < h_2$ entonces, $M(h_1) \leq M(h_2)$. Supongamos primero que M es estrictamente creciente y sea c_n una sucesión que tiende a infinito, tal que $\frac{c_n \log n}{n} \xrightarrow{n \rightarrow \infty} 0$. Definiendo $h_n \doteq M^{-1}\left(\frac{c_n \log n}{n}\right)$ tenemos que, $M(h_n) = \frac{c_n \log n}{n}$ lo que implica que $\frac{nM(h_n)}{\log n} = c_n \xrightarrow{n \rightarrow \infty} \infty$.

Si para algún n , $\frac{c_n \log n}{n} = M([a_n, b_n])$, con $a_n < b_n$, tomando como $h_n = a_n$ tenemos que $h_n = a_n = M^{-1}\left(\frac{c_n \log n}{n}\right)$ de donde se sigue nuevamente que $\frac{nM(h_n)}{\log n} = c_n \xrightarrow{n \rightarrow \infty} \infty$.

□

RESULTADOS Y HERRAMIENTAS ÚTILES

En este capítulo introduciremos conceptos, definiciones y resultados que serán utilizados a lo largo de toda la tesis. Estos resultados fueron extraídos principalmente de los siguientes libros [Taylor, 1973], [Wheeden and Zygmund, 1977], [Ferguson, 1996], [Folland, 1999], [Athreya and Lahiri, 2006] y [Ferraty and Vieu, 2006].

A.1. Espacios métricos

Definición A.1. Dado un conjunto \mathcal{F} , una **métrica** (o **distancia**) sobre \mathcal{F} es una función no negativa $d : \mathcal{F} \times \mathcal{F} \rightarrow \mathbb{R}_0^+$ tal que,

1. $d(x, y) = 0$ si y sólo si $x = y$;
2. $d(x, y) = d(y, x)$, para todo $x, y \in \mathcal{F}$;
3. $d(x, y) \leq d(x, z) + d(z, y)$, para todo $x, y, z \in \mathcal{F}$;

Diremos que el par (\mathcal{F}, d) es un **espacio métrico** si d es una métrica sobre el conjunto \mathcal{F} . Dado un espacio métrico (\mathcal{F}, d) , si $x \in \mathcal{F}$ y $r > 0$, entonces

$$\mathcal{B}(x, r) = \{y \in \mathcal{F} : d(x, y) < r\}, \tag{A.1}$$

denotará la **bola** (abierta) de centro x y radio r . Un subconjunto D de \mathcal{F} es **abierto** si para todo $x \in D$ existe $r > 0$ tal que $\mathcal{B}(x, r) \subseteq D$ y es **cerrado** si su complemento $D^c = \mathcal{F} - D$ es abierto. Un conjunto V es llamado **entorno** de un punto $x \in \mathcal{F}$ es cualquier subconjunto abierto de \mathcal{F} que contiene a x (en particular la bola abierta de centro x y radio r es un entorno de x). La unión de todos los subconjuntos abiertos de D es llamado el **interior** D . Denotaremos al interior de D por D° . Llamaremos **clausura** de D , y denotaremos \overline{D} , a la intersección de todos los conjuntos cerrados que contienen a D . La **frontera** ∂D de un conjunto D se define

como $\partial D = \overline{D} - D^\circ$. Diremos que D es **denso** en \mathcal{F} si $\overline{D} = \mathcal{F}$. Más precisamente, diremos que D es denso si para cada $x \in \mathcal{F}$ existe $y \in D$ tal que $d(x, y) \leq \epsilon$, para todo $\epsilon > 0$. Un conjunto $K \subset \mathcal{F}$ es **compacto** si todo cubrimiento por abiertos de K tiene un subcubrimiento numerable. Diremos que el espacio métrico (\mathcal{F}, d) es **separable** si contiene un conjunto denso numerable y que es **localmente compacto** si todo punto de \mathcal{F} tiene un entorno cuya clausura es compacta.

Si (\mathcal{F}_1, d_1) (\mathcal{F}_2, d_2) son dos espacios métricos, una función $f : \mathcal{F}_1 \rightarrow \mathcal{F}_2$ es **continua en un punto** $x \in \mathcal{F}_1$ si, para todo $\epsilon > 0$ existe $\delta = \delta(x, \epsilon) > 0$ tal que si $d_1(x, y) < \delta$ entonces, $d_2(f(x), f(y)) < \epsilon$. Diremos que f es **continua** si es continua en cada $x \in \mathcal{F}_1$ y que es **uniformemente continua** si δ no depende de x . Además, diremos que f es **Lipschitz continua** si existe una constante K , llamada **constante Lipschitz**, tal que $d_2(f(x), f(y)) \leq K d_1(x, y)$. Toda función Lipschitz es uniformemente continua.

Definición A.2. Sea $\mathcal{F} = \mathbb{R}^d$. Una **norma** en \mathbb{R}^d es una función $\|\cdot\| : \mathbb{R}^d \rightarrow \mathbb{R}_0^+$ tal que,

1. $\|x\| \geq 0$ y $\|x\| = 0$ si y sólo $x = 0$;
2. $\|\alpha x\| = |\alpha| \|x\|$ para todo $\alpha \in \mathbb{R}$;
3. $\|x + y\| \leq \|x\| + \|y\|$, para todo $x, y, z \in \mathcal{F}$;

Una norma en \mathbb{R}^d nos permite definir una distancia como $d(x, y) = \|x - y\|$.

A.2. Teoría de la medida

Definición A.3. Sea Ω un conjunto no vacío. Una **σ -álgebra** de conjuntos de Ω es una colección no vacía \mathcal{A} de subconjuntos de Ω tal que,

1. $\emptyset \in \mathcal{A}$;
2. si $A \in \mathcal{A}$, entonces $A^c \in \mathcal{A}$;
3. si $\{A_i\}_{i=1}^\infty \in \mathcal{A}$, entonces $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$.

Las condiciones 1. y 2. implican que $\Omega \in \mathcal{A}$, 2. y 3. implican que, si $\{A_i\}_{i=1}^\infty \in \mathcal{A}$, entonces $\bigcap_{i=1}^\infty A_i \in \mathcal{A}$. Además, es fácil ver que la intersección de toda σ -álgebra es también una σ -álgebra y que $\mathcal{P}(\Omega) = \{E : E \subseteq \Omega\}$ y $\{\emptyset, \Omega\}$ son σ -álgebras.

Si Ω es un conjunto y \mathcal{A} es una σ -álgebra de subconjuntos de Ω , entonces el par (Ω, \mathcal{A}) es llamado **espacio medible**. Dada una familia \mathcal{E} de subconjuntos de Ω , existe una mínima

σ -álgebra sobre Ω que contiene a \mathcal{E} , llamada **σ -álgebra generada por \mathcal{E}** . Existe al menos una σ -álgebra sobre Ω que contiene a \mathcal{E} el conjunto de partes de Ω , $\mathcal{P}(\Omega)$. La σ -álgebra generada por los subconjuntos abiertos de Ω es llamada **σ -álgebra de Borel** en Ω y es denotada por $\mathfrak{B}(\Omega)$. Todo $B \in \mathfrak{B}(\Omega)$ es llamado **Boreliano** o **conjunto de Borel**.

Sean Ω_1 y Ω_2 dos conjuntos no vacíos. Sea \mathcal{A}_i la σ -álgebra sobre Ω_i , $i = 1, 2$, entonces, la **σ -álgebra producto** sobre $\Omega_1 \times \Omega_2$ es la σ -álgebra generada por los **rectángulos** $\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}$ y se denota por $\mathcal{A}_1 \otimes \mathcal{A}_2$. Esta definición se extiende a una cantidad numerable de conjuntos y σ -álgebras.

Sea (Ω, \mathcal{A}) un espacio medible y (\mathcal{F}, d) un espacio métrico. Consideremos la función $f : \Omega \rightarrow \mathcal{F}$. Diremos que f es una **función medible** si, para todo $B \in \mathfrak{B}(\mathcal{F})$,

$$f^{-1}(B) = \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{A}.$$

Definición A.4. Sea (Ω, \mathcal{A}) un espacio medible. Una **medida** ν en \mathcal{A} es una función $\nu : \mathcal{A} \rightarrow [0, \infty)$ tal que,

1. $\nu(\emptyset) = 0$;
2. ν es σ -aditiva, esto es, si $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$ es una colección numerable de conjuntos tales que $A_i \cap A_j = \emptyset$, $i \neq j$, entonces

$$\nu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \nu(A_i).$$

De 2. se sigue que ν es monótona creciente, es decir, si $A_1 \subset A_2$ entonces, $\nu(A_1) \leq \nu(A_2)$. Dada una medida ν , diremos que es de **Borel** si su dominio es la σ -álgebra de Borel, es decir, si $\mathcal{A} = \mathfrak{B}(\Omega)$ y diremos que es **finita** si $\nu(\Omega) < \infty$. Además, si existe una colección de conjuntos $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$ con $\nu(A_i) < \infty$ para todo i , tal que $\Omega = \bigcup_{i=1}^{\infty} A_i$ entonces diremos que ν es **σ -finita**. Si (Ω, \mathcal{A}) es un espacio medible y ν es una medida en \mathcal{A} , la terna $(\Omega, \mathcal{A}, \nu)$ es llamada **espacio de medida** y un conjunto $A \in \mathcal{A}$ es llamado **conjunto ν -medible** o simplemente **conjunto medible**. Si ν es σ -finita, diremos que el espacio de medida es **σ -finito**.

El **soporte** de la medida ν se define como el conjunto $\{x \in \Omega : \nu(\mathcal{B}(x, \epsilon)) > 0, \forall \epsilon > 0\}$. Denotaremos a este conjunto por $\text{sop}(\nu)$.

Si ν_1 y ν_2 son medidas sobre (Ω, \mathcal{A}) , decimos que ν_2 es **absolutamente continua** con respecto a ν_1 , y escribimos $\nu_2 \ll \nu_1$, si $\nu_2(A) = 0$ para todo conjunto $A \in \mathcal{A}$ tal que $\nu_1(A) = 0$.

Sean $(\Omega_1, \mathcal{A}_1, \nu_1)$ y $(\Omega_2, \mathcal{A}_2, \nu_2)$ dos espacios de medida y sea $\mathcal{A}_1 \otimes \mathcal{A}_2$ la σ -álgebra producto sobre $\Omega_1 \times \Omega_2$. Definimos la **medida producto** $\nu_1 \times \nu_2$ como la única medida en $(\Omega_1 \times \Omega_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ tal que para todo $A_1 \in \mathcal{A}_1$ y $A_2 \in \mathcal{A}_2$,

$$\nu_1 \times \nu_2(A_1 \times A_2) = \nu_1(A_1)\nu_2(A_2).$$

A.3. Integración

La integral de una función f con respecto a una medida ν sobre el espacio de medida $(\Omega, \mathcal{A}, \nu)$ se define mediante los pasos usuales (ver [Wheeden and Zygmund, 1977]).

Una **función simple** $f : \Omega \rightarrow \mathbb{R}$ es una función de la forma

$$f(x) = \sum_{i=1}^k a_i \mathbb{I}_{A_i},$$

donde, a_1, \dots, a_k son números reales y A_1, \dots, A_k son conjuntos disjuntos ν -medibles. Definimos la **integral de una función simple** $f : \Omega \rightarrow \mathbb{R}$ con respecto a ν como

$$\int f d\nu = \int \left(\sum_{i=1}^k a_i \mathbb{I}_{A_i} \right) d\nu = \sum_{i=1}^k a_i \int \mathbb{I}_{A_i} d\nu = \sum_{i=1}^k a_i \nu(A_i).$$

La **integral de una función medible no negativa** se define como

$$\int_E f d\nu = \sup \left\{ \int_E g d\nu : g \text{ simple y } 0 \leq g \leq f \right\}.$$

Observemos que la integral de una función no negativa es siempre no negativa pero puede ser infinita. Finalmente, para calcular la integral de una función medible arbitraria f , debemos descomponerla en su **parte positiva** $f^+(x) = \max\{f(x), 0\}$ y en su **parte negativa** de $f^-(x) = \max\{-f(x), 0\}$. Luego, $f = f^+ - f^-$ y la **integral** de f es definida por

$$\int f(x) d\nu(x) = \int f d\nu = \int f^+ d\nu - \int f^- d\nu,$$

siempre que $\int f^+ d\nu$ y $\int f^- d\nu$ no sean simultáneamente infinitas. Diremos que una función f es **ν -integrable** (o simplemente integrable) si $\int |f| d\nu = \int f^+ d\nu + \int f^- d\nu < \infty$. Denotaremos al espacio de todas las funciones integrables como $L^1(\Omega, \mathcal{A}, \nu)$ o simplemente $L^1(\nu)$. Más

precisamente,

$$L^1(\nu) = \left\{ f : \Omega \rightarrow \mathbb{R} : f \text{ es medible y } \int |f| d\nu < \infty \right\}. \quad (\text{A.2})$$

Valen los siguientes teoremas de convergencia y diferenciación.

Teorema A.5. (Teorema de la convergencia dominada de Lebesgue) Sean $(\Omega, \mathcal{A}, \nu)$ un espacio de medida y $\{f_n\}$ una sucesión de funciones en $L^1(\nu)$ tal que,

1. $f_n \rightarrow f$ en c.t.p.
2. existe una función $g \in L^1(\nu)$ tal que $|f_n| \leq g$, en c.t.p. para todo n .

Entonces, $f \in L^1(\nu)$ y

$$\lim_{n \rightarrow \infty} \int f_n d\nu = \int f d\nu.$$

Teorema A.6. (Tonelli) Sean $(\Omega_1, \mathcal{A}_1, \nu_1)$ y $(\Omega_2, \mathcal{A}_2, \nu_2)$ dos espacios de medida σ -finitos y f una función medible no negativa sobre $\Omega_1 \times \Omega_2$. Entonces, las funciones $g(x) = \int f(x, y) d\nu_2(y)$ y $h(y) = \int f(x, y) d\nu_1(x)$ son funciones medibles en Ω_1 y Ω_2 respectivamente, y

$$\int f d(\nu_1 \times \nu_2) = \int_{\Omega_1} \left(\int_{\Omega_2} f(x, y) d\nu_2(y) \right) d\nu_1(x) = \int_{\Omega_2} \left(\int_{\Omega_1} f(x, y) d\nu_1(x) \right) d\nu_2(y).$$

Teorema A.7. (Desigualdad de Jensen) Sea $(\Omega, \mathcal{A}, \nu)$ un espacio de medida con $\nu(\Omega) = 1$. Sean f una función integrable y $\phi : \mathbb{R} \rightarrow \mathbb{R}$ una función convexa. Entonces,

$$\phi \left(\int f d\nu \right) \leq \int \phi(f) d\nu.$$

Teorema A.8. (Teorema de Radon-Nikodym) Sean ν_1 y ν_2 dos medidas σ -finitas sobre el espacio medible (Ω, \mathcal{A}) tal que $\nu_1 \ll \nu_2$. Entonces, existe una única función $f \in L^1(\nu_2)$ tal que

$$\nu_1(E) = \int_E f d\nu_2, \quad \forall E \in \mathcal{A}. \quad (\text{A.3})$$

Por otro lado, si ν_2 es una medida y $f \in L^1(\nu_2)$ entonces la medida ν_1 definida por (A.3) es absolutamente continua con respecto a ν_2 y es finita si y sólo si $f \in L^1(\nu_2)$. Cuando la medida ν_1 está definida de esta forma, llamamos a f **densidad** de ν_1 con respecto a ν_2 y escribimos $d\nu_1 = f d\nu_2$.

En lo que sigue consideremos $(\Omega, \mathcal{A}, \nu) = (\mathbb{R}, \mathfrak{B}(\mathbb{R}), \lambda)$ con λ la medida de Lebesgue en \mathbb{R} .

Teorema A.9. (Teorema de Diferenciación de Lebesgue) Sea $f \in L^1(\mathbb{R})$. Entonces, para casi todo $x \in \mathbb{R}$,

$$\lim_{\lambda(\mathcal{B}(x,r)) \searrow x} \frac{1}{\lambda(\mathcal{B}(x,r))} \int_{\mathcal{B}(x,r)} f(y) dy = f(x). \quad (\text{A.4})$$

Un punto x en el cual (A.4) se verifica se denomina **punto de Lebesgue** de f . Si f es Lipschitz entonces todo $x \in \mathbb{R}$ es punto de Lebesgue.

Teorema A.10. (Teorema del valor medio integral) Sea $[a, b] \subset \mathbb{R}$. Si f es continua en $[a, b]$ y diferenciable en (a, b) , existe $c \in [a, b]$ tal que

$$\int_a^b f(x) dx = f(c)(b - a).$$

Teorema A.11. (Teorema de Taylor) Sea $n \in \mathbb{Z}_0^+$ y f una función n veces continuamente diferenciable en el intervalo cerrado $[a, x]$ y $n + 1$ veces diferenciable en el intervalo abierto (a, x) . Entonces

$$f(x) = f(a) + \frac{f'(a)}{1!}(x - a) + \frac{f''(a)}{2!}(x - a)^2 + \dots + \frac{f^{(n)}(a)}{n!}(x - a)^n + R_n(x),$$

donde el resto $R_n(x)$ es la diferencia entre el polinomio de Taylor de grado n y la función original.

Teorema A.12. (Teorema del Binomio) Para todo $n \geq 0$,

$$(x + y)^n = \sum_{i=0}^n \binom{n}{i} x^i y^{n-i},$$

donde $\binom{n}{i} = \frac{n!}{i!(n-i)!}$ es el coeficiente binomial.

Corolario A.13. Si $x = y = 1$, dado que $\binom{n}{0} = \frac{n!}{0!(n-0)!} = 1$ entonces $\sum_{i=0}^n \binom{n}{i} = 2^n - 1$.

A.4. Los espacios L^p

Sea $(\Omega, \mathcal{A}, \nu)$ un espacio de medida. Si f es una función medible en Ω y $1 \leq p < \infty$ definimos al espacio L^p como

$$L^p(\Omega, \mathcal{A}, \nu) = L^p(\nu) = \left\{ f : \Omega \rightarrow \mathbb{R} : f \text{ es medible y } \int |f|^p d\nu < \infty \right\}.$$

Observemos que si $p = 1$ recuperamos el espacio L^1 definido en (A.2). Para $p = \infty$ definimos,

$$L^\infty(\Omega, \mathcal{A}, \nu) = L^\infty(\nu) = \{ f : \Omega \rightarrow \mathbb{R} : f \text{ es medible y } \inf \{ a \geq 0 : \nu(\{x : |f(x)| > a\}) = 0 \} < \infty \}.$$

Valen las siguientes desigualdades.

Teorema A.14. (Desigualdad de Hölder) Sea $(\Omega, \mathcal{A}, \nu)$ un espacio de medida. Sean $1 \leq p < \infty$, $f \in L^p(\nu)$ y $g \in L^q(\nu)$ donde $q = \frac{p}{p-1}$. Entonces,

$$\int |fg| d\nu \leq \left(\int |f|^p d\nu \right)^{1/p} \left(\int |g|^q d\nu \right)^{1/q}.$$

Corolario A.15. (*Desigualdad de Cauchy-Schwartz*) Si $f, g \in L^2(\nu)$ entonces,

$$\int |fg| d\nu \leq \left(\int |f|^2 d\nu \right)^{1/2} \left(\int |g|^2 d\nu \right)^{1/2}.$$

A.5. Probabilidad

Un espacio de medida (Ω, \mathcal{A}, P) es llamado **espacio de probabilidad** si $P(\Omega) = 1$. En este caso, la medida $P : \mathcal{A} \rightarrow [0, 1]$ se denomina **probabilidad**, el conjunto Ω **espacio muestral** y los elementos de \mathcal{A} **eventos**.

Definición A.16. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y (\mathcal{F}, d) un espacio métrico. Una función $X : \Omega \rightarrow \mathcal{F}$ es un **elemento aleatorio** en \mathcal{F} si es una función medible. Más precisamente, si para todo $B \in \mathfrak{B}(\mathcal{F})$, $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$ (o simplemente $X^{-1}(\mathfrak{B}) \subset \mathcal{A}$).

Notemos que si X es un elemento aleatorio, para cualquier colección $B_\alpha \in \mathfrak{B}$ se tiene que,

$$(X^{-1}(B_\alpha))^c = X^{-1}(B_\alpha^c), \quad \bigcap_{\alpha} X^{-1}(B_\alpha) = X^{-1}\left(\bigcap_{\alpha} B_\alpha\right), \quad \bigcup_{\alpha} X^{-1}(B_\alpha) = X^{-1}\left(\bigcup_{\alpha} B_\alpha\right),$$

luego, $X^{-1}(\mathfrak{B})$ es una σ -álgebra llamada la **σ -álgebra generada por el elemento aleatorio X** y es la menor σ -álgebra \mathcal{G} tal que X es medible.

Una **variable aleatoria** es un elemento aleatorio en $\mathcal{F} = \mathbb{R}$ y un **vector aleatorio** es un elemento aleatorio para en $\mathcal{F} = \mathbb{R}^d$.

Asociado a un elemento aleatorio podemos definir un nuevo espacio de probabilidad $(\mathcal{F}, \mathfrak{B}(\mathcal{F}), \mu)$ con μ definida como

$$\mu(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega : X(\omega) \in B\}) \doteq P(X \in B), \quad \forall B \in \mathfrak{B}(\mathcal{F}).$$

Es decir, el elemento aleatorio X induce la medida μ en la σ -álgebra de Borel de \mathcal{F} . La **medida de probabilidad** μ es llamada **distribución** del elemento aleatorio X . Dada una muestra X_1, X_2, \dots, X_n de X , definimos la **medida de probabilidad empírica** μ_n en $B \in \mathfrak{B}(\mathcal{F})$ como

$$\mu_n(B) \doteq \frac{1}{n} \sum_{i=1}^n \mathbb{I}_B(X_i).$$

Teorema A.17. (Continuidad de la medida de probabilidad)

1. Sea $\{A_i\}_{i \geq 1}$ una sucesión de conjuntos abiertos tales que $A_1 \subseteq A_2 \subseteq \dots$ entonces,

$$\lim_{n \rightarrow \infty} \mu(A_i) = \mu\left(\bigcup_{i=1}^{\infty} A_i\right).$$

2. Sea $\{B_i\}_{i \geq 1}$ una sucesión de conjuntos cerrados tales que $B_1 \supseteq B_2 \supseteq \dots$ entonces,

$$\lim_{n \rightarrow \infty} \mu(A_i) = \mu\left(\bigcap_{i=1}^{\infty} A_i\right).$$

Lema A.18. Sea (Ω, \mathcal{A}, P) un espacio de probabilidad y (\mathcal{F}, d) un espacio métrico separable. Entonces,

$$\mu(\text{sop}(\mu)) = P(X \in \text{sop}(\mu)) = 1.$$

Demostración. Sea $A = \text{sop}(\mu)$. Por definición de soporte resulta que,

$$A^c = \{X \in \mathcal{F} : \mu(\mathcal{B}(X, r_X)) = 0, \text{ para algún } r_X > 0\}.$$

Dado que \mathcal{F} es separable, existe un conjunto denso y numerable \mathcal{D} para el cual se tiene que, para todo $X \in A^c$, existe $Y \in \mathcal{D}$ tal que $d(X, Y) \leq r_X/3$. Esto implica que, $X \in \mathcal{B}(Y, r_X/2) \subset \mathcal{B}(X, r_X)$ pero como $X \in A^c$, $\mu(\mathcal{B}(X, r_X)) = 0$ entonces, $\mu(\mathcal{B}(Y, r_X/2)) = 0$. Luego,

$$A^c \subset \bigcup_{X \in A^c} \mathcal{B}(Y, r_X/2).$$

El lado derecho de esta inclusión es una unión numerable de conjuntos de medida nula, por lo tanto,

$$\mu(A^c) \leq \mu\left(\bigcup_{X \in A^c} \mathcal{B}(Y, r_X/2)\right) = \sum_{X \in A^c} \mu(\mathcal{B}(Y, r_X/2)) = 0,$$

de donde se sigue que $\mu(A) = 1$. □

Si X es un elemento aleatorio definido en un espacio de probabilidad (Ω, \mathcal{A}, P) con valores en un espacio métrico (\mathcal{F}, d) y $g : \mathcal{F} \rightarrow \mathbb{R}$ es una función medible entonces $Y = g(X)$ también es un elemento aleatorio definido en el espacio de probabilidad (Ω, \mathcal{A}, P) con valores en \mathbb{R} . En efecto, dado que g es medible, para cada $B \in \mathfrak{B}(\mathcal{F})$, $g(B) = \tilde{B} \in \mathfrak{B}(\mathbb{R})$ luego, dado que X es un elemento aleatorio, para todo $B \in \mathfrak{B}(\mathcal{F})$, $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{A}$ luego, para todo $\tilde{B} \in \mathfrak{B}(\mathbb{R})$,

$$Y^{-1}(\tilde{B}) = X^{-1}(g^{-1}(\tilde{B})) = X^{-1}(B) \in \mathcal{A}.$$

Sean (Ω, \mathcal{A}, P) un espacio de probabilidad y μ la medida de probabilidad de X . Definimos su **esperanza** como

$$\mathbb{E}(X) = \int_{\Omega} X dP = \int_{\mathcal{F}} x d\mu(x).$$

Cuando $\mathbb{E}(X) < \infty$ diremos que X es **integrable**. Para cualquier entero positivo p , definimos el **p -ésimo momento** del elemento aleatorio X como $\mathbb{E}(|X|^p)$. Luego, la esperanza de X es su primer momento. Si X tiene segundo momento finito, esto es, si $\mathbb{E}(|X|^2) < \infty$, definimos la **varianza** de X como $\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}^2(X)$. Dado que $\mathbb{E}(|X|^p) = \int |X|^p dP = \int |x|^p d\mu(x)$, tenemos que $L^p(\mu) = \{X : \Omega \rightarrow \mathcal{F} : \mathbb{E}(|X|^p) < \infty\}$ por lo que podemos utilizar las desigualdades dadas en la Sección A.4.

En lo que sigue $\mathcal{F} = \mathbb{R}$ y $d(x, y) = |x - y|$ será la métrica usual en \mathbb{R} por lo que X será una variable aleatoria. En este caso, la función de distribución de X , representada por F_X (o simplemente F) y la función de distribución empírica de X , representada por F_n , se definen respectivamente por

$$F(x) = P(X \leq x) \quad \text{y} \quad F_n(x) = \frac{\#\{X_i : X_i \leq x\}}{n}.$$

Diremos que una variable aleatoria es **discreta** si toma un número finito de valores y diremos que es **(absolutamente) continua** si existe una función $f_X \geq 0$, llamada **función de densidad** tal que para todo $x \in \mathbb{R}$,

$$F(x) = \int_{-\infty}^x f_X(u) du.$$

En este caso, la esperanza de X está dada por $\mathbb{E}(X) = \int x f_X(x) dx$. La **función de distribución conjunta** $F_{X,Y}$ del par (X, Y) donde X e Y son variables aleatorias, se define por $F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$ para todo $(x, y) \in \mathbb{R} \times \mathbb{R}$. La **función de densidad conjunta** de (X, Y) es la función $f_{X,Y} \geq 0$ tal que

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) du dv.$$

Diremos que las variables aleatorias son **independientes** si

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i) \quad \text{y} \quad \lim_{n \rightarrow \infty} F_{X_i}(x) = 1.$$

La **distribución marginal** y la **densidad marginal** de X se definen respectivamente por $F_X(x) = \lim_{y \rightarrow \infty} F_{X,Y}(x, y)$ y $f_X(x) = \int f_{X,Y}(x, y) dy$. La **densidad condicional** de Y dado

X representada por $f_{Y|X}(y|x)$, se define por

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)},$$

la **esperanza condicional** por

$$\mathbb{E}_Y(Y|X = x) = \int y f_{Y|X}(y|x) dy$$

y la **esperanza condicional** de X dado el evento B por

$$\mathbb{E}(X|B) = \frac{\int_B X dP}{P(B)}.$$

Observación A.19. Dado que $\int_B X dP = \int_{\Omega} \mathbb{I}_B X dP = \mathbb{E}(\mathbb{I}_B X)$ tenemos que

$$\mathbb{E}(X|B) = \frac{\mathbb{E}(\mathbb{I}_B X)}{P(B)}.$$

Teorema A.20. (Teorema de la probabilidad total) Sea X una variable aleatoria. Si A_1, \dots, A_n es una partición del espacio muestral Ω tal que $P(X \in A_i) > 0$ para todo $i = 1, \dots, n$. Entonces,

$$P(X) = \sum_{i=1}^n P(X|X \in A_i) P(X \in A_i).$$

Corolario A.21. Sea X una variable aleatoria integrable. Si A_1, \dots, A_n es una partición del espacio muestral Ω tal que $P(X \in A_i) > 0$ para todo $i = 1, \dots, n$. Entonces,

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X|X \in A_i) P(X \in A_i).$$

Proposición A.22. (Ley de la esperanza iterada) Si X es una variable aleatoria integrable e Y es otra variable aleatoria definida en el mismo espacio de probabilidad no necesariamente integrable entonces,

$$\mathbb{E}_X(X) = \mathbb{E}_Y(\mathbb{E}_{X|Y}(X|Y)).$$

Proposición A.23. Sean X, Y variables aleatorias definidos en (Ω, \mathcal{A}, P) con función de densidad conjunta $f_{X,Y}$ y sean $\psi : \mathbb{R} \rightarrow \mathbb{R}$ y $\phi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ funciones medibles. Entonces,

$$\mathbb{E}_{X,Y}(\psi(X)\phi(X,Y)) = \mathbb{E}(\psi(X)\mathbb{E}_{Y|X}(\phi(X,Y)|X)).$$

Demostración.

$$\begin{aligned}
\mathbb{E}_{X,Y}(\psi(X)\phi(X,Y)) &= \int \int \psi(x)\phi(x,y)f_{X,Y}(x,y)dx dy \\
&= \int \psi(x) \left(\int \phi(x,y)f_{X,Y}(x,y)dy \right) dx \\
&= \int f_X(x)\psi(x) \left(\int \phi(x,y)f_{Y|X}(y|x)dy \right) dx \\
&= \int f_X(x)\psi(x)\mathbb{E}_{Y|X}(\phi(x,Y)|X) dx \\
&= \mathbb{E}(\psi(X)\mathbb{E}_{Y|X}(\phi(X,Y)|X)).
\end{aligned}$$

□

A.6. Convergencias estocásticas

Dadas las variables aleatorias $\{X_n\}_{n \in \mathbb{N}}$ y X en (Ω, \mathcal{A}, P) , diremos que:

- $\{X_n\}_{n \in \mathbb{N}}$ **converge casi completamente** a X ($X_n \xrightarrow{a.co.} X$) si y solo si

$$\sum_{n \in \mathbb{N}} P(|X_n - X| > \epsilon) < \infty, \quad \forall \epsilon > 0.$$

- $\{X_n\}_{n \in \mathbb{N}}$ **converge casi seguramente** a X ($X_n \xrightarrow{a.s.} X$) si y solo si

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1.$$

- $\{X_n\}_{n \in \mathbb{N}}$ **converge en r -media** a X ($X_n \xrightarrow{r} X$) si y solo si

$$\lim_{n \rightarrow \infty} \mathbb{E}(|X_n - X|^r) = 0.$$

- $\{X_n\}_{n \in \mathbb{N}}$ **converge en probabilidad** a X , ($X_n \xrightarrow{p} X$) si y solo si

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0, \quad \forall \epsilon > 0.$$

- $\{X_n\}_{n \in \mathbb{N}}$ **converge en distribución** a X ($X_n \xrightarrow{\mathcal{D}} X$) si y solo si

$$\lim_{n \rightarrow \infty} F_n(x) = F(x),$$

para todo x punto de continuidad de F . En este caso decimos que F_n **converge débilmente** a F .

Proposición A.24. *Valen las siguientes implicaciones:*

- (i) $X_n \xrightarrow{a.co.} X \Rightarrow X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{\mathcal{D}} X$.
- (v) Si $X_n \xrightarrow{a.s.} X$ y $|X_n|^r \leq Z$, para alguna Z con $\mathbb{E}(Z) < \infty$ entonces, $X_n \xrightarrow{r} X$.

(vi) Si $X_n \xrightarrow{a.s.} X$, $X_n \geq 0$ y $\mathbb{E}(X_n) \rightarrow \mathbb{E}(X) < \infty$ entonces, $X_n \xrightarrow{r} X$ con $r = 1$.

Para la demostración de esta proposición ver [Ferraty and Vieu, 2006] pág 229 y [Ferguson, 1996] pág 6 y 9.

Proposición A.25. $X_n \xrightarrow{a.s.} X$ si y sólo si $\forall \epsilon > 0$,

$$P(|X_k - X| < \epsilon, \forall k \geq n) \xrightarrow{n \rightarrow \infty} 1.$$

Para la demostración de esta proposición ver [Ferguson, 1996] pág 5.

Diremos que la sucesión $X_n \xrightarrow{a.co.} X$ con **orden** u_n si y solo si existe $\epsilon_0 > 0$ tal que $\sum_n P(|X_n - X| > \epsilon_0 u_n) < \infty$.

A.7. Teoremas límites

Teorema A.26. (Teorema central del límite de Lindeberg) Sean X_1, X_2, \dots variables aleatorias independientes tales que $\mathbb{E}(X_i) = \mu_i$ y $\text{var}(X_i) = \sigma_i^2 < \infty$, donde σ_i^2 y por lo menos una $\sigma_i^2 > 0$. Sea $S_n = \sum_{i=1}^n X_i$ y $s_n = \sqrt{\text{var}(S_n)} = \sqrt{\sum_{i=1}^n \sigma_i^2}$. Entonces, para que

$$\frac{S_n - \mathbb{E}(S_n)}{s_n} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$$

es suficiente que se satisfaga la **condición de Lindeberg**

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \int_{\{|x - \mu_k| > \epsilon s_n\}} (x - \mu_k)^2 dF_{X_k}(x) = 0. \quad (\text{A.5})$$

Observación A.27. Si $s_n \rightarrow \infty$ entonces la condición de Lindeberg (A.5) se satisface y por lo tanto vale el Teorema A.26.

Teorema A.28. (Teorema central del límite para variables aleatorias i.i.d.) Sean X_1, X_2, \dots variables aleatorias independientes e idénticamente distribuidas, con media común μ y varianza común $\sigma^2 < \infty$. Sea $S_n = \sum_{i=1}^n X_i$ y $s_n = \sqrt{\text{var}(S_n)} = \sqrt{\sum_{i=1}^n \sigma^2} = \sqrt{n}\sigma$. Entonces,

$$\frac{S_n - \mathbb{E}(S_n)}{s_n} = \frac{S_n - n\mu}{\sqrt{n}\sigma} \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1).$$

Teorema A.29. (Ley de los grandes números) Sean X_1, X_2, \dots variables aleatorias integrables en el espacio de probabilidad (Ω, \mathcal{A}, P) y sea $S_n = \sum_{i=1}^n X_i$. Diremos que

- X_1, X_2, \dots satisfacen la **ley débil de los grandes números** si

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{p} 0.$$

- X_1, X_2, \dots satisfacen la ley fuerte de los grandes números si

$$\frac{S_n - \mathbb{E}(S_n)}{n} \xrightarrow{\text{a.s.}} 0.$$

A.8. Desigualdades exponenciales

A.8.1. De Bernstein

Teorema A.30. [Ferraty and Vieu, 2006] Sean X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas tales que $\mathbb{E}(X_i) = 0$ para todo i . Supongamos que $\forall m \geq 2$, $|\mathbb{E}(X_1^m)| \leq \frac{m!}{2}(a_i)^2 b^{m-2}$, donde $b \in \mathbb{R}$ y $A_n^2 = \sum_{i=1}^n (a_i)^2$. Entonces,

$$\forall \epsilon > 0, P\left(\left|\sum_{i=1}^n X_i\right| > \epsilon A_n\right) \leq 2 \exp\left\{-\frac{\epsilon^2}{2\left(1 + \frac{\epsilon b}{A_n}\right)}\right\}.$$

Corolario A.31. Sean X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas tales que $\mathbb{E}(X_i) = 0$ para todo i . Si $\forall m \geq 2$, $\mathbb{E}(|X_1^m|) \leq \frac{m!}{2} a^2 (m-1)$, entonces,

$$\forall \epsilon > 0, P\left(\left|\sum_{i=1}^n X_i\right| > \epsilon n\right) \leq 2 \exp\left\{-\frac{\epsilon^2 n}{2a^2(1 + \epsilon)}\right\}.$$

Corolario A.32. Sean X_1, X_2, \dots, X_n variables aleatorias independientes e idénticamente distribuidas tales que $\mathbb{E}(X_i) = 0$ para todo i . Supongamos que existe $M < \infty$ tal que, $|X_1| \leq M$, y denotemos con $\sigma^2 = \mathbb{E}(X_1^2)$, entonces,

$$\forall \epsilon > 0, P\left(\left|\sum_{i=1}^n X_i\right| > \epsilon n\right) \leq 2 \exp\left\{-\frac{\epsilon^2 n}{2\sigma^2\left(1 + \epsilon \frac{M}{\sigma^2}\right)}\right\} = 2 \exp\left\{-\frac{\epsilon^2 n}{2\sigma^2 + 2\epsilon M}\right\}.$$

A.8.2. Para variables aleatorias débilmente dependientes

Teorema A.33. [Doukhan and Neumann, 2006] Sean X_1, X_2, \dots, X_n variables aleatorias definidas en un espacio de probabilidad (Ω, \mathcal{A}, P) con $\mathbb{E}(X_i) = 0$ y $P(|X_i| \leq M) = 1$, para todo $i = 1, \dots, n$ y alguna constante $M < \infty$. Sean $S_n = \sum_{i=1}^n X_i$ y $\phi : \mathbb{N}^2 \rightarrow \mathbb{R}^+$ alguna de las siguientes funciones:

- $\phi(u, v) = 2v$;
- $\phi(u, v) = u + v$;
- $\phi(u, v) = uv$;
- $\phi(u, v) = \rho(u + v) + (1 - \rho)uv$, para algún $\rho \in (0, 1)$.

Supongamos que existen constantes $K, L_1, L_2 < \infty, \mu \geq 0$, y una sucesión no creciente de coeficientes reales $\{\alpha(n)\}_{n \geq 1}$ tales que, para toda u -upla (i_1, \dots, i_u) y para toda v -upla (j_1, \dots, j_v) con $1 \leq \dots i_u < j_1 + r = j_1 \leq \dots \leq j_v \leq n$ se cumple la siguiente desigualdad

$$|\text{cov}(X_{i_1} \dots X_{i_u}, X_{j_1} \dots X_{j_v})| \leq K^2 M^{u+v-2} \phi(u, v) \alpha(r),$$

donde

$$\sum_{j=0}^{\infty} (j+1)^k \alpha(j) \leq L_1 L_2^k (k!)^\mu, \quad \forall k \geq 0.$$

Entonces,

$$P(S_n \geq t) \leq \exp\left(-\frac{t^2/2}{\Sigma_n + X_n^{1/(\mu+2)} t^{(2\mu+3)/(\mu+2)}}\right),$$

donde Σ_n puede ser elegido mayor o igual a $\sigma_n^2 = \text{var}(S_n)$ y

$$X_n = 2(K \vee M) L_2 \left(\left(\frac{2^{4+\mu} n K^2 L_1}{\Sigma_n} \right) \vee 1 \right)$$

A.8.3. Para variables aleatorias α -mixing geométricas

Teorema A.34 . [Doukhan et al., 1984] Sea $\{X_1, \dots, X_n\}$ una sucesión de variables aleatorias geoméricamente α -mixing tales que $\mathbb{E}(X_i) = 0$, $|X_i| \leq 1$. Sea $\gamma = \frac{2}{1-\theta}$ y $\sigma = \sup_{i=1, \dots, n} \left\{ \mathbb{E}(|X_i|^\gamma)^{1/\gamma} \right\}$. Entonces, existen constantes C_1 y C_2 las cuales dependen sólo de los coeficientes “mixing”, tales que, para $0 < \theta < 1$,

$$\forall \epsilon > 0, \quad P\left(\sum_{i=1}^n X_i > \epsilon\right) \leq 2C_1 \theta^{-1} \exp\left\{-\frac{C_{2,n} \epsilon^{1/2}}{n^{1/4} \sigma^{1/2}}\right\},$$

donde $C_{2,n} = C_2$ si $n^{1/2} \sigma \leq 1$ y $C_{2,n} = C_2 n^{1/4} \sigma^{1/2}$ si $n^{1/2} \sigma > 1$.

A.9. Procesos estocásticos y campos aleatorios

Dado un espacio de probabilidad (Ω, \mathcal{A}, P) , un proceso estocástico es una colección parametrizada de variables aleatorias $\{\mathcal{X}_\alpha(\omega), \alpha \in \Lambda, \omega \in \Omega\}$ definidas en el mismo espacio (Ω, \mathcal{A}, P) con Λ un conjunto no vacío. Cada una de las variables aleatorias del proceso tiene su propia función de distribución de probabilidad y entre ellas pueden estar correlacionadas o no.

Cuando el conjunto Λ sea un subconjunto de los números reales, digamos T , diremos que la colección $\{\mathcal{X}(t, \omega), t \in T, \omega \in \Omega\}$ es un **proceso estocástico**. Cuando el conjunto Λ sea un subconjunto de \mathbb{R}^d , digamos \mathbf{S} , diremos que la colección $\{\mathcal{X}(\mathbf{s}, \omega), \mathbf{s} \in \mathbf{S}, \omega \in \Omega\}$ es un **campo aleatorio**.

Si bien las definiciones que damos a continuación están dadas para procesos estocásticos, lo mismo valen para campos aleatorios reemplazando T por \mathbf{S} y t por \mathbf{s} .

Un proceso estocástico es una función $\mathcal{X}_t(\omega) \doteq \mathcal{X}(t, \omega) : T \times \Omega \rightarrow \mathcal{F}$ tal que:

- Fijado un cierto $t_0 \in \Omega$, $\mathcal{X}(t, \omega_0) : T \rightarrow \mathcal{F}$ es función de t , lo que hemos definido como una realización del proceso (el uso que hacemos aquí de la letra t no es casual ya que en la mayoría de las aplicaciones esta variable suele representar el tiempo).
- Fijado $t_0 \in T$, $\mathcal{X}(t_0, \omega) : \Omega \rightarrow \mathbb{R}$ es una variable aleatoria, cuyo valor dependerá de ω .

Para simplificar la notación llamaremos a $\mathcal{X}(\omega, t)$ simplemente $\mathcal{X}(t)$ sin olvidar que ω está implícitamente en la expresión anterior.

Las posibles elecciones de los conjuntos T y \mathcal{F} nos llevan a definir distintos tipos de procesos:

- Hablaremos de **procesos en tiempo discreto** cuando el espacio origen T sea discreto, generalmente \mathbb{Z} ó \mathbb{N} . Las realizaciones serán ahora sucesiones $\{\mathcal{X}_n\}$ que varían con el tiempo.
- Hablaremos de **procesos en tiempo continuo** cuando el espacio origen T sea \mathbb{R} o un subconjunto de \mathbb{R} , por ejemplo, un intervalo $[a, b]$, la semirecta $[0, \infty)$, etc. En este caso las realizaciones serán funciones usuales del tipo $\mathcal{X}(t), t \in T$.
- Hablaremos de **procesos discretos** cuando el espacio de llegada \mathcal{F} sea discreto, generalmente \mathbb{Z} ó \mathbb{N} .
- Hablaremos de **procesos continuos** cuando cuando el espacio de llegada \mathcal{F} sea \mathbb{R} .

Para $\{t_1, \dots, t_k\} \in T$ y $A \in \mathcal{A}$, la **distribución (marginal) finito dimensional** del proceso se define como $\mu_{(t_1, \dots, t_k)}(A) = P((\mathcal{X}(t_1), \dots, \mathcal{X}(t_k)) \in A) = P(\omega \in \Omega : (\mathcal{X}(t_1, \omega), \dots, \mathcal{X}(t_k, \omega)) \in A)$. Por lo tanto, conocer la distribución de un proceso aleatorio, supone conocer la distribución de todo vector de la forma $(\mathcal{X}(t_1), \dots, \mathcal{X}(t_k))$ siendo $\{t_1, \dots, t_k\}$ cualquier colección finita de instantes.

Sobre un proceso estocástico también podemos definir distintas medidas (que serán generalmente funciones de t) que resumen parte de la información contenida en su distribución.

- **Valor medio** en el instante t :

$$\mu_{\mathcal{X}}(t) = \mathbb{E}(\mathcal{X}(t)).$$

- **Correlación** entre los instantes t_1 y t_2 :

$$R_{\mathcal{X}}(t_1, t_2) = \mathbb{E}(\mathcal{X}(t_1)\mathcal{X}(t_2)).$$

- **Covarianza** entre los instantes t_1 y t_2 :

$$C_{\mathcal{X}}(t_1, t_2) = R_{\mathcal{X}}(t_1, t_2) - \mu_{\mathcal{X}}(t_1)\mu_{\mathcal{X}}(t_2) = \mathbb{E}((\mathcal{X}(t_1) - \mu_{\mathcal{X}}(t_1))(\mathcal{X}(t_2) - \mu_{\mathcal{X}}(t_2))).$$

- **Varianza** en el instante t :

$$\sigma_{\mathcal{X}}^2(t) = C_{\mathcal{X}}(t, t).$$

- **Coefficiente de correlación** entre los instantes t_1 y t_2 :

$$\rho_{\mathcal{X}}(t_1, t_2) = \frac{C_{\mathcal{X}}(t_1, t_2)}{\sqrt{\sigma_{\mathcal{X}}^2(t_1)\sigma_{\mathcal{X}}^2(t_2)}}.$$

Cuando necesitemos probar si estas propiedades estadísticas se cumplen para todo el proceso, no sólo para un instante t , necesitaremos el concepto de **proceso estacionario**. Diremos que un proceso estocástico $\mathcal{X}(t)$ es **estacionario de primer orden** si su función de densidad permanece invariante con respecto a todo desplazamiento en el tiempo. Más precisamente, si f_{t_1} representa la densidad de $\mathcal{X}(t)$ al tiempo t_1 entonces,

$$f_{t_1}(x) = f_{t_1+\tau}(x),$$

para todo desplazamiento τ . La característica principal que identifica a todo proceso estacionario de primer orden, es el hecho que su media es una constante independiente de todo desplazamiento en el tiempo.

Diremos que un proceso \mathcal{X}_t tiene **incrementos independientes** si para toda colección de instantes t_1, \dots, t_k , con $k \geq 3$ los incrementos $\mathcal{X}(t_2) - \mathcal{X}(t_1), \dots, \mathcal{X}(t_k) - \mathcal{X}(t_{k-1})$ son independientes. Diremos que \mathcal{X}_t tiene **incrementos estacionarios** si para todo par de instantes $t_1 < t_2$ la distribución de $\mathcal{X}(t_2) - \mathcal{X}(t_1)$ es independiente de t_1 y t_2 (aunque sí puede depender de la diferencia $t_2 - t_1$).

En lo que sigue, presentamos algunos procesos que son de interés en estadística.

Ejemplo A.35. (Proceso Gaussiano) Un proceso estocástico $\{\mathcal{X}(t) : t \geq 0\}$ se llama **proceso Gaussiano** si para cualquier conjunto finito de índices $\{t_1, \dots, t_k\}$ y cualquier conjunto finito de números reales $\{a_1, \dots, a_k\}$ la variable aleatoria $\sum_{i=1}^k a_i \mathcal{X}_{t_i}$ tiene distribución normal univariada.

Ejemplo A.36. (Proceso de Markov) Un proceso estocástico $\{\mathcal{X}(t) : t \geq 0\}$ se llama **proceso de Markov o Markoviano** si para cualquier sucesión finita de estados $\{t_1, \dots, t_k, t_{k+1}\}$

$$P(\mathcal{X}_{t_{k+1}} | \mathcal{X}_{t_1}, \dots, \mathcal{X}_{t_k}) = P(\mathcal{X}_{t_{k+1}} | \mathcal{X}_{t_k}).$$

Es decir, la distribución de probabilidad condicional del proceso al tiempo t_{k+1} dada la historia entera del proceso hasta el tiempo t_k , depende solamente del estado del proceso en el tiempo t_k .

Ejemplo A.37. (El movimiento Browniano) Un proceso estocástico a valores reales $\{\mathcal{W}(t) : t \geq 0\}$ se llama **movimiento Browniano** si satisface las siguientes tres condiciones:

- (i) $\mathcal{W}(0) = 0$;
- (ii) para cada $t \geq 0$, $\mathcal{W}(t) \sim \mathcal{N}(0, t)$;
- (iii) $\mathcal{W}(t)$ tiene incrementos estacionarios independientes.

Entonces, $\{\mathcal{W}(t) : t \geq 0\}$ es un proceso Gaussiano con media $\mu_{\mathcal{W}}(t) = 0$ y función de autocovarianza $C_{\mathcal{W}}(t_1, t_2) = \min\{t_1, t_2\}$. También puede demostrarse que las trayectorias de \mathcal{W} son continuas con probabilidad uno. Por lo tanto, un movimiento Browniano es un proceso Gaussiano, tiene trayectorias continuas e incrementos estacionarios independientes lo que implica que es Markoviano.

CONCLUSIONES GENERALES

Para un proceso estocástico estacionario a tiempo continuo, hemos definido un estimador de la densidad marginal de tipo k -NN a partir de una muestra i.i.d. utilizando la noción de medida de ocupación y hemos obtenido tasas paramétricas (de orden $n^{-1/2}$) de convergencia. Hemos extendido la definición del estimador al caso no estacionario y si bien en este caso hemos obtenido menores tasas de convergencia (de orden $n^{-1/4}$), teniendo en cuenta que en dimensión finita la velocidad de convergencia depende de la dimensión d del espacio (decreciendo con d) y que nuestro contexto es infinito dimensional, podemos considerar que dichas tasas son muy buenas. Cabe destacar que la irregularidad de las trayectorias es un aspecto clave para obtener estas velocidades de convergencia tanto en el caso estacionario como el no estacionario.

A partir una sucesión dependiente de campos aleatorios d -dimensionales, hemos encontrado velocidades de convergencia de orden $n^{-\alpha}$ con $\alpha < 1/4$ de un estimador de la densidad calculado como una extensión del estimador anterior. La dependencia que presenta la sucesión de campos es una nueva noción, llamada (α, ϕ) -débil, la cual probamos que es más débil que las condiciones de dependencia mixing existentes y que es una clase particular de dependencia $(\alpha, \mathcal{G}, \psi)$ -débil. Además probamos que sucesiones de campos aleatorios (α, ϕ) -débilmente dependientes generan sucesiones de variables aleatorias $(\alpha, \mathcal{G}, \psi)$ -débilmente dependientes y es por ello que desigualdades de tipo Bernstein para este último tipo de dependencia son utilizadas para obtener dichas velocidades de convergencia. En este caso, las desigualdades exponenciales que disponemos no nos han permitido obtener tasas paramétricas de convergencia.

Finalmente, basados en el resultado de consistencia universal para estimadores de regresión en \mathbb{R}^d presentado por [Stone, 1977], probamos un resultado de consistencia similar en el contexto infinito dimensional y lo utilizamos para probar la consistencia de estimadores ya conocidos como son el estimador de k -vecinos más cercanos y el estimador de núcleo. Si bien en el contexto infinito dimensional existen resultados individuales de consistencia para estos estimadores, las

pruebas presentadas aquí mediante la aplicación del resultado general son considerablemente más sencillas que las ya existentes.

BIBLIOGRAFÍA

- [Abraham et al., 2006] Abraham, C., Biau, G., and Cadre, B. (2006). On the kernel rule for function classification. *Ann. Inst. Statist. Math.*, 58(3):619–633.
- [Akaike, 1954] Akaike, H. (1954). An approximation to the density function. *Annals Inst. Stat. Math.*, 6:127–132.
- [Athreya and Lahiri, 2006] Athreya, K. and Lahiri, S. (2006). *Measure theory and probability theory*. Springer.
- [Azaïs and Wschebor, 2009] Azaïs, J. and Wschebor, M. (2009). *Level sets and extrema of random processes and fields*. Wiley.
- [Baíllo and Grané, 2009] Baíllo, A. and Grané, A. (2009). Local linear regression for functional predictor and scalar response. *J. Multivariate Anal.*, 100(1):102–111.
- [Banon, 1978] Banon, G. (1978). Nonparametric identification for diffusion processes. *SIAM J. Control Optim.*, 16(3):380–395.
- [Banon and Nguyen, 1981] Banon, G. and Nguyen, H. T. (1981). Recursive estimation in diffusion model. *SIAM J. Control Optim.*, 19(5):676–685.
- [Biau et al., 2005] Biau, G., Bunea, F., and Wegkamp, M. (2005). Functional classification in hilbert spaces. *IEEE Trans. Inform. Theory*, 51(6):2163–2172.
- [Biau et al., 2010] Biau, G., Cérou, F., and Guyader, A. (2010). Rates of convergence of the functional k -nearest neighbor estimate. *IEEE Trans. Inform. Theory*, 56(4):2034–2040.
- [Blanke, 2004] Blanke, D. (2004). Adaptive sampling schemes for density estimation. *Journal of Statistical Planning and Inference*, 136(9):2898–2917.
- [Blanke and Bosq, 1997] Blanke, D. and Bosq, D. (1997). Accurate rates of density estimators for continuous-time processes. *Statist. Probab. Lett.*, 33(2):185–191.
- [Boente and Fraiman, 1988] Boente, G. and Fraiman, R. (1988). Consistency of a nonparametric estimate of a density function for dependent variables. *J. Multivariate Anal.*, 25(1):90–99.
- [Bosq, 1997] Bosq, D. (1997). Parametric rates of nonparametric estimators and predictors for continuous time processes. *Ann. Stat.*, 25:982–1000.
- [Bosq, 1998] Bosq, D. (1998). *Nonparametric statistics for stochastic processes. Estimation and Prediction*. Lecture Notes in Stat., Vol. 110, 2nd. ed., Springer, New York.
- [Bosq and Davydov, 1999] Bosq, D. and Davydov, Y. (1999). Local time and density estimation in continuous time. *Math. Methods Statist.*, 8(1):22–45.
- [Bowman, 1984] Bowman, A. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360.

- [Cacoullos, 1966] Cacoullos, T. (1966). Estimation of a multivariate density. *Ann. Inst. Statist. Math.*, 18:179–189.
- [Carbon et al., 1996] Carbon, M., Hallin, M., and Tran, L. (1996). Kernel density estimation for random fields: the L_1 theory. *J. Nonparametr. Statist.*, 6(2-3):157–170.
- [Carbon et al., 1997] Carbon, M., Hallin, M., and Tran, L. (1997). Kernel density estimation for random fields (density estimation for random fields). *Statist. Probab. Lett.*, 36(2):115–125.
- [Cardot et al., 2003] Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica*, 13(3):571–591.
- [Cardot et al., 2004] Cardot, H., Goia, A., and Sarda, P. (2004). Testing for no effect in functional linear regression models, some computational approaches. *Comm. Statist. Simulation Comput.*, 33(1):179–199.
- [Castellana and Leadbetter, 1986] Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Process. Appl.*, 21(2):179–193.
- [Cèrou and Guyader, 2006] Cèrou, F. and Guyader, A. (2006). Nearest neighbor classification in infinite dimension. *ESAIM Probab. Stat.*, 10:340–355.
- [Cover and Hart, 1967] Cover, T. and Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Trans. Inform. Theory*, IT-13(22):21–27.
- [Cuevas et al., 2002] Cuevas, A., Febrero, M., and Fraiman, R. (2002). Linear functional regression: the case of fixed design and functional response. *Canad. J. Statist.*, 30(2):285–300.
- [Dabo-Niang, 2004] Dabo-Niang, S. (2004). Kernel density estimator in an infinite-dimensional space with a rate of convergence in the case of diffusion process. *Appl. Math. Lett.*, 17(4):381–386.
- [Devroye, 1981a] Devroye, L. (1981a). On the almost everywhere convergence of nonparametric regression function estimates. *Ann. Statist.*, 9(6):1310–1319.
- [Devroye, 1981b] Devroye, L. (1981b). On the inequality of cover and hart in nearest neighbor discrimination. *IEEE Trans. on Pat. Anal. and Mac. Int.*, PAMI-3(1):75–78.
- [Devroye et al., 1996] Devroye, L., Györfi, L., and Lugosi, G. (1996). *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag.
- [Devroye, 1978] Devroye, L. P. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canad. J. Statist.*, 6(2):179–191.
- [Devroye and Wagner, 1980] Devroye, L. P. and Wagner, T. J. (1980). Distribution-free consistency results in nonparametric discrimination and regression function estimation. *Ann. Statist.*, 8(2):231–239.
- [Dony and Einmahl, 2006] Dony, J. and Einmahl, U. (2006). Weighted uniform consistency of kernel density estimators with general bandwidth sequences. *Electron. J. Probab.*, 11:no. 33, 844–859 (electronic).
- [Doukhan, 1994] Doukhan, P. (1994). *Mixing: properties and examples*. Lecture Notes in Stat., Vol. 85, Springer, Berlin.
- [Doukhan et al., 1984] Doukhan, P., Leon, J., and Portal, F. (1984). Vitesse de convergence dans le théorème central limite pour des variables aleatoires mtlangentes a valeurs dans un espace de hilbert. *R. Acad. Sci. Paris*, 289:305–308.
- [Doukhan and Louhichi, 1999] Doukhan, P. and Louhichi, S. (1999). A new dependence condition and applications to moment inequalities. *Stoch. Process. Appl.*, 84:313–342.

-
- [Doukhan and Neumann, 2006] Doukhan, P. and Neumann, M. (2006). Probability and moment inequalities for sums of weakly dependent random variables, with applications. *Stoch. Process. Appl.*, 117:878–903.
- [Duda and Stork, 2000] Duda, R.O. Hart, P. and Stork, D. (2000). *Pattern Classification, 2nd edition*. Wiley.
- [Epanechnikov, 1969] Epanechnikov, V. (1969). Nonparametric estimation of a multivariate probability density. *Theory of Prob. and its Appl.*, 14:153–158.
- [Ferguson, 1996] Ferguson, T. (1996). *A course in large sample theory*. Chapman and Hall.
- [Ferraty and Romain, 2010] Ferraty, F. and Romain, Y. (2010). *The Oxford Handbook of Functional Data Analysis*. Oxford Handbooks.
- [Ferraty and Vieu, 2003] Ferraty, F. and Vieu, P. (2003). Curves discrimination: a nonparametric functional approach. *Comput. statist. and data anal.*, 44(1-2):161–173.
- [Ferraty and Vieu, 2006] Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis. Theory and Practice*. Springer, New York.
- [Fix and Hodges, 1951] Fix, E. and Hodges, J. (1951). Discriminatory analysis. nonparametric discrimination: consistency properties. *Technical Report 4, Project Number 21-49-004, USAF School of Aviation Medicine, Randolph Field, Texas*.
- [Folland, 1999] Folland, G. B. (1999). *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons Inc., second edition. Modern techniques and their applications, A Wiley-Interscience Publication.
- [Geman and Horowitz, 1980] Geman, D. and Horowitz, J. (1980). Occupation densities. *The Annals of Probab.*, 8(1):1–67.
- [Giné and Guillou, 2002] Giné, E. and Guillou, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(6):907–921. En l’honneur de J. Bretagnolle, D. Dacunha-Castelle, I. Ibragimov.
- [González Manteiga and Vieu, 2007] González Manteiga, W. and Vieu, P. (2007). Statistics for functional data. *Comput. Stat. Data Anal.*, 51:4788–4792.
- [Hall and Horowitz, 2007] Hall, P. and Horowitz, J. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.*, 35(1):70–91.
- [Hallin et al., 2001] Hallin, M., Lu, Z., and Tran, L. (2001). Density estimation for spatial linear processes. *Bernoulli*, 7(4):657–668.
- [Hallin et al., 2004] Hallin, M., Lu, Z., and Tran, L. (2004). Kernel density estimation for spatial processes: the L_1 theory. *Ann. Statist.*, 88:61–75.
- [Härdle and Müller, 1997] Härdle, W. and Müller, M. (1997). Multivariate and semiparametric kernel regression.
- [Hastie et al., 1995] Hastie, T., Buja, A., and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann. Statist.*, 23(1):73–102.
- [Hastie et al., 2001] Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning*. Springer-Verlag, New York.
- [Kutoyants, 2004] Kutoyants, Y. (2004). On invariant density estimation for ergodic diffusion processes. *SORT.*, 28(2):111–124.
- [Kutoyants, 1997] Kutoyants, Y. A. (1997). On unbiased density estimation for ergodic diffusion. *Statist. Probab. Lett.*, 34(2):133–140.

- [Labrador, 2006] Labrador, B. (2006). Almost sure convergence of the k_T -occupation time density estimator. *C. R. Math. Acad. Sci. Paris*, 343(10):665–669.
- [Labrador, 2008] Labrador, B. (2008). Strong pointwise consistency of the k_T -occupation time density estimator. *Statist. Probab. Lett.*, 78(9):1128–1137.
- [Llop et al., 2011] Llop, P., Forzani, L., and Fraiman, R. (2011). On local times, density estimation and supervised classification from functional data. *J. Multivariate Anal.*, 102(1):73–86.
- [Loftsgaarden and Quesenberry, 1965] Loftsgaarden, D. and Quesenberry, C. (1965). A nonparametric estimate of a multivariate density function. *Annals of Math. Stat.*, 36(3):1049–1051.
- [Marron and Nolan, 1988] Marron, J. S. and Nolan, D. (1988). Canonical kernels for density estimation. *Statist. Probab. Lett.*, 7(3):195–199.
- [MatLab, 2009] MatLab (2009). *Versión 7.8.0 (R2009a)*. The MathWorks Inc., Natick, Massachusetts.
- [Moore and Henrichon, 1969] Moore, D. and Henrichon, E. (1969). Uniform consistency of some estimates of a density function. *Ann. Math. Statist.*, 40(4):1499–1502.
- [Moore and Yackel, 1977] Moore, D. and Yackel, J. (1977). Consistency properties of nearest neighbor density function estimators. *Ann. Statist.*, 5(1):143–154.
- [Mynbaev and Martins-Filho, 2010] Mynbaev, K. and Martins-Filho, C. (2010). Bias reduction in kernel density estimation via Lipschitz condition. *J. Nonparametr. Stat.*, 22(1-2):219–235.
- [Nadaraya, 1964] Nadaraya, E. (1964). On estimating regression. *Theory prob. and appl.*, 9:141–142.
- [Nadaraya, 1965] Nadaraya, E. (1965). On non nonparametric estimates for density functions and regression curves. *Theory prob. and appl.*, 10:297–302.
- [Nguyen, 1979] Nguyen, H. (1979). Density estimation in a continuous-time stationary markov process. *Ann. Statist.*, 7(2):341–348.
- [Parzen, 1962] Parzen, E. (1962). On the estimation of a probability density function and the mode. *Ann. Math. Statist.*, 33:1065–1076.
- [Preda and Saporta, 2005a] Preda, C. and Saporta, G. (2005a). Clusterwise pls regression on a stochastic process. *Comput. Statist. and Data Analysis.*, 49(1):99–108.
- [Preda and Saporta, 2005b] Preda, C. and Saporta, G. (2005b). Pls regression on a stochastic process. *Comput. Statist. and Data Analysis.*, 48:149–158.
- [Preda et al., 2007] Preda, C., Saporta, G., and Lévêder, C. (2007). Pls classification of functional data. *Comput. Statist.*, 22(2):223–235.
- [R Development Core Team, 2006] R Development Core Team (2006). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- [Ramsay and Dalzell, 1991] Ramsay, J. and Dalzell, C. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society B*, 53:539–572.
- [Ramsay and Silverman, 1997] Ramsay, J. and Silverman, B. (1997). *Functional Data Analysis*. Springer Series in Statistics, New York.
- [Ramsay and Silverman, 2002] Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis. Method and case studies*. Springer Series in Statistics, New York.

-
- [Ramsay and Silverman, 2005] Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis. 2nd Edition*. Springer Series in Statistics, New York.
- [Ribeiro Jr. and Diggle, 2001] Ribeiro Jr., P. and Diggle, P. (2001). geoR: a package for geostatistical analysis. *R-NEWS*, 1(2):15–18.
- [Rosenblatt, 1956] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc.Natl. Acad. Sci. USA*, 42:43–47.
- [Rudemo, 1982] Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scand. J. Statist.*, 9(2):65–78.
- [Shin, 2008] Shin, H. (2008). An extension of fisher’s discriminant analysis for stochastic processes. *J. Multivariate Anal.*, 99:1191–1216.
- [Spiegelman and Sacks, 1980] Spiegelman, C. and Sacks, J. (1980). Consistent window estimation in nonparametric regression. *Ann. Statist.*, 8(2):240–246.
- [Stone, 1977] Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, 5(4):595–645. With discussion and a reply by the author.
- [Tang et al., 2008] Tang, X., Liu, Y., Zhang, J., and Kainz, W. (2008). *Advances in Spatio-Temporal Analysis*. ISPRS Book Series, Vol. 5.
- [Taylor, 1973] Taylor, S. (1973). *Introduction to measure and integration*. Cambridge University Press.
- [Tran and Yakowitz, 1993] Tran, L. and Yakowitz, S. (1993). Nearest neighbor estimators for random fields. *J. Multivariate Anal.*, 44(1):23–46.
- [Tran, 1990] Tran, L. T. (1990). Kernel density estimation on random fields. *J. Multivariate Anal.*, 34(1):37–53.
- [Vieu, 1993] Vieu, P. (1993). Bandwidth selection for kernel regression: a survey. In *Computer intensive methods in statistics (Louvain-la-Neuve, 1991)*, Statist. Comput., pages 134–149. Physica, Heidelberg.
- [Wagner, 1973] Wagner, T. (1973). Strong consistency of a nonparametric estimate of a density function. *IEEE Trans. Syst. Man Cybern*, pages 289–290.
- [Watson, 1964] Watson, G. (1964). Smooth regression analysis. *Sankhya Series A*, 26:359–372.
- [Weisberg, 2005] Weisberg, S. (2005). *Applied linear regression*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition.
- [Wheeden and Zygmund, 1977] Wheeden, R. and Zygmund, A. (1977). *Measure and integral. An introduction to real analysis*. CRC Press.

ÍNDICE ALFABÉTICO

- L^1 , 101
- L^p , 102
- L^∞ , 102
- σ -álgebra, 98
 - de Borel, 99
 - generada por conjuntos, 99
 - producto, 99

- a.co., 107
- a.s., 107
- ancho de banda, 2
- aprendizaje supervisado, 16

- Bayes
 - error de, 16
 - problema de, 16
 - regla de, 16
- bola, 97
- borelianos, 99

- campo aleatorio, 18, 110
- clasificación supervisada, 15
- clasificador, 16
- conjunto
 - abierto, 97
 - cerrado, 97
 - clausura de un, 97
 - compacto, 98
 - denso, 98
 - frontera de un, 97
 - interior de un, 97
 - medible, 99

- convergencia
 - casi completa, 107
 - casi segura, 107
 - en r -media, 107
 - en distribución, 107
 - en probabilidad, 107
 - orden de, 108

- datos funcionales, 18
- densidad, 105
 - condicional, 106
 - conjunta, 105
 - marginal, 106
- dependencia
 - $(\mathcal{G}, \alpha, \psi)$ -débil, 61
 - (α, ϕ) -débil, 62
 - α -mixing, 60
- desigualdad
 - de Bernstein, 109
 - de Cauchy-Schwartz, 103
 - de Hölder, 103
 - de Jensen, 101
 - exponencial para α -mixing, 110
 - exponencial para dependientes, 109
- discriminación, 15
- distancia, 97
- distribución, 105
 - conjunta, 105
 - empírica, 105
 - marginal, 106

- elemento aleatorio, 103

- distribución de un, 103
- entorno, 97
- error, 80
- error cuadrático
 - integrado, 14
 - medio, 15
 - medio integrado, 15
- espacio
 - σ -finito, 99
 - de medida, 99
 - de probabilidad, 103
 - localmente compacto, 98
 - métrico, 97
 - medible, 98
 - muestral, 103
 - separable, 98
- esperanza, 105
 - condicional, 106
- estimador
 - consistente en media cuadrática, 15
- estimador de densidad
 - de k -NN en \mathbb{R} , 5
 - de núcleo en \mathbb{R} , 4
 - histograma, 3
 - para campos estac. centrados, 63
 - para campos no estacionarios, 69
 - para proc. estacionarios, 21
 - para proc. no estacionarios, 44
- estimador de regresión
 - de k -NN en \mathbb{R} , 9
 - de k -NN para datos funcionales, 86
 - de núcleo en \mathbb{R} , 9
 - de núcleo para datos funcionales, 88
 - regresograma, 8
- eventos, 103
- función
 - continua, 98
 - continua en un punto, 98
 - de autocorrelación, 112
 - de autocovarianza, 112
 - de correlación de un, 112
 - integrable, 100
 - Lipschitz continua, 98
 - medible, 99
 - simple, 100
 - uniformemente continua, 98
- independencia, 106
- integral
 - de una función medible, 100
 - de una función no negativa, 100
 - de una función simple, 100
- leave-one-out, 15
- Ley
 - débil de los grandes números, 108
 - de la esperanza iterada, 106
 - fuerte de los grandes números, 109
- métrica, 97
- medida, 99
 - σ -finita, 99
 - absolutamente continua, 100
 - de Borel, 99
 - finita, 99
 - producto, 100
 - soporte de, 99
 - soporte de una, 99
- medida de probabilidad, 103
 - continuidad de una, 104
 - empírica, 104
 - inducida por un e.a., 103
- momentos, 105
- núcleo regular, 88
- Nadaraya-Watson, 9
- norma, 98

-
- operador de regresión, 80
 - parámetro de suavización, 12
 - pesos, 8
 - de probabilidad, 81
 - normales, 80
 - proceso estocástico, 110
 - (movimiento) Browniano, 113
 - con incrementos estacionarios, 112
 - con incrementos independientes, 112
 - continuo, 111
 - de Markov, 113
 - discreto, 111
 - en tiempo continuo, 111
 - en tiempo discreto, 111
 - estacionario de primer orden, 112
 - Gausiano, 113
 - punto de lebesgue, 102
 - rectángulos, 99
 - regla de clasificación, 16
 - débilmente consistente, 16
 - fuertemente consistente, 16
 - universalmente consistente, 16
 - soporte, 99
 - Teorema
 - central del límite de Lindeberg, 108
 - central del límite para v.a. i.i.d., 108
 - de diferenciación de Lebesgue, 101
 - de la convergencia dominada, 101
 - de Radon-Nikodym, 101
 - de Taylor, 102
 - de Tonelli, 101
 - del Binomio, 102
 - del Valor Medio Integral, 102
 - tiempo local, 21
 - validación cruzada, 14
 - valor medio, 112
 - variable aleatoria, 103
 - varianza, 105, 112
 - vector aleatorio, 103