

Tesis de Maestría

Regresión no paramétrica para datos funcionales  
no estacionarios

Laura Aspirot

Orientador: Gonzalo Perera

Maestría en Ingeniería Matemática.  
Facultad de Ingeniería, Universidad de la República.  
Montevideo, Uruguay.

5 de septiembre de 2008

## Resumen

En este trabajo se estudia el modelo de regresión  $Y = \phi(X) + \varepsilon$  donde la variable  $Y$  es real y la variable explicativa  $X$  es funcional, es decir que  $X \in \mathcal{D}$  donde  $\mathcal{D}$  es un espacio de funciones. Se considera el estimador basado en núcleos tipo Nadaraya-Watson para la función  $\phi$ , definido para cada  $x \in \mathcal{D}$  por

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

donde  $K$  es un núcleo,  $h_n$  es una sucesión positiva que tiende a cero y  $\|\cdot\|$  es una seminorma en el espacio de funciones  $\mathcal{D}$ . Las observaciones  $\{(X_i, Y_i) : i = 1, \dots, n\}$  son no estacionarias y siguen un modelo de dependencia. Los principales resultados son la consistencia y la distribución asintótica del estimador para el modelo no estacionario y dependiente propuesto para la variable  $X$ . Este estudio está motivado por un problema de telecomunicaciones, donde el tráfico en un enlace en una red es dependiente y no estacionario. Los resultados sobre el estimador permiten utilizarlo para estimar la calidad de servicio de extremo a extremo para un flujo de datos multimedia, por ejemplo voz o video en tiempo real, a lo largo de un camino en una red.

# Índice general

<b>1. Introducción</b>	<b>2</b>
<b>2. Regresión no paramétrica</b>	<b>8</b>
2.1. Estimador para variables en $\mathbb{R}^d$ . . . . .	9
2.2. Estimador para variables funcionales . . . . .	12
2.3. Extensiones al caso no estacionario y dependiente . . . . .	17
<b>3. Teorema central del límite</b>	<b>19</b>
3.1. Conjuntos asintóticamente medibles . . . . .	20
3.2. Teorema central del límite . . . . .	22
3.2.1. Teorema central del límite para variables reales . . . . .	22
3.2.2. Teorema central del límite para variables vectoriales . . . . .	24
3.3. Variables asintóticamente medibles . . . . .	26
<b>4. Consistencia y distribución asintótica</b>	<b>28</b>
4.1. Consistencia . . . . .	30
4.2. Distribución asintótica . . . . .	36
<b>5. Aplicaciones</b>	<b>43</b>
5.1. Descripción del problema . . . . .	43
5.2. Un ejemplo básico . . . . .	47
5.3. Datos simulados . . . . .	50
5.4. Datos reales . . . . .	52

# Capítulo 1

## Introducción

Se consideran dos variables aleatorias  $X$  e  $Y$  que satisfacen  $Y = \phi(X) + \varepsilon$ , donde  $\varepsilon$  es otra variable aleatoria centrada e independiente de  $X$ . Este modelo se conoce como modelo de regresión, la función  $\phi$  se llama regresor y la variable aleatoria  $X$  es la variable explicativa. En este modelo la función  $\phi$  puede tener diferentes características, se pueden considerar funciones lineales, o se pueden hacer hipótesis más generales, por ejemplo sobre la regularidad de la función, y considerar funciones continuas o derivables. Por otra parte se puede asumir determinada distribución para las variables (modelo paramétrico) o puede no hacerse ninguna hipótesis explícita sobre la distribución de los datos (modelo no paramétrico).

Uno de los problemas consiste en estimar la función  $\phi$  a partir de observaciones de las variables  $(X, Y)$ . Cuando  $X$  es una variable aleatoria en  $\mathbb{R}^d$  y la variable  $Y$  es real la función  $\phi$  puede estimarse a partir de observaciones  $\{(X_i, Y_i) : i = 1, \dots, n\}$  mediante el estimador basado en núcleos

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

donde se considera  $0/0 = 0$ .  $K$  es un núcleo, es decir una función  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  que cumple que  $K(x) \geq 0 \forall x \in \mathbb{R}^d$  y que  $\int_{\mathbb{R}^d} K(x)dx = 1$ .  $h_n$  es una sucesión de reales positivos que tiende a cero con  $n$  y se denomina ventana, y junto con el núcleo determina el peso de que tiene cada una de las observaciones en la estimación. Este estimador se conoce como estimador de Nadaraya-Watson.

Para interpretar de una manera informal este estimador basta observar que en el modelo  $Y = \phi(X) + \varepsilon$ , con  $\varepsilon$  centrada e independiente de  $X$ ,  $\phi(X)$  es la esperanza condicional  $E(Y|X)$ . En el caso en que  $X$  es una variable aleatoria discreta la estimación de  $\phi(x) = E(Y|X = x)$  se obtiene promediando los  $Y_i$  tales que  $X_i = x$ . En el caso más general, la estimación de  $\phi(x) = E(Y|X = x)$  se hace promediando los  $Y_i$  tales que  $X_i$  está “cerca” de  $x$  y esto lleva a un

estimador de la forma

$$\sum_{i=1}^n Y_i w_n(X_i, x)$$

donde  $w_n(X_i, x)$  son “pesos” que indican cuáles de los datos  $X_i$  están “cerca” de  $x$  y para el estimador de Nadaraya-Watson están determinados por el núcleo  $K$  y la ventana  $h_n$  mediante

$$w_n(X_i, x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}.$$

Otra interpretación surge de suponer que la variable  $(X, Y)$  tiene densidad conjunta en  $\mathbb{R}^{d+1}$ , en ese caso el estimador  $\phi_n$  es el estimador plug-in de la esperanza condicional  $E(Y|X = x)$  cuando se sustituye la densidad por la estimación de densidades por núcleos. Estos aspectos se describen con más detalle en el capítulo 2.

Por otra parte, con el desarrollo de la estadística para datos funcionales (Ramsay y Silverman (2005) presentan un abordaje extenso del tema datos funcionales y Ferraty y Vieu (2006) consideran especialmente el caso no paramétrico), aparece el modelo de regresión no paramétrico para variables funcionales, es decir un modelo de regresión donde la variable explicativa es una función aleatoria  $X$ . En este caso el estimador de la función  $\phi$ , introducido por Ferraty, Goia y Vieu (2002), es

$$\phi_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\|x-X_i\|}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\|x-X_i\|}{h_n}\right)}$$

donde la diferencia con el caso real es que  $\|\cdot\|$  es una seminorma en un espacio de funciones. En ese trabajo los autores obtienen la convergencia completa y tasas de convergencia del estimador para variables débilmente dependientes y estacionarias. Por otra parte Masry (2005) obtiene la normalidad asintótica del estimador para variables débilmente dependientes y con estacionariedad de segundo orden.

En este trabajo se estudia la regresión no paramétrica para un modelo de datos funcionales no estacionarios. La variable  $X$  es una función aleatoria que sigue determinado modelo de dependencia que se explica con detalle en el capítulo 4, que consiste en una mezcla (no estacionaria y dependiente) de variables estacionarias débilmente dependientes. Para este modelo se obtienen resultados análogos a los mencionados en los trabajos existentes, es decir la consistencia y la distribución asintótica del estimador. La distribución asintótica para datos dependientes y no estacionarios se basa en un teorema central del límite para arreglos triangulares de campos aleatorios estacionarios en  $\mathbb{Z}^d$  (Tablar 2007) y en los resultados mencionados anteriormente de Ferraty, Goia y Vieu y de Masry.

Este estudio está motivado por las características no estacionarias y de dependencia del tráfico en una red de datos. El problema que se pretende abordar

consiste en estimar la calidad de servicio de extremo a extremo para un tráfico de datos multimedia (voz o video) en una red. La calidad de servicio de extremo a extremo está determinada por ciertos parámetros que afectan la transmisión, como ser la pérdida de paquetes de datos, el retardo o la variación del retardo. Estos parámetros afectan la calidad, especialmente para las aplicaciones de transmisión de voz y video en tiempo real. Estas aplicaciones son relativamente nuevas en las redes, y a diferencia de otras aplicaciones, como correo o transferencia de archivos, tienen requerimientos más altos para dichos parámetros, y por ejemplo para una conversación o para una video llamada, los retardos son mucho más importantes que para una transferencia de archivos.

El conocimiento de la calidad de servicio que puede brindar una determinada red es de interés tanto para los usuarios como para los proveedores de servicios. Los usuarios conociendo información precisa pueden corroborar si se están cumpliendo los requerimientos que se les garantizaron, mientras que para los proveedores de servicios puede ser información útil para análisis, diseño de la red, ingeniería de tráfico, tarificación de los servicios, etc..

Muchos de los modelos para las redes están basados en teoría de colas, donde existen numerosas referencias de las aplicaciones para redes de datos (Baccelli & Bremaud 2003) (Kleinrock & Gail 1996). El tráfico llega a un enlace o nodo de la red y es procesado según la capacidad del enlace o almacenado en una cola hasta que se procesa y se dirige a otro nodo de la red. Bajo ciertas hipótesis, como por ejemplo suponer un solo cuello de botella, esto es una sola cola, las pérdidas en la red se pueden calcular con un modelo de cola finita o aproximar asumiendo una cola infinita y estudiando el tamaño de la cola, para ver cuándo supera el tamaño máximo real. En estos modelos la hipótesis de que el tráfico llega según un proceso de Poisson es fundamental para poder hacer cálculos explícitos, pero esta hipótesis restringe los modelos a estudiar. Por otra parte al imponer distribuciones más generales en el modelo se hace más difícil obtener resultados en términos de distribución de los retardos o pérdidas y solamente se obtiene información de los comportamientos en media. Sin embargo el retardo medio, por ejemplo, no brinda la información necesaria para estimar la calidad de servicio para las aplicaciones de voz o video, porque interesa más conocer la probabilidad de que éste supere determinado valor y se pueden tener distintos comportamientos en este sentido para un mismo retardo medio.

Otros modelos analíticos se basan en la teoría de grandes desvíos. Por ejemplo las pérdidas, que se pueden aproximar como el resultado de tener un tamaño de cola superior a determinado valor, son eventos del tipo grandes desvíos. Surgen así varios resultados, válidos en diferentes hipótesis. Se consideran distintos regímenes asintóticos y algunos modelos se describen a continuación. Si se considera el número de usuarios tendiendo a infinito, con el crecimiento de la capacidad y del tamaño máximo de la cola proporcional al número de usuarios, este modelo se denomina “régimen de muchas fuentes” (Kelly 1996). Otro modelo considera el número de usuarios tendiendo a infinito, con la capacidad creciendo proporcionalmente al número de usuarios, pero de modo que el tamaño máximo de la cola crece más lentamente que el número de usuarios. Este régimen, donde el cociente entre el tamaño máximo de la cola y el número de usuarios tiende a

cero (“régimen de buffer pequeño” (Ozturk, Mazumdar & Likhanov 2004)) se utiliza para analizar por ejemplo el corazón de la red. La validez en toda la red depende de la hipótesis de la existencia de un único cuello de botella y por otra parte algunos trabajos que permiten extender los resultados de grandes desvíos a redes dependen fuertemente de la topología de la red o utilizan algoritmos complejos que también dependen de esta topología.

Otro enfoque del problema consiste en realizar simulaciones. Sin embargo muchos de los aspectos que interesan estudiar se consideran eventos raros, es decir que son eventos muy poco probables como por ejemplo la pérdida de paquetes en la red. El problema central consiste en simular adecuadamente este tipo de eventos, utilizándose diversas técnicas para esto (De Boer 2000). Por otra parte la realización de simulaciones generalmente es muy costosa en términos de cálculo y también en el tiempo necesario para su realización.

En todos los análisis mencionados antes, los autores asumen como hipótesis cierta independencia o dependencia débil de los datos, por ejemplo respecto del tiempo, o independencia entre los usuarios cuando el número de usuarios tiende a infinito. Por otra parte también existen modelos que consideran la dependencia y afirman que no es posible modelar el comportamiento de Internet sin considerar dependencias largas en el tiempo. En estos casos la predicción depende fuertemente del ajuste a un modelo concreto (Willinger, Paxon, Reidi & Taqqu 2001).

Por las dificultades de encontrar modelos analíticos que tengan buen desempeño y también gracias a mayores posibilidades tecnológicas muchos de los estudios se hacen en base a medidas explícitas dentro de la red. Uno de los objetivos al estudiar el comportamiento de Internet ha sido el intento de estandarización para las medidas de varios parámetros, como ser pérdidas, retardo, variación del retardo, como puede verse en distintas normas internacionales, desarrolladas principalmente por el grupo IPPM (IP Performance Metrics) (IPPM 2008) de la IETF (Internet Engineering Task Force) y también por la ITU (International Telecommunication Union).

Existen diversas herramientas de software tanto para mediciones activas, donde se introduce tráfico adicional y controlado en la red a efectos de medir, como pasivas, donde el software instalado en algunos puntos de la red es el que realiza las mediciones. Parte de la información que se obtiene de mediciones en distintos nodos de la red a veces está sistematizada y disponible, como en el caso de CAIDA (Cooperative Association for Internet Data Analysis) (CAIDA 2008).

Las herramientas de medición pasiva permiten obtener información en cada nodo de la red o en una red del mismo propietario pero no son adecuadas para las medidas de extremo a extremo. Estas herramientas muchas veces son usadas por los proveedores de servicios para analizar cada cierto tiempo el estado de la red, contabilizando por ejemplo la cantidad de tráfico que llega a determinado nodo, el tipo de aplicación, etc..

Las medidas activas se basan en el envío de tráfico de prueba, donde se envían paquetes destinados a realizar las medidas, con distintas características. Por ejemplo para estudiar la capacidad del cuello de botella en una conexión se puede enviar tráfico de modo de provocar congestión. Por otra parte interesa

medir sin afectar la calidad de la red, de modo que generalmente se busca añadir el menor tráfico posible. Para medir la capacidad del cuello de botella también existen técnicas basadas en pares de paquetes. Estas técnicas asumen que si se envían paquetes de a pares, eligiendo adecuadamente el tiempo entre envíos para los pares, la mayoría de estos pares de paquetes permanecen juntos en la cola en el cuello de botella, de modo que el tiempo entre paquetes en el destino permite calcular la capacidad del enlace. Por otra parte para medir los retardos o las pérdidas existen diversas técnicas basadas en el envío de ráfagas de paquetes livianos. En algunos casos el tráfico de prueba utilizado está asociado a algún modelo para el tráfico en Internet y los paquetes de datos tienen determinada distribución para su tamaño o para los tiempos de envío, por ejemplo siguiendo un proceso de Poisson, aunque también se utilizan ráfagas de paquetes periódicos, trazas de tráfico, etc..

Luego de obtenidas las mediciones hay un posterior tratamiento estadístico de los datos, con el objetivo de predecir comportamientos futuros, explicar lo observado o validar un modelo. En muchos casos se realiza un tratamiento descriptivo. Para muchas de las aplicaciones los comportamientos en media, como por ejemplo el retardo medio para aplicaciones de voz, pueden no servir como herramienta para distinguir entre calidades de servicio diferentes.

El problema abordado en este trabajo consiste, a grandes rasgos, en estimar la calidad de servicio para una conexión entre dos extremos de una red en la que se desea cursar tráfico multimedia (voz o video). Para esto se asume que existe una función que relaciona el estado de la red con la calidad de servicio para este tráfico. El estado de la red está determinado por el tráfico que se esté cursando en ese momento, así como la calidad de servicio del tráfico que se desea analizar se verá afectada por el resto del tráfico que esté presente en la red. Se considera que el estado de la red se puede caracterizar mediante tráfico de prueba, donde por ejemplo el tiempo a la llegada entre dos paquetes de prueba brinda una medida de cuánto tráfico hay en la red, ya que la separación entre los tiempos de llegada depende del volumen de tráfico de la red que fue encolado entre los dos paquetes de prueba. Este tráfico de prueba consiste en una serie de experimentos con ráfagas periódicas de que no generan demasiada carga en la red. Del mismo modo se asume que la calidad de servicio para el tráfico multimedia se determina también con tráfico de prueba multimedia, donde en este caso este tráfico genera mayor carga sobre la red. La función que relaciona el estado de la red con la calidad de servicio se obtiene a partir de los datos mencionados, es decir que conociendo para una secuencia de datos la relación entre el estado de la red y la calidad de servicio se infiere la función que los relaciona. Luego para tener una estimación de la calidad de servicio sólo será necesario conocer el nuevo estado de la red y usar la función estimada a partir de las observaciones. Esto permite conocer la calidad de servicio sin tener que enviar tráfico de prueba multimedia, y por lo tanto sin sobrecargar la red. Es decir que los parámetros de calidad de servicio estarán representados por la variable  $Y$  y se busca estimarlos a partir de las mediciones del estado de la red que determinarán la variable  $X$ . Se desea estimar la función  $\phi$  que relaciona estas dos variables mediante observaciones de las variables  $(X, Y)$  y luego, para observaciones nuevas de la variable  $X$  estimar



$Y$  conociendo únicamente esta observación y la estimación previa de  $\phi$ . En el capítulo 5 se describe con detalle el tipo de medidas que se realizan.

El trabajo se organiza de la siguiente manera. En el capítulo 2 se revisan los conceptos básicos y los resultados existentes en regresión no paramétrica tanto para datos reales como para datos funcionales. En el capítulo 3 se presentan resultados sobre campos aleatorios que serán fundamentales como herramienta para extender los resultados sobre regresión para datos funcionales débilmente dependientes y estacionarios a un caso dependiente y no estacionario, que es el resultado principal de este trabajo y se presenta, junto con la descripción detallada del problema en el capítulo 4. En el capítulo 5 se presentan los resultados aplicados a simulaciones de un enlace en una red de datos.

Los resultados que se presentan aquí forman parte de dos trabajos que se citan a continuación. El primero incluye algunos resultados sobre consistencia del estimador y resultados de su aplicación a simulaciones y datos reales, el segundo consiste en los resultados sobre la distribución asintótica del estimador tipo Nadaraya-Watson para variables funcionales.

Aspirot, L., Bazzano, B., Belzarena, P., Perera, G.. End-To-End Quality of Service Prediction Based On Functional Regression. Third International Working Conference on Performance Modelling and Evaluation of Heterogeneous Networks (HET-NET's), Ilkley, 2005.

Aspirot, L., Bertin, K., Perera, G.. Asymptotic normality of the Nadaraya-Watson estimator for non-stationary functional data and applications to telecommunications. X Congreso Latinoamericano de Probabilidad y Estadística Matemática (CLAPEM), Lima, 2007. Enviado para evaluación al Journal of Nonparametric Statistics en 2007.

## Capítulo 2

# Regresión no paramétrica

En este capítulo se describen algunos resultados sobre regresión no paramétrica y se plantean sus posibles extensiones, que serán el contenido del capítulo 4. En la sección 2.1 se presentan resultados para variables en  $\mathbb{R}^d$  y en la sección 2.2 se presentan resultados para el caso funcional. En la sección 2.3 se plantea el problema de extender los resultados existentes para datos no estacionarios y dependientes, motivado en el problema de telecomunicaciones que se quiere analizar en este trabajo.

En el modelo de regresión se representa la relación entre dos variables aleatorias  $X$  e  $Y$  mediante la ecuación

$$Y = \phi(X) + \varepsilon,$$

donde  $\varepsilon$  es una variable aleatoria, centrada e independiente de  $X$  que representa los errores y  $\phi$  es una función no aleatoria. Este modelo incluye varios problemas diferentes. Uno de los problemas consiste en conocer la función  $\phi$ . Esta función se puede estimar a partir de observaciones de las variables  $(X, Y)$ . Luego mediante nuevas observaciones se puede validar el modelo o también se pueden hacer predicciones sobre la variable  $Y$ , es decir que conociendo  $X$  y la estimación de  $\phi$  se estiman futuros valores de  $Y$ .

Para el problema de regresión se pueden asumir diferentes modelos para la función  $\phi$ , por ejemplo se puede considerar que  $\phi$  es una función lineal, este modelo se conoce como regresión lineal y existe una gran variedad de aplicaciones. En ese caso para conocer la función basta estimar un cantidad finita de parámetros, siendo este un modelo paramétrico. Un modelo no paramétrico para la función  $\phi$  consiste en hacer hipótesis generales sobre esta función (continuidad, derivabilidad, etc.) pero la función no queda determinada por una cantidad finita de parámetros. También se pueden realizar hipótesis sobre la distribución de las variables  $(X, Y)$  de modo que los datos sigan un modelo paramétrico, mientras que en el caso no paramétrico se asumen hipótesis generales sobre la distribución, (momentos, decaimiento de las colas, dependencia débil, etc.).

Una referencia para una introducción general al problema de regresión no paramétrica en el caso de variables en  $\mathbb{R}^d$  es el libro de Nadaraya (1989), del cual se resumen resultados, donde las demostraciones están presentadas para el caso de variables independientes e idénticamente distribuidas (i.i.d.) en  $\mathbb{R}$ , pero son análogas para el caso de variables i.i.d. en  $\mathbb{R}^d$ . Las referencias para el caso funcional se encuentran en diversos artículos (Ferraty, Goia & Vieu 2002) (Ferraty & Vieu 2004) (Masry 2005) (Oliveira 2005) (Ferraty & Vieu 2006) y existen otras referencias relacionadas, por ejemplo estimación de densidades o estimación de la moda para datos funcionales, que se citan más adelante en este capítulo.

## 2.1. Estimador para variables en $\mathbb{R}^d$

Se considera una variable aleatoria  $X$  en  $\mathbb{R}^d$ , una variable aleatoria  $Y$  en  $\mathbb{R}$ , una función  $\phi : \mathbb{R}^d \rightarrow \mathbb{R}$  y la regresión  $Y = \phi(X) + \varepsilon$ , con  $\varepsilon$  otra variable aleatoria en  $\mathbb{R}$ , centrada e independiente de  $X$ . Un estimador para inferir la función  $\phi$  a partir de observaciones  $\{(X_i, Y_i) : i = 1, \dots, n\}$  fue introducido por Nadaraya y por Watson en 1964 (Nadaraya 1964), (Watson 1964). Este estimador se conoce como estimador de Nadaraya-Watson de la función  $\phi$  y está dado por la siguiente fórmula para la función  $\phi_n : \mathbb{R}^d \rightarrow \mathbb{R}$

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

si  $\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \neq 0$  y si  $\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) = 0$  se define  $\phi_n(x) = 0$ .  $K$  es un núcleo, es decir una función  $K : \mathbb{R}^d \rightarrow \mathbb{R}$  que cumple que  $K(x) \geq 0 \forall x \in \mathbb{R}^d$  y  $\int_{\mathbb{R}^d} K(x) dx = 1$ , y  $h_n$  es una sucesión de reales positivos, que tiende a 0 con  $n$ , que se denomina ventana.

Para interpretar el estimador se puede ver que  $\phi(X)$  es la esperanza condicional  $E(Y|X)$ ,

$$E(Y|X) = E(\phi(X) + \varepsilon|X) = E(\phi(X)|X) + E(\varepsilon|X)$$

y como  $\varepsilon$  es independiente de  $X$  y centrada  $E(Y|X) = \phi(X) + E(\varepsilon) = \phi(X)$ .

Sobre la fórmula del estimador, si se consideran variables discretas, para estimar  $E(Y|X)$  basta estimar  $\phi(x) = E(Y|X = x)$  para  $x$  en el recorrido de  $X$ . En este caso  $E(Y|X = x)$  se puede estimar con el promedio de los  $Y_i$  tales que  $X_i = x$ , es decir,

$$\phi_n(x) = \sum_{i=1}^n Y_i 1_{\{X_i=x\}}$$

donde  $1_A$  es la indicatriz del conjunto  $A$ . Cuando la variable  $X$  no es discreta si se estima  $\phi(x) = E(Y|X = x)$  promediando los  $Y_i$  tales que  $X_i$  está “cerca”

de  $x$  y se obtiene un estimador de la forma

$$\phi_n(x) = \sum_{i=1}^n Y_i w_n(X_i, x)$$

donde  $w_n(X_i, x)$  son “pesos” que indican si los datos están “cerca” o “lejos” de  $x$  para alguna medida de distancia definida en el espacio de las variables  $X$ . En este caso los pesos están determinados por

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

donde la ventana regula el peso que tendrá cada dato en la estimación.

Otra interpretación del estimador es la siguiente. Suponiendo que la variable  $(X, Y)$  tiene densidad conjunta  $v$  en  $\mathbb{R}^{d+1}$  el estimador  $\phi_n$  es el estimador plug-in de la esperanza condicional  $E(Y|X = x)$  cuando se sustituye la densidad por la estimación de densidades por núcleos.

$$\phi(x) = E(Y|X = x) = \frac{\int_{\mathbb{R}} yv(x, y)dy}{\int_{\mathbb{R}} v(x, y)dy}$$

El estimador plug-in que se obtiene reemplazando la densidad  $v$  por la estimación de densidades por núcleos  $v_n$  es

$$\phi_n(x) = \frac{\int_{\mathbb{R}} yv_n(x, y)dy}{\int_{\mathbb{R}} v_n(x, y)dy}$$

donde

$$v_n(x, y) = \frac{1}{nh_n^{d+1}} \sum_{i=1}^n \tilde{K}\left(\frac{x-X_i}{h_n}, \frac{y-Y_i}{h_n}\right),$$

siendo  $\tilde{K}$  un núcleo en  $\mathbb{R}^{d+1}$ . Definiendo  $K(x) = \int_{\mathbb{R}} \tilde{K}(x, y)dy$  y con el cambio de variable  $\frac{y-Y_i}{h_n} = u$  se tiene que

$$\begin{aligned} \int_{\mathbb{R}} v_n(x, y)dy &= \frac{1}{nh_n^{d+1}} \sum_{i=1}^n \int_{\mathbb{R}} \tilde{K}\left(\frac{x-X_i}{h_n}, \frac{y-Y_i}{h_n}\right) dy \\ &= \frac{1}{nh_n^d} \sum_{i=1}^n \int_{\mathbb{R}} \tilde{K}\left(\frac{x-X_i}{h_n}, u\right) du \\ &= \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \end{aligned}$$

Análogamente, con el mismo cambio de variable  $\frac{y-Y_i}{h_n} = u$  y la hipótesis adicional

de que  $\int_{\mathbb{R}} u \tilde{K}(x, u) du = 0$  se obtiene

$$\begin{aligned} \int_{\mathbb{R}} y v_n(x, y) dy &= \frac{1}{nh_n^{d+1}} \sum_{i=1}^n \int_{\mathbb{R}} y \tilde{K} \left( \frac{x - X_i}{h_n}, \frac{y - Y_i}{h_n} \right) dy \\ &= \frac{1}{nh_n^d} \sum_{i=1}^n \int_{\mathbb{R}} (Y_i + uh_n) \tilde{K} \left( \frac{x - X_i}{h_n}, u \right) du \\ &= \frac{1}{nh_n^d} \sum_{i=1}^n Y_i K \left( \frac{x - X_i}{h_n} \right) \end{aligned}$$

El estimador plug-in definido antes es

$$\phi_n(x) = \frac{\int_{\mathbb{R}} y v_n(x, y) dy}{\int_{\mathbb{R}} v_n(x, y) dy} = \frac{\frac{1}{nh_n^d} \sum_{i=1}^n Y_i K \left( \frac{x - X_i}{h_n} \right)}{\frac{1}{nh_n^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h_n} \right)}$$

Se denotan la siguientes expresiones como  $g_n(x)$  y  $f_n(x)$

$$\begin{aligned} g_n(x) &= \frac{1}{nh_n^d} \sum_{i=1}^n Y_i K \left( \frac{x - X_i}{h_n} \right) \\ f_n(x) &= \frac{1}{nh_n^d} \sum_{i=1}^n K \left( \frac{x - X_i}{h_n} \right) \end{aligned}$$

de modo que el estimador se escribe como  $\phi_n(x) = \frac{g_n(x)}{f_n(x)}$ .

La normalización de las variables  $f_n(x)$  y  $g_n(x)$  se realiza mediante  $nh_n^d$ , es decir que depende de la dimensión del espacio, más concretamente del volumen de la bola de centro 0 y radio  $h_n$ . El problema de dimensionalidad, es decir que para estimar en dimensiones mayores son necesarios más datos porque la concentración de una cantidad fija de datos disminuye al aumentar la dimensión, queda reflejado en esta normalización. Gran parte de los resultados existentes asumen densidades para las variables  $(X, Y)$ . Esta hipótesis es necesaria por la forma en que se realiza la normalización, al considerarse  $g_n(x)$  y  $f_n(x)$  normalizados por  $nh_n^d$ . Bajo hipótesis más generales, sin asumir densidades se puede realizar la normalización utilizando los valores esperados, es decir normalizando por  $E \left[ K \left( \frac{x - X_i}{h_n} \right) \right]$ , obteniéndose así la convergencia en probabilidad (Greblicki, Krzyzak & Pawlak 1984).

Para el caso donde se asumen densidades se demuestra la convergencia en probabilidad, es decir que

$$\phi_n(x) \xrightarrow{P} \phi(x),$$

cuando  $n \rightarrow \infty$ , donde  $\xrightarrow{P}$  indica convergencia en probabilidad. Este resultado se basa en las convergencias en media cuadrática de  $f_n(x)$  y de  $g_n(x)$ , en que el estimador es asintóticamente insesgado, y del mismo modo que para los

resultados de distribución asintótica, la demostración se concluye realizando descomposiciones adecuadas del estimador  $\phi_n(x)$ .

También se prueba convergencia y normalidad asintótica, es decir que

$$\sqrt{nh_n}(\phi_n(x) - \phi(x)) \xrightarrow{w} N(0, \sigma^2(x)),$$

donde  $\xrightarrow{w}$  indica convergencia en distribución y  $N(0, \sigma^2(x))$  representa una variable con distribución gaussiana con media 0 y varianza  $\sigma^2(x)$ . Bajo diferentes hipótesis de momentos para las variables  $(X, Y)$  y diferentes velocidades de convergencia a 0 de la ventana se demuestran estos resultados. En todos los casos la velocidad de convergencia a 0 de la ventana debe ser pequeña comparada con la velocidad con la que  $n$  tiende a  $\infty$  y además depende de la dimensión del espacio (por ejemplo en  $\mathbb{R}^d$  es necesario que  $\lim_{n \rightarrow \infty} nh_n^d = 0$ ). Bajo otras hipótesis se prueban resultados más fuertes, como es la convergencia uniforme casi segura, es decir que

$$\sup_{a \leq x \leq b} |\phi_n(x) - \phi(x)| \xrightarrow{c.s.} 0,$$

donde  $\xrightarrow{c.s.}$  indica convergencia casi segura cuando  $n \rightarrow \infty$ . También se obtiene la distribución asintótica gaussiana de modo uniforme, es decir que si se define la función  $\hat{\phi}_n$  de modo que para  $a \leq x \leq b$ ,  $i = 1, \dots, s_n$ ,  $\hat{\phi}_n(x) = \phi(x_i)$  para todo  $x_i \leq x < x_{i+1}$ , bajo ciertas hipótesis se puede tener de manera explícita la distribución asintótica de

$$\sqrt{nh_n} \max_{a \leq x \leq b} \left| \frac{\hat{\phi}_n(x) - \phi(x)}{\sigma_n(x)} \right|$$

donde  $\sigma_n^2(x)$  es un estimador de la varianza. Este resultado permite la construcción de intervalos de confianza para toda la función  $\phi$  a partir de las estimaciones en los puntos  $x_i$  con  $i = 1, \dots, s_n$  cuando la varianza asintótica es desconocida.

Resultados análogos de convergencia y normalidad asintótica se obtienen para el caso de variables débilmente dependientes (más específicamente, para variables  $\alpha$ -mixing) (Robinson 1983). En el caso de variables dependientes los resultados sobre distribución asintótica se basan en el método de bloques de Bernstein, que permite aproximar una suma de variables dependientes por una suma de variables independientes y luego aplicar un teorema central del límite para arreglos triangulares de las variables independientes.

## 2.2. Estimador para variables funcionales

Para el caso en que, para cada  $n \geq 1$ ,  $X_n$  es una función aleatoria en un espacio de funciones  $\mathcal{D}$  con una seminorma  $\|\cdot\|$ , Ferraty, Goia y Vieu (2002) introducen el estimador  $\phi_n$  para el modelo de regresión, definido por

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

donde se considera  $0/0 = 0$ .

En esta sección se enunciarán sin demostración los resultados de consistencia y distribución asintótica del estimador bajo diferentes hipótesis. Todos los trabajos que aquí se refieren buscan alguna manera de resolver el problema de la dimensionalidad, ya que a diferencia del caso real se está trabajando en espacios de funciones, que tienen dimensión infinita. El problema que abordan los diferentes autores consiste en encontrar algún tipo de condición que garantice la concentración de las observaciones.

En el trabajo de Ferraty et al. (2002) el estimador finalmente es usado para la estimación de series temporales, usando un conjunto continuo de valores anteriores de la serie temporal. Los autores trabajan con variables débilmente dependientes y estacionarias y abordan el problema de dimensionalidad mediante la dimensión fractal del proceso funcional, obteniendo convergencia completa y tasas de convergencia para el estimador. Se resumen aquí los principales resultados.

**Teorema 2.1** (Teorema 4.1 (Ferraty et al. 2002)). *Sean  $\mathcal{D}$  un espacio vectorial de funciones con una seminorma  $\|\cdot\|$  y  $(X_n, Y_n)_{n \geq 1}$  una sucesión de variables estacionarias donde  $X_n \in \mathcal{D}$ ,  $Y_n \in \mathbb{R}$ , y se cumplen las siguientes hipótesis.*

**H 2.2.** *La función  $\phi$  es continua en  $x$ .*

**H 2.3.**  *$K$  es una función real, no negativa, con soporte en  $[0, \xi]$  para  $\xi > 0$  y tal que  $\int_{\mathbb{R}} K(u) du = 1$ .*

**H 2.4.**  *$K$  es una función Lipschitz, es decir que existe  $L > 0$  tal que, para todo  $u, v \in [0, \xi]$ ,  $|K(u) - K(v)| \leq L|u - v|$ .*

**H 2.5.** *Existen  $p \in \mathbb{N}$ ,  $p > 0$  y  $C > 0$  tales que, para todo  $t > 0$ ,*

$$P(|Y| > t) \leq Ct^{-p}.$$

**H 2.6.** *La sucesión  $(X_n, Y_n)_{n \in \mathbb{Z}}$  es  $\alpha$ -mixing con coeficiente de mixing  $\alpha$  y existen  $a > 2$  y  $c > 0$  tales que, para todo  $k \in \mathbb{N}$ ,*

$$\alpha(k) \leq ck^{-a}.$$

**H 2.7.** *Existen los segundos momentos condicionales y están acotados para todo  $i, j \geq 1$ , es decir que*

$$E(|Y_i Y_j| | X_i, X_j) \leq M < \infty.$$

**H 2.8.** *Para cada  $x \in \mathcal{D}$  existen  $\delta(x) > 0$  y  $c(x) > 0$  tales que*

$$\lim_{h \rightarrow 0^+} \frac{P(\|X - x\| \leq h)}{h^{\delta(x)}} = c(x)$$

*y para todo  $i \neq j$ , existe  $\delta_0(x) > 0$  tal que*

$$\lim_{h \rightarrow 0^+} \frac{P(\|X_i - x\| \leq h, \|X_j - x\| \leq h)}{h^{\delta(x) + \delta_0(x)}} = c_{i,j}(x),$$

*con  $0 < c_1(x) \leq c_{i,j}(x) \leq c_2(x) < \infty$ .*

**H 2.9.** Se cumplen las hipótesis H 2.5, H 2.6, H 2.8 con  $p > 4$ ,  $a > \frac{\delta(x)}{\delta^*(x)}$ , donde  $\delta^*(x) = \min\{\delta(x), \delta_0(x)\}$ , y existen  $\theta, C_1, C_2 > 0$  tales que

$$C_2 n \frac{4a + p(3 - a)}{\delta(x)(a + 1)p} + \theta \leq h_n \leq C_1 n \frac{q}{q\delta(x) - a\delta^*(x)} - \theta,$$

con  $2/p = 1 - 1/q$ .

Entonces  $\phi_n(x)$  converge completamente a  $\phi(x)$  cuando  $n \rightarrow \infty$ , es decir que para todo  $\varepsilon > 0$

$$\sum_{n=1}^{\infty} P(|\phi_n(x) - \phi(x)| > \varepsilon) < \infty.$$

Bajo hipótesis similares, pero que se cumplen de manera uniforme para todo  $x \in \mathcal{D}$ , y si  $\phi$  es uniformemente continua en  $\mathcal{S} \subset \mathcal{D}$ , los teoremas anteriores se generalizan a resultados uniformes, es decir que  $\sup_{x \in \mathcal{S}} |\phi_n(x) - \phi(x)|$  converge completamente a 0 cuando  $n \rightarrow \infty$ .

Además de este teorema existen resultados sobre la velocidad de convergencia del estimador (Ferraty et al. 2002). Los autores también realizan varios comentarios sobre las hipótesis. Algunos comentarios sobre las hipótesis y sobre la demostración se presentan en lo que sigue.

Las hipótesis H 2.5 y H 2.7 son sobre el decaimiento de las colas de la variable  $Y$  y sobre los momentos condicionales.

La hipótesis H 2.6 es una hipótesis de dependencia débil de las variables. El coeficiente de  $\alpha$ -mixing para la sucesión de variables aleatorias  $(X_n, Y_n)$  se define como

$$\alpha(k) = \sup\{|P(A \cap B) - P(A)P(B)| : A \in \sigma_{-\infty}^0, B \in \sigma_k^{+\infty}\},$$

donde  $\sigma_{-\infty}^0$  y  $\sigma_k^{+\infty}$  son las  $\sigma$ -álgebras generadas por  $\{(X_n, Y_n) : n \leq 0\}$  y  $\{(X_n, Y_n) : n \geq k\}$  respectivamente.

En este artículo los autores introducen la dimensión fractal (H 2.8) del proceso  $X$  para resolver el problema de dimensionalidad. Estas hipótesis son sobre la concentración de la variable  $X$  y sobre la distribución conjunta de  $(X_i, X_j)$ . La hipótesis sobre dimensión fractal del proceso  $X$  garantiza la concentración de las observaciones alrededor de  $x$  necesaria para estimar  $\phi(x)$ . La normalización en el estimador para demostrar la convergencia depende de esta dimensión fractal, es decir que la normalización que para  $X \in \mathbb{R}^d$  (asumiendo densidades) es  $nh_n^d$  en este caso es  $nh_n^{\delta(x)}$  donde  $\delta(x)$  está definida en la hipótesis H 2.8. Se puede interpretar que localmente el comportamiento es como en un espacio de dimensión finita, con dimensión  $\delta(x)$ . Análogamente al caso en  $\mathbb{R}^d$ , este valor, junto con las condiciones de dependencia débil, determina la velocidad de convergencia de la ventana a 0 para obtener la convergencia y las tasas de convergencia. La demostración está basada en el cálculo del límite de  $E(f_n(x))$  y de  $E(g_n(x))$ , donde se prueba que existe una función  $f(x)$  tal que

$$\lim_{n \rightarrow \infty} E(f_n(x)) = f(x),$$



$$\lim_{n \rightarrow \infty} E(g_n(x)) = \phi(x)f(x),$$

y luego en la convergencia completa a 0 de  $f_n(x) - E(f_n(x))$  y de  $g_n(x) - E(g_n(x))$ . Una descomposición adecuada de  $\phi_n(x) - \phi(x)$  permite reducir el problema al estudio de la convergencia de  $f_n(x)$  y  $g_n(x)$ . A diferencia del caso real, donde  $E(f_n(x))$  converge a la densidad de la variable  $X$  en el punto  $x$ , en este contexto el límite  $f(x)$  depende de  $\delta(x)$ .

Para la demostración de la convergencia se utilizan desigualdades exponenciales (donde la cota depende en forma exponencial de los momentos de  $S_n$ ) para acotar la probabilidad  $P(|S_n| > \varepsilon)$ , siendo  $S_n$  una suma de variables débilmente dependientes ( $\alpha$ -mixing) y luego desigualdades de covarianza de variables mixing para acotar los momentos de  $S_n$ . La convergencia completa se obtiene además fijando la velocidad de convergencia a cero de la ventana  $h_n$ . En esa velocidad intervienen los parámetros que determinan el decaimiento del coeficiente de mixing y de las colas de la variable  $Y$ , donde la relación entre estos parámetros está dada por la hipótesis H 2.9.

Masry (2005) prueba la normalidad asintótica de  $\phi_n$  para variables débilmente dependientes asumiendo estacionariedad de segundo orden. El resultado se enuncia a continuación.

**Teorema 2.10** (Corolario 2 (Masry 2005)). *Sea  $\mathcal{D}$  un espacio vectorial de funciones con una seminorma  $\|\cdot\|$  y sea  $(X_n, Y_n)_{n \geq 1}$  una sucesión de variables donde  $X_n \in \mathcal{D}$ ,  $Y_n \in \mathbb{R}$  y se cumplen las siguientes hipótesis.*

**H 2.11.** *La función  $K$  es positiva con soporte en  $[0, 1]$ .*

**H 2.12.** *La función  $\phi$  cumple que para todo  $u, v \in \mathcal{D}$*

$$|\phi(u) - \phi(v)| \leq L\|u - v\|^\beta,$$

*donde  $\beta > 0$  y  $L$  es una constante positiva.*

**H 2.13.** *La función  $g_2$  definida por*

$$g_2(u) = \text{var}(Y_j | X_j = u)$$

*es independiente de  $j$  y es continua en un entorno de  $x$ .*

**H 2.14.** *Existe  $\nu > 2$  tal que  $E|Y_i|^\nu < \infty$  y la función  $g_\nu$  definida por*

$$g_\nu(u) = E(|Y_i - \phi(x)|^\nu | X_i = u)$$

*no depende de  $i$  y es continua en un entorno de  $x$ .*

**H 2.15.** *La función  $g$  definida para  $i \neq j$  por*

$$g(u, v) = E((Y_i - \phi(x))(Y_j - \phi(x)) | X_i = u, X_j = v)$$

*es continua en algún entorno de  $(x, x)$ .*

**H 2.16.** Existen funciones no negativas  $\psi$  y  $\tilde{\psi}$  definidas en  $\mathbb{R}^+$ , con  $\psi(0) = \tilde{\psi}(0) = 0$ , continuas en un entorno de 0, y  $c_1$  y  $c_2$  también no negativas definidas en  $\mathcal{D}$  tales que para todo  $h > 0$ , cuando  $u \rightarrow 0$

$$P[\|X_i - x\| \leq h] = c_1(x)\psi(h),$$

$$\sup_{i \neq j} P(\|X_i - x\| \leq h, \|X_j - x\| \leq h) \leq \tilde{\psi}(h)c_2(x),$$

donde además  $\frac{\tilde{\psi}(h)}{\psi^2(h)}$  está acotado.

**H 2.17.** La función  $\psi$  es derivable y cumple, para  $j = 1, 2$  que

$$\lim_{h \rightarrow 0} \frac{h}{\psi(h)} \int_0^1 K^j(u)\psi'(hu)du = C_j,$$

donde  $C_j$  es una constante positiva.

**H 2.18.** La sucesión  $(X_n, Y_n)_{n \in \mathbb{Z}}$  es  $\alpha$ -mixing. El coeficiente de mixing  $\alpha$  cumple que existe  $\nu > 2$  y  $\delta > 1 - 2/\nu$  tales que

$$\sum_{k=1}^{\infty} k^\delta \alpha(k)^{1-2/\nu} < \infty.$$

**H 2.19.** La ventana  $h_n$  cumple que  $\lim_{n \rightarrow \infty} h_n = 0$  y

$$\lim_{n \rightarrow \infty} nh_n^{2\beta} \psi(h_n) = 0.$$

Entonces

$$\sqrt{n\psi(h_n)} (\phi_n(x) - \phi(x)) \xrightarrow{w} N(0, \sigma^2(x)),$$

donde  $\sigma^2(x) = \frac{C_2 g_2(x)}{C_1^2 c_1(x)}$ .

La distribución asintótica en el caso de variables dependientes se obtiene mediante el método de bloques de Bernstein, que permite reducir el problema al teorema central del límite de Lindeberg. Las hipótesis de dependencia débil son análogas a las hipótesis para la consistencia y no se asume estacionariedad en el sentido estricto. Bajo hipótesis más débiles para la convergencia de la ventana se obtiene la distribución gaussiana asintótica para el estimador más un término de sesgo, y para eliminar el término de sesgo se utiliza la hipótesis H 2.19. La hipótesis H 2.17 permite el cálculo de la varianza asintótica.

La demostración en este caso se basa en determinada descomposición del estimador, donde la normalización se hace utilizando los valores esperados, es decir normalizando mediante  $E \left[ K \left( \frac{\|x - X_i\|}{h_n} \right) \right]$ , análogamente a lo realizado en el caso real por Greblicki et al. (1984).

Por otra parte también existen resultados para el caso i.i.d. de Ferraty, Mas y Vieu (2007) sobre distribución asintótica, con hipótesis similares a las de

Masry sobre la concentración de las observaciones. Este trabajo los autores consideran particularmente el cálculo explícito de las constantes que aparecen en sus resultados, y resaltan su importancia para las aplicaciones. Del mismo modo tratan el problema de la elección de la ventana, que se obtiene minimizando el error cuadrático medio para valores del estimador obtenidos mediante bootstrap. El bootstrap para este problema se realiza calculando el estimador  $\phi_n$  a partir de una muestra,  $(X_1, Y_1), \dots, (X_n, Y_n)$  con una ventana  $h_n$  elegida y calculando los residuos para esa muestra, es decir

$$\varepsilon_i = Y_i - \phi_n(X_i),$$

para cada  $i = 1, \dots, n$ . Se obtiene una muestra bootstrap de estos residuos  $\varepsilon_1^*, \dots, \varepsilon_n^*$ , y para un estimador  $\tilde{\phi}_n$  calculando con la muestra original y otra ventana  $h_n$  se calcula

$$Y_i^* = \tilde{\phi}_n(X_i) + \varepsilon_i^*.$$

Luego con la nueva muestra  $(X_1, Y_1^*), \dots, (X_n, Y_n^*)$  y con la ventana original  $h_n$  se obtiene un estimador bootstrap  $\phi_n^*$ . La justificación teórica de la validez del bootstrap para estos datos está indicada como una línea de trabajo a desarrollar por los autores.

Se encuentran también, además de los ya citados, trabajos relacionados sobre la estimación densidades para datos funcionales (Gasser, Hall & Presnell 1998), para la estimación de la moda (Dabo-Niang, Ferraty & Vieu 2004), para la normalidad asintótica de estimadores de núcleos para datos funcionales (Oliveira 2005).

### 2.3. Extensiones al caso no estacionario y dependiente

El problema que motiva este trabajo consiste en estimar ciertos parámetros de calidad de servicio para un tráfico multimedia en una conexión entre dos extremos en la red, donde la calidad de servicio está representada por la variable  $Y$ , es función del estado de la red que determinará la variable  $X$  y se desea estimar la función  $\phi$  en el modelo  $Y = \phi(X) + \varepsilon$ .

En este caso, si bien hay relación entre el estado de la red y la calidad de servicio para un tráfico multimedia, es difícil establecer un modelo para la función  $\phi$ , por eso se enfoca este problema en forma no paramétrica.

También se pueden realizar hipótesis sobre la distribución de las variables  $(X, Y)$  de modo que los datos sigan un modelo paramétrico. Sin embargo para este tipo de datos tampoco parece adecuado un modelo paramétrico. Existen diferentes modelos en la literatura para el tráfico en Internet (modelos con dependencias largas, colas pesadas, Poisson, etc.), pero muchas veces dependen de características particulares de la red, y no parece existir un modelo para el tráfico de Internet que sea suficientemente general y que brinde resultados válidos para situaciones diferentes.

Los datos correspondientes a la variable  $X$  que se obtienen experimentalmente del problema de telecomunicaciones son datos de grandes dimensiones y para el análisis de estos datos es adecuado considerar el vector de datos como una función, es decir que cada observación de la variable aleatoria  $X$  en vez de tratarse como un vector de  $\mathbb{R}^d$  se considerará como función en algún espacio de funciones. Actualmente la posibilidad de obtener y almacenar datos de grandes dimensiones hace que el análisis de éstos sea un tema donde se desarrollan muchos trabajos, y el enfoque de datos funcionales es una alternativa para su tratamiento.

Por otra parte uno de los problemas que aparece para estos datos sobre tráfico en Internet, tanto si se consideran datos funcionales o datos en  $\mathbb{R}^d$ , es que los datos son no estacionarios. El estado de la red presenta un comportamiento no estacionario, dependiendo del momento del día, presentando comportamientos periódicos, donde solamente se puede asumir que las observaciones son estacionarias en escalas de tiempo cortas. A modo de ejemplo existen trabajos donde para el corazón de Internet se puede suponer válida la hipótesis de que el tráfico se distribuye como un proceso de Poisson en escalas cortas de tiempo, mientras que se observa tráfico no estacionario y con dependencias largas en escalas de tiempo mayores, y finalmente un modelo de Poisson dependiente del tiempo permite representar las características de estas diferentes escalas de tiempo (Karagiannis, Molle, Faloutsos & Broido 2004). El carácter estacionario o no de los datos (tanto si provienen de mediciones activas como de mediciones pasivas) depende de la escala de tiempo como del tipo de magnitud que se esté considerando (pérdidas, tráfico por unidad de tiempo) así como también del tipo de red que se esté considerando. Algunos trabajos abordan el análisis de estos diferentes casos con el objetivo de determinar la validez de la predicción de características futuras en base a medidas en determinado momento (Zhang, Paxson & Shenker 2000), (Zhang & Duffield 2001).

En este trabajo se consideran datos funcionales no estacionarios según determinado modelo que se describe en el capítulo 4 para representar el estado de la red y se estudia la regresión no paramétrica para estimar parámetros de calidad de servicio. La extensión de los resultados teóricos presentados en este capítulo a este modelo es el contenido de dicho capítulo. Los teoremas que allí se prueban se basan en los resultados de regresión no paramétrica para datos funcionales que se describieron en este capítulo y en un teorema central de límite para arreglos triangulares de campos estacionarios en  $\mathbb{Z}^d$ . Este teorema y varios conceptos relacionados que se utilizarán en el capítulo 4 son el contenido del capítulo 3.

## Capítulo 3

# Teorema central del límite para arreglos triangulares

En este capítulo se presentan los resultados de distribución asintótica para sumas de variables aleatorias donde la existencia de la distribución límite depende de la geometría de los conjuntos de índices considerados. Estos resultados serán básicos para el capítulo 4, donde se estudiará una mezcla de variables aleatorias no estacionarias y dependientes. Se considerarán en ese caso las sumas en los conjuntos de nivel de una de las componentes de la mezcla, de modo de obtener así variables estacionarias. La geometría de los conjuntos de nivel determinará la existencia de un teorema central del límite. En lo que sigue se presentan las generalidades de los resultados acerca de la relación entre el teorema central del límite y la geometría de los conjuntos, y luego las características que debe tener una variable aleatoria para que sus conjuntos de nivel permitan la existencia de una distribución asintótica gaussiana para sumas de variables con índices en dichos conjuntos.

Se considera una variable  $X = (X_n)_{n \geq 1}$ , donde  $X_n \in \mathbb{R}$ , y para cada subconjunto  $A \subset \mathbb{N}$  se define la suma

$$S_n(A, X) = \frac{1}{\sqrt{n}} \sum_{i \in A_n} X_i,$$

donde  $A_n = A \cap [1, n]$ . Para variables vectoriales, es decir a valores en  $\mathbb{R}^m$ , donde  $X = (X_n)_{n \geq 1}$ , con

$$X_n = (X_n^1, X_n^2, \dots, X_n^m),$$

se considera la suma en  $m$  subconjuntos de  $\mathbb{N}$ ,  $A^1, A^2, \dots, A^m$ , disjuntos dos a dos, de las variables reales  $X^1, X^2, \dots, X^m$  con  $X^k = (X_n^k)_{n \geq 1}$  para  $k = 1, 2, \dots, m$ , donde se considera el vector

$$S_n(A, X) = (S_n(A^1, X^1), S_n(A^2, X^2), \dots, S_n(A^m, X^m)).$$

Perera (1997) (2002) estudia la distribución asintótica de  $S_n(A, X)$  para el caso de campos aleatorios en  $\mathbb{Z}^d$ , es decir  $(X_n)_{n \in \mathbb{Z}^d}$ , concluyendo que la existen-

cia de un teorema central del límite depende fuertemente de la geometría de los conjuntos, donde la varianza para el teorema central del límite se calcula explícitamente en función de determinados parámetros que representan esta geometría. Por otra parte estos resultados son extendidos por Tablar (2007), que presenta el teorema central del límite para arreglos triangulares, es decir cuando se tiene  $X^n = (X_k^n)_{k \in \mathbb{Z}^d}$ , un campo en  $\mathbb{Z}^d$  a valores reales para cada  $n \in \mathbb{N}$ . En este trabajo se enuncian los resultados para variables aleatorias (es decir considerando un conjunto de índices en  $\mathbb{N}$ ), que son los que luego se utilizan en el capítulo 4, y al igual que en  $\mathbb{Z}^d$  la existencia de un teorema central del límite depende fuertemente de la geometría de los conjuntos. Por otra parte los resultados en los trabajos antes mencionados están presentados para campos a valores en  $\mathbb{R}$ , pero también son válidos para campos a valores en  $\mathbb{R}^m$ . Estos resultados se utilizarán en el capítulo 4, donde será necesaria la formulación para variables a valores en  $\mathbb{R}^m$ . La extensión de los resultados de normalidad asintótica para variables a valores en  $\mathbb{R}^m$  consiste en probar la normalidad asintótica para cualquier combinación lineal del vector aleatorio, de manera que el problema se reduce al caso real (Cramer-Wald). En la sección 3.1 se enuncian propiedades de los subconjuntos de  $\mathbb{N}$  que caracterizarán la existencia de un teorema central del límite para  $S_n(A, X)$  o  $S_n(A, X^n)$  y en la sección 3.2 se enuncia el resultado de teorema central del límite para arreglos triangulares para variables a valores reales y su extensión para variable a valores en  $\mathbb{R}^m$ . Un caso particular de conjuntos que será objeto de estudio consiste en considerar los conjuntos de nivel de una variable aleatoria  $Z$ . Determinadas propiedades de la variable determinan las propiedades geométricas de los conjuntos de nivel de modo de garantizar un teorema central del límite para estos conjuntos en particular. En la sección 3.3 se describen estas variables. Esta caracterización permite considerar en el capítulo 4 un modelo de dependencia y no estacionariedad para una mezcla de variables considerando las sumas en los conjuntos de nivel de una de las componente de la mezcla.

### 3.1. Conjuntos asintóticamente medibles

Las siguientes definiciones son introducidas por Perera (1997) (2002), con ejemplos de este tipo de conjuntos, donde se realiza la caracterización de estos en base a la regularidad de su frontera.

**Definición 3.1.** Un subconjunto  $A \subset \mathbb{N}$  se denomina conjunto asintóticamente medible en  $\mathbb{N}$  si para todo  $k \in \mathbb{N}$  existe el siguiente límite

$$F(k, A) = \lim_{n \rightarrow \infty} F_n(k, A) = \lim_{n \rightarrow \infty} \frac{\text{card}\{A_n^c \cap (A_n - k)\}}{n}$$

donde  $A^c$  es el complemento de  $A$ . La función  $F(\cdot, A)$  se llama función frontera del subconjunto  $A$ .

En el siguiente ejemplo se presenta un conjunto que no es asintóticamente medible.

**Ejemplo 3.2.** Sea el conjunto  $A$  definido por

$$A = \cup_{k=0}^{\infty} [10^{2k}, 10^{2k+1}) \cap 2\mathbb{N}.$$

Se tiene que

$$A = \cup_{k=0}^{\infty} I_k \cap 2\mathbb{N}$$

donde  $I_k = [10^{2k}, 10^{2k+1})$  y se cumple que la distancia entre dos intervalos consecutivos  $I_{k-1}$  e  $I_k$  es  $10^{2k-1} \cdot 9$ . Se consideran las sucesiones  $10^{2k}$ ,  $10^{2k+1}$  y  $F_{10^{2k}}(1, A)$ ,  $F_{10^{2k+1}}(1, A)$ . Entonces

$$F_{10^{2k}}(1, A) \leq \frac{\text{card}(A_{10^{2k}})}{10^{2k}}$$

y como  $\text{card}(A_{10^{2k}}) \leq \frac{10^{2k-1}}{2}$ , resulta que

$$F_{10^{2k}}(1, A) \leq \frac{10^{2k-1}}{2 \cdot 10^{2k}} = \frac{1}{20}.$$

Por otra parte  $A_{10^{2k+1}}^c \cap (A_{10^{2k+1}} - 1) = \cup_{i=0}^k [10^{2i}, 10^{2i+1}) \cap 2\mathbb{N} + 1$  de donde

$$F_{10^{2k}}(1, A) \geq \frac{\text{card}(I_k)}{10^{2k+1}}$$

donde  $\text{card}(I_k) = \frac{10^{2k} \cdot 9}{2}$  y entonces

$$F_{10^{2k}}(1, A) \geq \frac{10^{2k} \cdot 9}{2 \cdot 10^{2k+1}} = \frac{9}{20}.$$

Entonces la sucesión  $F_n(1, A)$  no tiene límite y el conjunto  $A$  no es asintóticamente medible.

**Definición 3.3.** La familia de subconjuntos de  $\mathbb{N}$ ,  $\{A^i : i = 1, \dots, m\}$  es una familia asintóticamente medible si para todo  $i \in \{1, \dots, m\}$  el subconjunto  $A^i$  es asintóticamente medible y para todo  $k \in \mathbb{N}$  y para todo  $i, j \in \{1, \dots, m\}$  existe el siguiente límite

$$F(k, A^i, A^j) = \lim_{n \rightarrow \infty} F_n(k, A^i, A^j) = \lim_{n \rightarrow \infty} \frac{\text{card}\{A_n^i \cap (A_n^j - k)\}}{n}.$$

Un ejemplo de familia asintóticamente medible se obtiene considerando los conjuntos de nivel de una variable aleatoria con determinadas propiedades.

**Ejemplo 3.4.** Se consideran  $(Z_n)_{n \in \mathbb{N}}$  una sucesión de variables aleatorias, y  $\{B^j : j = 1, \dots, m\}$  una familia de borelianos, disjuntos dos a dos, y se define el conjunto  $A^j$  mediante  $A^j = \{i \in \mathbb{N} : Z_i \in B^j\}$ . Entonces

$$\begin{aligned} \lim_{n \rightarrow \infty} F_n(k, A^i, A^j) &= \lim_{n \rightarrow \infty} \frac{\text{card}\{A_n^i \cap (A_n^j - k)\}}{n} \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h=1}^n 1_{\{Z_h \in B^i\}} 1_{\{Z_{h+k} \in B^j\}}, \end{aligned}$$

donde la última igualdad se cumple porque  $k$  es fijo y  $n \rightarrow \infty$ . Si por ejemplo  $(Z_n)_{n \in \mathbb{N}}$  es una sucesión i.i.d. por la ley fuerte de los grandes números se tiene que

$$\frac{1}{n} \sum_{h=1}^n 1_{\{Z_h \in B^i\}} 1_{\{Z_{h+k} \in B^j\}} \xrightarrow{c.s.} E(1_{\{Z_0 \in B^i\}} 1_{\{Z_0 \in B^j\}})$$

donde

$$E(1_{\{Z_0 \in B^i\}} 1_{\{Z_0 \in B^j\}}) = P(Z_0 \in B^i) P(Z_0 \in B^j).$$

## 3.2. Teorema central del límite

En esta sección se presenta el teorema central del límite para arreglos triangulares para variables centradas y estacionarias a valores en  $\mathbb{R}$  (Tablar 2007). Un arreglo triangular es una doble sucesión de variables aleatorias  $X_k^n$  donde  $k \rightarrow \infty$  cuando  $n \rightarrow \infty$ . Se enuncia el teorema para variables a valores reales, se presentan las hipótesis para el caso a valores en  $\mathbb{R}^m$  y se justifica la extensión del resultado a este caso. La demostración del teorema para variables a valores en  $\mathbb{R}^m$  consistirá en verificar las hipótesis del caso real para cualquier combinación lineal del vector aleatorio en  $\mathbb{R}^m$ . Para una variable  $X^n = (X_k^n)_{k \in \mathbb{N}}$  centrada estacionaria, a valores en  $\mathbb{R}^m$ , para cada  $n \in \mathbb{N}$ , las componentes se denotan por

$$X^n = (X^{1,n}, \dots, X^{m,n})$$

y para cada  $k \in \mathbb{N}$ ,

$$X_k^n = (X_k^{1,n}, \dots, X_k^{m,n}).$$

Se considera además  $X^{n,J} = (X^{1,n,J}, \dots, X^{m,n,J})$  la variable  $X^n$  truncada por  $J$ , definida para  $k \in \mathbb{N}$  y  $J > 0$  como

$$X_k^{n,J} = X_k^n 1_{\{\|X_k^n\| \leq J\}} - E \left[ X_k^n 1_{\{\|X_k^n\| \leq J\}} \right],$$

donde  $\|\cdot\|$  es la norma euclídea en  $\mathbb{R}^m$ .

### 3.2.1. Teorema central del límite para variables reales

Aquí se presenta el teorema central del límite para arreglos triangulares, (ver teorema 4.15 (Tablar 2007) para la demostración en el caso de campos en  $\mathbb{Z}^d$ ), enunciado para variables con índices en  $\mathbb{N}$  y a valores en  $\mathbb{R}$ , y luego la extensión para variables con índices en  $\mathbb{N}$  y a valores en  $\mathbb{R}^m$ .

**Teorema 3.5.** *Sea  $X^n = (X_k^n)_{k \in \mathbb{N}}$  una variable aleatoria para cada  $n \in \mathbb{N}$ . Si se verifican las siguientes hipótesis:*

**H 3.6.** *Para cada  $n \in \mathbb{N}$ ,  $X^n$  es una variable centrada, estacionaria, con segundo momento finito y tal que  $\sum_{k \in \mathbb{N}} |E\{X_0^n X_k^n\}| < \infty$ .*

**H 3.7.** *Sea  $X^{n,J}$  la truncación por  $J$  de la variable aleatoria  $X^n$  luego:*



1. Existe  $\gamma(k) \geq 0$  tal que  $\sum_{k \in \mathbb{N}} \gamma(k) < \infty$  y para todo  $k \in \mathbb{N}$ ,  $n \in \mathbb{N}$  vale

$$\left| E \left\{ X_0^{n,J} X_k^{n,J} \right\} \right| \leq \gamma(k)$$

2. Existe  $b(J)$  tal que  $\lim_{J \rightarrow \infty} b(J) = 0$  y para cada  $B \subset \mathbb{N}$

$$E \left[ (S_n(B, X^n) - S_n(B, X^{n,J}))^2 \right] \leq \frac{b(J) \text{card}(B_n)}{n}.$$

**H 3.8.** Existe una sucesión  $C(J) < \infty$ , para todo  $J > 0$ , tal que para cada  $B \subset \mathbb{N}$  vale

$$E \left[ S_n(B, X^{n,J})^4 \right] \leq C(J) \left( \frac{\text{card}(B_n)}{n} \right)^2.$$

**H 3.9.** Existe una función  $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , tal que  $\lim_{x \rightarrow +\infty} h(x) = 0$  y una función  $g(J, t) < \infty$  para todo  $J > 0$  fijo,  $g$  está acotada en la segunda variable, es decir  $\sup_{t \in \mathbb{R}^m} g(J, t) = g_J < \infty$ , que verifica

$$\left| E \left[ e^{iS_n(B \cup C, \langle t, X^{n,J} \rangle)} \right] - E \left[ e^{iS_n(B, \langle t, X^{n,J} \rangle)} \right] E \left[ e^{iS_n(C, \langle t, X^{n,J} \rangle)} \right] \right| \leq g(J, t) h(d(B, C)),$$

para todo par de conjuntos disjuntos  $B, C \subset \mathbb{N}$  y para todo  $n \in \mathbb{N}$ ,  $t \in \mathbb{R}$ .

**H 3.10.** Existen sucesiones  $\gamma^J(k)$  y  $\gamma(k)$ ,  $k \in \mathbb{N}$  tal que para todo  $J > 0$  se tiene que

$$\lim_{n \rightarrow \infty} E \left\{ X_0^{n,J} X_k^{n,J} \right\} = \gamma^J(k),$$

y

$$\lim_{J \rightarrow \infty} \gamma^J(k) = \gamma(k).$$

Entonces, para cualquier conjunto  $A$  asintóticamente medible,

$$S_n(A, X^n) \xrightarrow{w} N(0, \sigma^2(A)),$$

donde

$$\sigma^2(A) = \sum_{k \in \mathbb{N}} \gamma(k) F(k, A).$$

La hipótesis H 3.7 permite restringirse a las variables truncadas y de este modo trabajar con un campo centrado, estacionario y acotado. Las hipótesis H 3.8 y H 3.9 son hipótesis generales sobre dependencia débil, donde por ejemplo la hipótesis de  $\alpha$ -mixing y la desigualdad de covarianza implica la hipótesis H 3.9. En los trabajos de Perera (1997) (2002) se presentan discusiones más extensas de condiciones suficientes que implican las hipótesis anteriores para diferentes modelos de dependencia débil, así como también se discuten condiciones equivalentes, por ejemplo en términos de cuantiles. El resultado se obtiene con el método de bloques de Bernstein, donde se prueba que la suma  $S_n(A, X^n)$  tiene

la misma distribución asintótica que una suma de variables independientes. El método de bloques consiste en considerar la suma de las variables en bloques separados de modo que la distancia entre bloques tiende a  $\infty$  cuando  $n \rightarrow \infty$  pero a su vez es despreciable respecto al tamaño de cada bloque. De este modo las sumas en cada bloque asintóticamente se comportan como variables independientes. Luego el resultado final se obtiene aplicando el teorema central del límite para arreglos triangulares en el caso independiente.

### 3.2.2. Teorema central del límite para variables vectoriales

**Teorema 3.11.** *Sea  $X^n = (X_k^n)_{k \in \mathbb{N}}$  una sucesión de variables aleatorias centradas estacionarias a valores en  $\mathbb{R}^m$ . Si se verifican las siguientes hipótesis:*

**H 3.12.** *Para todo  $i, j \in \{1, \dots, m\}$  y  $n \in \mathbb{N}$ ,*

$$\sum_{k \in \mathbb{N}} \left| E \left\{ X_0^{i,n} X_k^{j,n} \right\} \right| < \infty.$$

**H 3.13.** *Existe una sucesión  $\gamma(k) \geq 0$  tal que  $\sum_{k \in \mathbb{N}} \gamma(k) < \infty$  y para todo  $k \in \mathbb{N}$ ,  $n \in \mathbb{N}$ ,  $i, j \in \{1, \dots, m\}$  y  $J > 0$  se tiene*

$$\left| E \left\{ X_0^{i,n,J} X_k^{j,n,J} \right\} \right| \leq \gamma(k).$$

**H 3.14.** *Existe una sucesión  $b(J)$  tal que  $\lim_{J \rightarrow \infty} b(J) = 0$  y para todo  $B \subset \mathbb{N}$ ,  $i \in \{1, \dots, m\}$ ,  $n \in \mathbb{N}$  y  $J > 0$*

$$E \left[ \left( S_n(B, X^{i,n}) - S_n(B, X^{i,n,J}) \right)^2 \right] \leq \frac{b(J) \text{card}(B_n)}{n}.$$

**H 3.15.** *Existe una sucesión  $C(J) < \infty$  para todo  $J > 0$  tal que para todo  $B \subset \mathbb{N}$ ,  $n \in \mathbb{N}$ ,  $J > 0$ ,  $i \in \{1, \dots, m\}$  se tiene que*

$$E \left[ S_n(B, X^{i,n,J})^4 \right] \leq C(J) \left( \frac{\text{card}(B_n)}{n} \right)^2.$$

**H 3.16.** *Existe una función  $h : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ , tal que  $\lim_{x \rightarrow +\infty} h(x) = 0$  y una función  $g : \mathbb{R}^+ \times \mathbb{R}^m \rightarrow \mathbb{R}^+$ , tal que  $g(J, t) < \infty$  para todo  $J > 0$  fijo y se cumple  $\sup_{t \in \mathbb{R}^m} g(J, t) = g_J < \infty$ , tal que*

$$\begin{aligned} \left| E \left[ e^{i S_n(B \cup C, \langle t, X^{n,J} \rangle)} \right] - E \left[ e^{i S_n(B, \langle t, X^{n,J} \rangle)} \right] E \left[ e^{i S_n(C, \langle t, X^{n,J} \rangle)} \right] \right| \\ \leq g(J, t) h(d(B, C)), \end{aligned}$$

para todos los  $B, C \subset \mathbb{N}$  disjuntos, para todo  $n \in \mathbb{N}$ ,  $J > 0$  y  $t \in \mathbb{R}^m$ , donde  $\langle \cdot, \cdot \rangle$  es el producto escalar en  $\mathbb{R}^m$ .

**H 3.17.** *Existen sucesiones  $\gamma^J(i, j, k)$  y  $\gamma(i, j, k)$ ,  $i, j \in \{1, \dots, m\}$ ,  $k \in \mathbb{N}$  tales que para todo  $J > 0$  se tiene que*

$$\lim_{n \rightarrow \infty} E \left\{ X_0^{i,n,J} X_k^{j,n,J} \right\} = \gamma^J(i, j, k),$$

y

$$\lim_{J \rightarrow \infty} \gamma^J(i, j, k) = \gamma(i, j, k)$$

para  $i, j \in \{1, \dots, m\}$  y  $k \in \mathbb{N}$ .

Entonces para cualquier familia asintóticamente medible  $\{A^i : i = 1, \dots, m\}$  de subconjuntos de  $\mathbb{N}$  disjuntos dos a dos ( $A^i \cap A^j = \emptyset$  si  $i \neq j$ ), cuando  $n$  tiende a  $\infty$  se cumple que:

$$(S_n(A^1, X^{1,n}), \dots, S_n(A^m, X^{m,n})) \xrightarrow{w} N_m(0, \Sigma),$$

donde

$$\Sigma(i, j) = \gamma(i, j, 0)F(i, j, 0) + \sum_{k \geq 1} \{\gamma(i, j, k)F(i, j, k) + \gamma(j, i, k)F(j, i, k)\} \quad (3.1)$$

para todo  $i, j \in \{1, \dots, m\}$  y

$$F(i, j, k) = \lim_{n \rightarrow \infty} \frac{\text{card}\{A_n^i \cap (A_n^j - k)\}}{n}.$$

*Demostración.* El resultado se obtiene a partir del teorema 3.5. Para probar que el campo vectorial

$$S_n = (S_n(A^1, X^{1,n}), \dots, S_n(A^m, X^{m,n})) \xrightarrow{w} N_m(0, \Sigma)$$

basta probar que  $\langle \lambda, S_n \rangle \xrightarrow{w} N(0, \lambda^t \Sigma \lambda)$  con  $\lambda^t = (\lambda_1, \dots, \lambda_m)$  ( (Billingsley 1995), pág. 335).

$$\begin{aligned} \langle S_n, \lambda \rangle &= \sum_{k=1}^m \lambda_k S_n(A^k, X^{k,n}) \\ &= \sum_{k=1}^m \frac{1}{\sqrt{n}} \lambda_k \sum_{i \in A_n^k} X_i^{k,n} \\ &= \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \sum_{k=1}^m \lambda_k X_i^{k,n} 1_{\{i \in A_n^k\}} \\ &= \frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \check{X}_i^n \\ &= S_n(\{0, \dots, n-1\}, \check{X}) \end{aligned}$$

donde el proceso  $\check{X}^n = \{\check{X}_i^n : i \in \mathbb{N}\}$  se define por  $\check{X}_i^n = \lambda_k X_i^{k,n}$  si  $i \in A_n^k$ . Para demostrar el teorema basta observar que el proceso a valores reales  $\check{X}^n$  verifica todas las hipótesis del teorema 3.5. Para obtener la matriz de covarianzas  $\Sigma$  basta calcular  $\Sigma(i, j)$ , donde

$$\Sigma(i, j) = \lim_{n \rightarrow \infty} E [S_n(A^i, X^{i,n}) S_n(A^j, X^{j,n})].$$

Se tiene que

$$E(S_n(A^i, X^{i,n})S_n(A^j, X^{j,n})) = \frac{1}{n} \sum_{h \in A_n^i, l \in A_n^j} E(X_h^i X_l^j)$$

Además  $E(X_h^i X_l^j) = \gamma(i, j, l-h)$  si  $l-h \geq 0$  y  $E(X_h^i X_l^j) = \gamma(j, i, h-l)$  si  $l-h < 0$ , de donde  $E[S_n(A^i, X^{i,n})S_n(A^j, X^{j,n})]$  está compuesta por la suma de tres términos

$$\begin{aligned} & \frac{1}{n} \sum_{h \in A_n^i \cap A_n^j} \gamma(i, j, 0), \\ & \frac{1}{n} \sum_{h \in A_n^i} \sum_{l \in A_n^j} \sum_{l-h > 0} \gamma(i, j, l-h), \\ & \frac{1}{n} \sum_{h \in A_n^i} \sum_{l \in A_n^j} \sum_{l-h < 0} \gamma(j, i, h-l). \end{aligned}$$

Entonces calculando los límites respectivos se tiene que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h \in A_n^i \cap A_n^j} \gamma(i, j, 0) = \lim_{n \rightarrow \infty} \gamma(i, j, 0) \frac{\text{card}(A_n^i \cap A_n^j)}{n} = \gamma(i, j, 0)F(i, j, 0),$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h \in A_n^i} \sum_{l \in A_n^j} \sum_{l-h > 0} \gamma(i, j, l-h) &= \lim_{n \rightarrow \infty} \sum_{k \geq 1} \gamma(i, j, k) \frac{\text{card}\{A_n^i \cap (A_n^j - k)\}}{n} \\ &= \sum_{k \geq 1} \gamma(i, j, k)F(i, j, k), \end{aligned}$$

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{h \in A_n^i} \sum_{l \in A_n^j} \sum_{l-h > 0} \gamma(j, i, h-l) &= \lim_{n \rightarrow \infty} \sum_{k \geq 1} \gamma(j, i, k) \frac{\text{card}\{A_n^j \cap (A_n^i - k)\}}{n} \\ &= \sum_{k \geq 1} \gamma(j, i, k)F(j, i, k), \end{aligned}$$

de donde se obtiene  $\Sigma(i, j)$  determinada por la ecuación (3.1).  $\square$

### 3.3. Variables asintóticamente medibles

En este trabajo interesa considerar como conjuntos particulares para la suma de variables aleatorias los conjuntos de nivel de otra variable  $Z$ . En esta sección se presentan las hipótesis que hacen que los conjuntos de nivel de una variable  $Z$  sean familias asintóticamente medibles. Se presentan las definiciones de variables asintóticamente medibles y regulares introducidas por Perera (2002) para campos en  $\mathbb{Z}^d$ .

**Definición 3.18.** La variable  $Z = (Z_n)_{n \in \mathbb{N}}$  a valores reales es una variable asintóticamente medible en  $\mathbb{N}$  si existe una medida de probabilidad aleatoria  $R_0$  en  $\mathcal{B}$ , donde  $\mathcal{B}$  es la  $\sigma$ -álgebra de Borel en  $\mathbb{R}$ , tal que

$$\frac{1}{n} \sum_{m=1}^n 1_{\{Z_m \in B\}} \xrightarrow{c.s.} R_0(B)$$

y para todo  $i \in \mathbb{N} - \{0\}$  existe una medida aleatoria  $R_k$  en  $\mathcal{B}_2$ , donde  $\mathcal{B}_2$  es la  $\sigma$ -álgebra de Borel en  $\mathbb{R}^2$ , tal que para todo  $B, C \in \mathcal{B}$

$$\frac{1}{n} \sum_{m=1}^n 1_{\{Z_m \in B\}} 1_{\{Z_{m-k} \in C\}} \xrightarrow{c.s.} R_k(B \times C)$$

Si la medida límite no es aleatoria se dice que  $Z$  es un campo regular en  $\mathbb{N}$ .

En la siguiente proposición se calcula la función frontera para los conjuntos de nivel de una variable  $Z$ , siendo el ejemplo 3.4 un caso particular de esto.

**Proposición 3.19.** *Se consideran  $(Z_n)_{n \in \mathbb{N}}$  una sucesión de variables aleatorias, y  $\{B^j : j = 1, \dots, m\}$  una familia de borelianos, disjuntos dos a dos, y se define el conjunto  $A^j$  mediante  $A^j = \{i \in \mathbb{N} : Z_i \in B^j\}$  para cada  $j = 1, \dots, m$ . Entonces condicionado a  $Z$  la familia  $\{A^j : j = 1, \dots, m\}$  es una familia asintóticamente medible con*

$$\begin{aligned} F(k, A^i, A^j) &= R_k(B^i \times B^j), \\ F(0, A^i, A^j) &= R_0(B^i) 1_{\{i=j\}}. \end{aligned}$$

Si además  $Z$  es regular  $F(k, A^i, A^j)$  y  $F(0, A^i, A^j)$  no dependen de  $Z$  y la familia  $\{A^j : j = 1, \dots, m\}$  es una familia asintóticamente medible.

*Demostración.* Se tiene que

$$F_n(k, A^i, A^j) = \frac{\text{card}(A_n^i \cap (A_n^j - k))}{n},$$

de donde  $l \in A_n^i \cap (A_n^j - k)$  si  $Z_l \in B^i$  y  $Z_{l+k} \in B^j$ , de modo que asintóticamente, como  $k$  es fijo  $F_n(k, A^i, A^j)$  se comporta como  $\frac{1}{n} \sum_{h=0}^{n-1} 1_{\{Z_h \in B^i\}} 1_{\{Z_{h+k} \in B^j\}}$ , de donde, condicionando a  $Z$ , que es asintóticamente medible, resulta

$$F(k, A^i, A^j) = R_k(B^i \times B^j).$$

Del mismo modo

$$F_n(0, A^i, A^j) = \frac{\text{card}(A_n^i \cap A_n^j)}{n}$$

y  $l \in A_n^i \cap A_n^j$  si  $Z_l \in B^i$ ,  $Z_l \in B^j$ , entonces  $F_n(0, A^i, A^j) = 0$  si  $i \neq j$  y para  $i = j$  tomando límite condicionado a  $Z$  se obtiene

$$F(0, A^i, A^j) = R_0(B^i).$$

Cuando  $Z$  es regular las medidas  $R_k$  y  $R_0$  no dependen de  $Z$  y se obtiene el mismo resultado sin condicionar.  $\square$

## Capítulo 4

# Consistencia y distribución asintótica del estimador

En este capítulo se estudia la consistencia y la distribución asintótica del estimador de la función de regresión  $\phi$  en el modelo

$$Y = \phi(X) + \varepsilon,$$

donde  $X \in \mathcal{D}$ , un espacio vectorial de funciones con una seminorma  $\|\cdot\|$ ,  $\phi$  es una función  $\phi : \mathcal{D} \rightarrow \mathbb{R}$ ,  $Y \in \mathbb{R}$ , y la variable  $\varepsilon$  es centrada e independiente de  $X$ .

Se quiere estudiar la consistencia y distribución asintótica estacionaria del estimador obtenido en base a las observaciones  $\{(X_i, Y_i) : i = 1, \dots, n\}$ , definido por

$$\phi_n(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{\|x-X_i\|}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\|x-X_i\|}{h_n}\right)}.$$

El objetivo es generalizar los resultados sobre consistencia del estimador demostrados por Ferraty et al. (2002) y sobre distribución asintótica demostrados por Masry (2005) para un modelo en el que  $X$  es una mezcla no estacionaria y dependiente de procesos estacionarios débilmente dependientes. Más precisamente se asume que existen dos procesos  $\xi = (\xi_n)_{n \geq 1}$ ,  $Z = (Z_n)_{n \geq 1}$  independientes entre sí, tales que  $\xi$  es estacionario con valores en  $\mathcal{F}$ , siendo  $\mathcal{F}$  un espacio de funciones y  $Z$  toma valores en un subconjunto  $\{z_1, \dots, z_m\}$  de  $\mathbb{R}$  y una función  $\varphi : \mathcal{F} \times \mathbb{R} \rightarrow \mathcal{D}$  tal que  $X$  cumple que para cada  $n \geq 1$

$$X_n = \varphi(\xi_n, Z_n).$$

Se asumen condiciones generales de dependencia que se obtienen de condiciones de dependencia débil del proceso  $\xi$ , por ejemplo hipótesis de mixing, pero el proceso  $Z$  puede ser no estacionario y dependiente. La consistencia del estimador se obtiene mediante argumentos similares a los de Ferraty et al.(2002),

condicionando a la variable  $Z$  y luego hipótesis generales para esta variable permiten obtener el resultado sin condicionar. Para obtener la normalidad asintótica también se considerará el estimador condicionado a los valores de la variable  $Z$ . Este modelo permite considerar observaciones no estacionarias e incluso no débilmente dependientes. Perera (2002) introduce el modelo para aplicaciones a problemas con datos estacionales. Se tienen diferentes tipos de variables para cada  $k = \{1, 2, \dots, m\}$ . En la aplicación a telecomunicaciones estos tipos de variables pretenden modelar las variaciones a lo largo del tiempo de los estados de una red, donde se puede asumir que por ejemplo hay determinadas horas del día donde el tráfico que atraviesa un enlace es mucho mayor que en otras horas, pero que dentro de ciertas escalas de tiempo este mismo tráfico es estacionario. Para cada uno de los valores de  $Z$  se tiene una variable estacionaria, de modo que el objetivo será aplicar los resultados de teorema central del límite para arreglos triangulares del capítulo 3 para un vector que tiene en sus componentes a los valores del estimador condicionados a cada valor de  $Z$ . Por otra parte para resolver el problema de dimensión infinita al trabajar con variables funcionales se utilizan las mismas técnicas que Masry (2005). Para cada valor del conjunto  $\{z_1, \dots, z_m\}$  se asumen hipótesis sobre la concentración de la variable  $\varphi(\xi, z_k)$ , alrededor de  $x$ , donde a grandes rasgos se asume que existen funciones no negativas  $\psi_k$  definidas en  $\mathbb{R}^+$ , con  $\psi_k(0) = 0$ , continuas en un entorno de 0, y  $c_k$  también no negativa definida en  $\mathcal{D}$  tales que para todo  $h > 0$ , cuando  $h \rightarrow 0$

$$P(\|\varphi(\xi, z_k) - x\| \leq h) = c(x)\psi_k(h).$$

Las funciones  $\psi_k$  determinarán la normalización del estimador, dependiendo de las características de cada componente de la mezcla.

El estimador  $\phi_n$  se puede escribir como

$$\phi_n(x) = \frac{g_n(x)}{f_n(x)},$$

donde

$$g_n(x) = \frac{1}{n\psi(h_n)} \sum_{i=1}^n Y_i K_n(X_i),$$

$$f_n(x) = \frac{1}{n\psi(h_n)} \sum_{i=1}^n K_n(X_i)$$

con

$$K_n(u) = K\left(\frac{\|u - x\|}{h_n}\right),$$

para todo  $u \in \mathcal{D}$  y donde  $\psi(h_n)$  es la normalización, y está definida más adelante, pero se obtiene a partir de las funciones  $\psi_k$  que determinan la concentración de cada componente de la mezcla, es decir la concentración de cada variable  $\varphi(\xi, z_k)$ , con  $k = 1, \dots, m$ .

En el teorema 4.9 se prueba la consistencia del estimador y en el teorema 4.17 se obtiene la normalidad asintótica de

$$\phi_n(x) - \frac{E(g_n(x))}{E(f_n(x))}.$$

En lo que sigue se presentan los resultados de consistencia en la sección 4.1 y de normalidad asintótica en la sección 4.2.

## 4.1. Consistencia

En esta sección se demuestra a la convergencia casi segura del estimador. Para esto se calcula primero el límite de  $E(f_n(x))$  y de  $E(g_n(x))$ , siguiendo las ideas de Ferraty et al.(2002), condicionando a los valores de la variable  $Z$ . Luego de la demostración se presentan algunos comentarios sobre las hipótesis.

**Proposición 4.1.** *Sea  $X = (X_n)_{n \geq 1}$  y el modelo  $Y = \phi(X) + \varepsilon$ , con  $\varepsilon$  centrada e independiente de  $X$  tal que se cumplen las siguientes hipótesis.*

**H 4.2.** *Existen dos procesos  $\xi = (\xi_n)_{n \in \mathbb{N}}$ ,  $Z = (Z_n)_{n \in \mathbb{N}}$ , independientes entre sí, tales que  $\xi$  es estacionario con valores en  $\mathcal{D}$ ,  $Z$  toma valores en un subconjunto  $\{z_1, \dots, z_m\}$  de  $\mathbb{R}$  y existe una función  $\varphi : \mathcal{D} \times \mathbb{R} \rightarrow \mathcal{D}$  tal que*

$$X_n = \varphi(\xi_n, Z_n).$$

**H 4.3.** *Existen funciones positivas  $\psi, \psi_1, \dots, \psi_m$  definidas en  $\mathbb{R}^+ \times \mathcal{D}$ , funciones  $c_1, \dots, c_k$  definidas en  $\mathcal{D}$  y un subconjunto  $\Delta$  de  $\{1, \dots, m\}$  tal que para todo  $h > 0$*

$$P[\|\varphi(\xi_1, z_k) - x\| \leq h] = c_k(x)\psi_k(h, x),$$

donde  $\lim_{h \rightarrow 0} \frac{\psi_k(h, x)}{\psi(h, x)} = 1$  si  $k \in \Delta$ , y  $\lim_{h \rightarrow 0} \frac{\psi_k(h, x)}{\psi(h, x)} = 0$  si  $k \in \Delta^c$ . Para simplificar notación se omite la dependencia de  $x$  escribiendo  $\psi(h_n)$  en vez de  $\psi(h_n, x)$ .

**H 4.4.** *Las funciones  $u \mapsto \psi_k(u, x)$  son derivables en  $\mathbb{R}^+$ , con derivada  $\psi'_k(u, x)$  y cumplen*

$$\lim_{h \rightarrow 0} \frac{h}{\psi_k(h, x)} \int_0^1 K(u)\psi'_k(uh, x)du = d_k(x),$$

donde las funciones  $d_k$  están definidas en  $\mathcal{D}$ .

**H 4.5.** *Para  $k \in \{1, \dots, m\}$  existe el siguiente límite*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k)$$

y se denota  $p_k$ .

**H 4.6.** *La función  $\phi$  es continua.*

**H 4.7.** *La función  $K$  es positiva con soporte en  $[0, 1]$ .*

**H 4.8.** *La ventana  $h_n$  cumple que  $\lim_{n \rightarrow \infty} h_n = 0$  y que  $\lim_{n \rightarrow \infty} n\psi(h_n) = 0$ .*

Entonces:



1. el estimador  $f_n(x)$  cumple que

$$\lim_{n \rightarrow \infty} E(f_n(x)) = f(x),$$

donde  $f(x) > 0$  y la función  $f$  está definida para todo  $u \in \mathcal{D}$  por

$$f(u) = \sum_{k \in \Delta} p_k d_k(u) c_k(u),$$

2. el estimador  $g_n(x)$  cumple que

$$\lim_{n \rightarrow \infty} E(g_n(x)) = \phi(x) f(x).$$

*Demostración.*

$$E(f_n(x)) = \frac{1}{n\psi(h_n)} \sum_{i=1}^n E(K_n(X_i))$$

Como  $\xi$  y  $Z$  son independientes se cumple

$$E(K_n(X_i)) = E\{E(K_n(X_i)|Z_i)\} = \sum_{k=1}^m E\{K_n(\varphi(\xi_i, z_k))\} P(Z_i = z_k)$$

Como para cada  $k$   $(\varphi(\xi_i, z_k))_{i \geq 1}$  es una sucesión estacionaria se tiene que

$$E(f_n(x)) = \sum_{k=1}^m \left( \frac{1}{\psi(h_n)} E(K_n(\varphi(\xi_1, z_k))) \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \right).$$

La hipótesis H 4.3 implica que la densidad de  $\|\varphi(\xi_1, z_k) - x\|$  es la función  $u \mapsto c_k(x) \psi'_k(u, x)$  y entonces

$$E[K_n(\varphi(\xi_1, z_k))] = h_n c_k(x) \int_0^1 K(u) \psi'_k(uh_n, x) du$$

Luego usando las hipótesis H 4.3, H 4.4, H 4.8 se tiene que

$$\lim_{n \rightarrow \infty} \frac{1}{\psi(h_n)} E[K_n(\varphi(\xi_1, z_k))] = d_k(x) c_k(x) 1_{\{k \in \Delta\}},$$

y usando la hipótesis H 4.5

$$\begin{aligned} \lim_{n \rightarrow \infty} E(f_n(x)) &= \lim_{n \rightarrow \infty} \sum_{k=1}^m \left( \frac{1}{\psi(h_n)} E(K_n(\varphi(\xi_1, z_k))) \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \right) \\ &= \sum_{k \in \Delta} p_k d_k(x) c_k(x) \end{aligned}$$

Análogamente se tiene que

$$E(g_n(x)) = \sum_{k=1}^m \left( \frac{1}{\psi(h_n)} E(\phi(\varphi(\xi_1, z_k)) K_n(\varphi(\xi_1, z_k))) \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \right).$$

Entonces

$$E(g_n(x)) = \phi(x)E(f_n(x)) + R_n,$$

donde

$$\begin{aligned} R_n &= \sum_{k=1}^m \left( \frac{1}{\psi(h_n)} E[\{\phi(\varphi(\xi_1, z_k)) - \phi(x)\} K_n(\varphi(\xi_1, z_k))] \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \right) \\ &\leq \sup_{u: \|x-u\| \leq h_n} |\phi(u) - \phi(x)| E(f_n(x)) \end{aligned}$$

y el resultado  $\lim_{n \rightarrow \infty} E(g_n(x)) = \phi(x)f(x)$  se obtiene a partir de la continuidad de  $\phi$  y de  $\lim_{n \rightarrow \infty} E(f_n(x)) = f(x)$ .  $\square$

A continuación se hacen algunos comentarios sobre las hipótesis y sobre el resultado de la proposición, estableciendo las analogías que existen con el caso real.

La hipótesis H 4.3 es sobre la concentración de las observaciones en un entorno de  $x$ . Para cada valor de  $k$ ,  $\psi_k(h)$  representa la relación de la concentración de la variable con  $h$ , por ejemplo para una variable en  $\mathbb{R}^d$  con densidad, para  $h$  en un entorno de 0 se tiene que la distribución de la variable es proporcional a  $h^d$  por el teorema del valor medio. La función  $\psi$  que aparece en la normalización para el cálculo del estimador determina cuáles de los componentes de la mezcla intervienen efectivamente en el estimador. Los únicos componentes de la mezcla que finalmente determinan el comportamiento del estimador son aquellos  $\varphi(\xi_n, z_k)$  con  $k \in \Delta$ . Localmente, es decir alrededor de  $x$ , el comportamiento de las variables  $\varphi(\xi_n, z_k)$  es similar al de un espacio de dimensión finita donde, para  $h$  tendiendo a 0,  $\psi(h)$  es análogo al volumen de la bola de centro  $x$  y radio  $h$ . Las componentes de la mezcla que determinan el comportamiento asintótico son las que corresponden a la menor dimensión del espacio en estas aproximaciones.

La hipótesis H 4.4 implica que la variable  $\|\varphi(\xi, z_k) - x\|$  tiene densidad  $c_k(x)\psi'_k$ . En el caso de variables en  $\mathbb{R}^d$  con densidad las funciones  $d_k$  son constantes para todo  $x \in \mathbb{R}^d$ .

La hipótesis H 4.5 Esto garantiza cierta estacionariedad “en media”. Esta hipótesis se verifica por ejemplo para variables periódicas, y más generalmente se verifica para las variables asintóticamente medibles presentadas en el capítulo 4. Si  $Z$  es asintóticamente medible se cumple que existe una medida aleatoria  $R_0$  definida en los borelianos que cumple que

$$\frac{1}{n} \sum_{i=1}^n 1_{\{Z_i = z_k\}} \xrightarrow{c.s.} R_0(B^i),$$

siendo  $B^i = \{1 \leq i \leq n : Z_i = z_k\}$ . y por el teorema de convergencia monótona se tiene que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E(1_{\{Z_i = z_k\}}) = E(R_0(B^i)).$$

El siguiente ejemplo, similar al ejemplo 3.2 de un conjunto que no es asintóticamente medible, es un ejemplo de una variable  $Z$  que no cumple la hipótesis H 4.5.

$$Z_i = \begin{cases} 1 & \text{si } i \in \cup_{k=0}^{\infty} [10^{2k}, 10^{2k+1}) \cap \mathbb{N}. \\ 0 & \text{si no.} \end{cases}$$

Se consideran las sucesiones  $10^{2k}$  y  $10^{2k+1}$  con  $k \in \mathbb{N}$ . Entonces

$$\lim_{k \rightarrow \infty} \frac{1}{10^{2k}} \sum_{i=1}^{10^{2k}} P(Z_i = z_j) \leq \frac{10^{2k-1}}{10^{2k}} = \frac{1}{10}$$

y por otra parte

$$\lim_{k \rightarrow \infty} \frac{1}{10^{2k+1}} \sum_{i=1}^{10^{2k+1}} P(Z_i = z_j) \geq \sum_{i=10^{2k}+1}^{10^{2k+1}} P(Z_i = z_j) = \frac{10^{2k} \cdot 9}{10^{2k+1}} = \frac{9}{10},$$

de donde  $\frac{1}{n} \sum_{i=1}^n P(Z_i = z_k)$  no tiene límite.

Las hipótesis H 4.6, H 4.7, H 4.8 son usuales en estimaciones para regresión basadas en núcleos.

Si se considera la proposición anterior en el caso de variables i.i.d. con densidad en  $\mathbb{R}^d$  la función  $f$  es la densidad de  $X$ . En este caso aparece la mezcla de las funciones que indican la concentración de cada componente, donde se consideran únicamente las variables  $\varphi(\xi, z_k)$  con  $k \in \Delta$ . Por ejemplo, si  $Z$  toma dos valores  $z_1$  y  $z_2$  y se cumple que  $\psi_1(h) = h^{d_1}$  y  $\psi_2(h) = h^{d_2}$  con  $d_1 < d_2$ , considerando  $\psi(h) = h^{d_1}$  se tiene que  $f(u) = p_1 d_1(u) c_1(u)$ . Se puede interpretar que en la función  $f$  únicamente intervienen las componentes de la mezcla que localmente corresponden a un espacio de menor dimensión.

Para demostrar la consistencia se considera el estimador condicionado a cada uno de los valores de la variable  $Z$ , y se trabaja con una variable vectorial a valores en  $\mathbb{R}^{2m}$  como sigue.

Se considera la variable  $\tilde{X}^n = (\tilde{X}^{1,n}, \dots, \tilde{X}^{2m,n})$  a valores en  $\mathbb{R}^{2m}$  definida para  $i \geq 1$  de la siguiente manera.

- para  $l \in \{1, \dots, m\}$

$$\begin{aligned} \tilde{X}_i^{l,n} &= \frac{1}{\sqrt{\psi(h_n)}} K_n(\varphi(\xi_i, z_l)) (\phi(\varphi(\xi_i, z_l)) + \varepsilon_i) \\ &\quad - \frac{1}{\sqrt{\psi(h_n)}} E [K_n(\varphi(\xi_i, z_l)) \phi(\varphi(\xi_i, z_l))], \end{aligned}$$

- para  $l \in \{m+1, \dots, 2m\}$

$$\tilde{X}_i^{l,n} = \frac{1}{\sqrt{\psi(h_n)}} K_n(\varphi(\xi_i, z_l)) - \frac{1}{\sqrt{\psi(h_n)}} E [K_n(\varphi(\xi_i, z_l))]$$

Para probar la consistencia se introducen hipótesis adicionales que se presentan en el siguiente teorema.

**Teorema 4.9.** *Se asumen las hipótesis de la proposición 4.1 y las siguientes hipótesis adicionales*

**H 4.10.** *Para todo  $B \subset \mathbb{N}$ ,  $n \in \mathbb{N}$ ,  $i \in \{1, \dots, 2m\}$  se tiene*

$$E \left[ S_n \left( B, \tilde{X}^{i,n} \right)^4 \right] \leq C \left( \frac{\text{card}(B_n)}{n} \right)^2$$

**H 4.11.** *La ventana  $h_n$  cumple que*

$$\lim_{n \rightarrow \infty} h_n = 0$$

y para algún  $0 < \beta < 1$

$$\lim_{n \rightarrow \infty} \psi(h_n) n^{\beta/2} = 0.$$

Entonces el estimador  $\phi_n(x)$  converge casi seguramente a  $\phi(x)$  cuando  $n \rightarrow \infty$ .

*Demostración.* Como  $\phi_n(x) = \frac{g_n(x)}{f_n(x)}$  basta probar que  $f_n(x)$  converge casi seguramente a  $f(x)$  y que  $g_n(x)$  converge casi seguramente a  $\phi(x)f(x)$ . Se tiene que

$$\begin{aligned} f_n(x) - f(x) &= f_n(x) - E(f_n(x)) + E(f_n(x)) - f(x) \\ g_n(x) - \phi(x)f(x) &= g_n(x) - E(g_n(x)) + E(g_n(x)) - \phi(x)f(x) \end{aligned}$$

y de la proposición 4.1

$$\lim_{n \rightarrow \infty} E(f_n(x)) - f(x) = 0,$$

$$\lim_{n \rightarrow \infty} E(g_n(x)) - \phi(x)f(x) = 0,$$

entonces basta probar que  $f_n(x) - E(f_n(x))$  y  $g_n(x) - E(g_n(x))$  convergen a cero casi seguramente. Para probar que  $f_n - E(f_n(x))$  converge a cero casi seguramente basta probar la convergencia completa, es decir que para todo  $\varepsilon > 0$  la serie  $\sum P(|f_n(x) - E(f_n(x))| > \varepsilon)$  es convergente.

$$\begin{aligned} f_n(x) - E(f_n(x)) &= \frac{1}{n\psi(h_n)} \sum_{i=1}^n [K_n(X_i) - E(K_n(X_i))] \\ &= \frac{1}{\sqrt{n\psi(h_n)}} S_n \end{aligned}$$

donde

$$S_n = \frac{1}{\sqrt{n\psi(h_n)}} \sum_{i=1}^n [K_n(X_i) - E(K_n(X_i))]$$

Usando la desigualdad de Markov se tiene que

$$P(|f_n(x) - E(f_n(x))| > \varepsilon) \leq \frac{E[f_n(x) - E(f_n(x))]^4}{\varepsilon^4} = \frac{E(S_n)^4}{(n\psi(h_n))^2 \varepsilon^4}$$

Para calcular  $E(S_n^4)$  condicionamos a toda la trayectoria del proceso  $Z$  donde una trayectoria del proceso  $Z$  es la variable  $Z^\infty$  que cumple que  $Z^\infty = z^\infty$  si  $Z_i = z_{j_i}$  para todo  $i \geq 1$ . Es decir que una trayectoria es una variable aleatoria que toma valores en el espacio de sucesiones a valores en  $\{z_1, \dots, z_m\}$ .

$$\begin{aligned} E(S_n^4) &= E[E(S_n^4|Z^\infty)] \\ &= \int_T E(S_n^4|Z^\infty = z^\infty) d\mu \end{aligned}$$

siendo  $T$  el espacio de trayectorias, es decir que  $T$  es el conjunto de las sucesiones toman valores en  $\{z_1, z_2, \dots, z_m\}$ . Para probar la convergencia completa basta probar que para un conjunto de trayectorias con probabilidad 1 se cumple que  $E(S_n|Z^\infty = z^\infty) \leq a_n$  donde  $\sum \frac{a_n}{(n\psi(h_n))^2}$  converge. Calculando para cada trayectoria

$$\begin{aligned} E(S_n^4|Z^\infty = z^\infty) &= E \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\{K_n(\varphi(\xi_i, z_{l_i})) - E[K_n(\varphi(\xi_i, z_{l_i}))]\}}{\sqrt{\psi(h_n)}} \right]^4 \\ &= E \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i^{l_i, n} \right]^4 \end{aligned}$$

con  $l_i \in \{m+1, \dots, 2m\}$ . Agrupando los sumandos según los valores de  $Z$  se tiene

$$\sum_{i=1}^n \tilde{X}_i^{l_i, n} = \sum_{k=m+1}^{2m} \sum_{i \in A_n^k} \tilde{X}_i^{k, n}$$

siendo  $A_n^k = \{i : Z_i = z_k\} \cap [1, n]$  y entonces

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{X}_i^{l_i, n} &= \sum_{k=m+1}^{2m} \frac{1}{\sqrt{n}} \sum_{i \in A_n^k} \tilde{X}_i^{k, n} \\ &= \sum_{k=m+1}^{2m} S_n(A^k, \tilde{X}^{k, n}) \end{aligned}$$

Se tiene que, aplicando la desigualdad de Cauchy-Schwartz y la hipótesis H 4.10,

$$\begin{aligned} E \left[ \sum_{k=m+1}^{2m} S_n(A^k, \tilde{X}^{k, n}) \right]^4 &\leq m \max_{k \in \{m+1, \dots, 2m\}} E \left[ S_n(A^k, \tilde{X}^{k, n})^4 \right] \\ &\leq C \max_{k \in \{m+1, \dots, 2m\}} \left( \frac{\text{card}(A_n^k)}{n} \right)^2 \end{aligned}$$

Entonces, como  $\text{card}(A_n^k) \leq n$ , se tiene que  $E(S_n^4|Z^\infty = z^\infty) \leq C$ . La demostración anterior, donde se probó que  $E(S_n^4|Z^\infty = z^\infty) \leq C$  es independiente de la trayectoria. Para la convergencia completa es necesario elegir  $h_n$  de

modo que la serie

$$\sum \frac{1}{(n\psi(h_n))^2}$$

sea convergente, por ejemplo  $(n\psi(h_n))^2 \geq n^\alpha$  con  $\alpha > 1$ , es decir  $\psi(h_n) \geq \frac{1}{n^{\beta/2}}$ , con  $0 < \beta < 1$  y usando la hipótesis H 4.11 se tiene la convergencia completa de  $f_n(x) - E[f_n(x)]$  a cero. La convergencia completa de  $g_n(x) - E[g_n(x)]$  a cero se prueba de la misma manera, considerando las variables  $\tilde{X}^{k,n}$  con  $k \in \{1, \dots, m\}$ .  $\square$

En el teorema anterior la hipótesis H 4.10 es análoga a la hipótesis H 3.15 en el capítulo 3, mientras que la hipótesis H 4.11 brinda una velocidad adecuada de la ventana para obtener la consistencia del estimador.

## 4.2. Distribución asintótica

La distribución asintótica del estimador se demuestra en varios pasos, donde primero se halla la distribución asintótica de

$$\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x)) \right)$$

Para esto se considera la variable  $\tilde{X}^n$  a valores en  $\mathbb{R}^{2m}$  (donde cada componente corresponde a un valor fijo de la variable  $Z$ ) y se usa el teorema 3.11. Luego se demuestra la distribución gaussiana asintótica para

$$\sqrt{n\psi(h_n)} \left( \phi_n(x) - \frac{E(g_n(x))}{E(f_n(x))} \right)$$

y finalmente para

$$\sqrt{n\psi(h_n)} (\phi_n(x) - \phi(x)).$$

**Proposición 4.12.** Para  $l \in \{1, \dots, m\}$ , sea  $A^l = \{k \in \mathbb{N}, Z_k = z_l\}$  y  $A^{l+m} = A^l$ . Se asumen las siguientes hipótesis

**H 4.13.**  $\tilde{X}^n = (\tilde{X}^{1,n}, \dots, \tilde{X}^{2m,n})$  cumple las hipótesis del teorema 3.11 del capítulo 3,

**H 4.14.**  $Z$  es asintóticamente medible.

Entonces, condicionado a  $Z$ , cuando  $n$  tiende a  $\infty$

$$(S_n(A^1, \tilde{X}^{1,n}), \dots, S_n(A^{2m}, \tilde{X}^{2m,n})) \xrightarrow{w} N_{2m}(0, \Sigma),$$

donde para  $i, j \in \{1, \dots, 2m\}$

$$\Sigma(i, j) = \gamma(i, j, 0)F(i, j, 0) + \sum_{k \in \mathbb{N}} \{\gamma(i, j, k)F(i, j, k) + \gamma(j, i, k)F(j, i, k)\},$$

con  $\gamma(i, j, k)$ ,  $k \in \mathbb{N}$ , definida como en la hipótesis H 3.17 del capítulo 3 para el proceso  $\tilde{X}^n$  y

$$F(i, j, k) = \lim_{n \rightarrow \infty} \frac{\text{card}\{A_n^i \cap (A_n^j - k)\}}{n}.$$

*Demostración.* La demostración de este teorema consiste en aplicar la proposición 3.19 a  $Z$  y el teorema 3.11 a  $\tilde{X}^n$  y a los conjuntos de nivel de la variable  $Z$ .  $\square$

Las hipótesis de dependencia débil se obtiene a partir de hipótesis de dependencia débil para  $\xi$ , por ejemplo mixing, pues, para cada  $k$ , las  $\sigma$ -álgebras generadas por la sucesiones  $\varphi(\xi_i, z_k)$ ,  $\tilde{X}_i^{k,n}$  y  $\tilde{X}_i^{k+m,n}$  coinciden. La hipótesis H 4.14 permite condicionar a los conjuntos de nivel de la variable  $Z$ .

Para obtener la distribución asintótica de

$$\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x)) \right)$$

se asumen hipótesis sobre la variable  $Z$  de modo que el resultado anterior sea válido sin condicionar y luego se obtiene la distribución utilizando las propiedades de las combinaciones lineales de vectores gaussianos.

**Proposición 4.15.** *Se asumen las hipótesis de la proposiciones 4.1 y 4.12 y la siguiente hipótesis*

**H 4.16.** *El proceso  $Z$  es regular y cumple que para cada  $k \in \{1, \dots, m\}$  que*

$$\sqrt{n\psi(h_n)} \left( \frac{1}{n} \sum_{i=1}^n 1_{\{Z_i=z_k\}} - \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \right) \xrightarrow{p} 0. \quad (4.1)$$

*Entonces, cuando  $n$  tiende a  $\infty$*

$$\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x)) \right) \xrightarrow{w} N_2(0, A),$$

*donde*

$$A = \begin{pmatrix} a_1(x) & a_2(x) \\ a_2(x) & a_3(x) \end{pmatrix}$$

*con*

$$a_1(x) = \sum_{i,j=1}^m \Sigma(i, j), \quad a_2(x) = \sum_{i=1}^m \sum_{j=m+1}^{2m} \Sigma(i, j), \quad a_3(x) = \sum_{i,j=m+1}^{2m} \Sigma(i, j).$$

*Demostración.* Condicionado a  $Z$

$$\begin{aligned} & \sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)|Z), f_n(x) - E(f_n(x)|Z) \right) \\ &= \left( \sum_{j=1}^m S_n(A^j, \tilde{X}_{n,j}), \sum_{j=m+1}^{2m} S_n(A^j, \tilde{X}_{n,j}) \right), \end{aligned}$$

es decir que se obtiene a partir del vector

$$S_n = (S_n(A^1, \tilde{X}^{1,n}), \dots, S_n(A^{2m}, \tilde{X}^{2m,n}))$$

mediante una función lineal dada por

$$\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)|Z), f_n(x) - E(f_n(x)|Z) \right)^t = MS_n^t,$$

siendo

$$M = \begin{pmatrix} 1 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 1 & \cdots & 1 \end{pmatrix}.$$

Aplicando la proposición 4.12

$$\left( \sum_{j=1}^m S_n(A^j, \tilde{X}_{n,j}), \sum_{j=m+1}^{2m} S_n(A^j, \tilde{X}_{n,j}) \right) \xrightarrow{w} N_2(0, A),$$

donde la distribución límite se obtiene condicionada a  $Z$  y la matriz de covarianzas se obtiene mediante  $M\Sigma M^t$ , donde estos resultados son un caso particular para funciones lineales de las propiedades de la convergencia en distribución al aplicar funciones continuas. Para  $Z$  regular, la matriz de covarianzas  $A$  es no aleatoria y se tiene

$$\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)|Z), f_n(x) - E(f_n(x)|Z) \right) \xrightarrow{w} N_2(0, A).$$

Entonces

$$\begin{aligned} (g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x))) &= (g_n(x) - E(g_n(x)|Z), f_n(x) - E(f_n(x)|Z)) \\ &\quad + (E(g_n(x)|Z) - E(g_n(x)), E(f_n(x)|Z) - E(f_n(x))). \end{aligned}$$

Se tiene que

$$E(f_n(x)|Z) - E(f_n(x)) = \sum_{j=1}^{2m} \frac{E[K_n(\varphi(\xi_1, z_j))]}{\psi(h_n)} \left( \frac{\text{card}(A_n^j)}{n} - \frac{1}{n} \sum_{i=1}^n P[Z_i = j] \right).$$

Usando la hipótesis H 4.16, y la proposición 4.1, donde se obtiene

$$\lim_{n \rightarrow \infty} \frac{1}{\psi(h_n)} E[K_n(\varphi(\xi_1, z_k))] = d_k(x) c_k(x) 1_{\{k \in \Delta\}},$$

resulta que

$$\sqrt{n\psi(h_n)} (E(f_n(x)|Z) - E(f_n(x)))$$

converge en probabilidad a 0 cuando  $n \rightarrow \infty$ . Del mismo modo

$$\sqrt{n\psi(h_n)} (E(g_n(x)|Z) - E(g_n(x)))$$

converge en probabilidad a 0 cuando  $n \rightarrow \infty$ . Como la convergencia a 0 en probabilidad implica la convergencia en distribución a 0 se obtiene el resultado.  $\square$



La hipótesis H 4.16 se puede obtener si por ejemplo las variables  $1_{\{Z_i=z_k\}}$  verifican un teorema central del límite. En el libro de Billingsley (1995), capítulo 5, se encuentran las demostraciones de las propiedades de la convergencia en distribución utilizadas.

El resultado anterior y una descomposición adecuada del estimador permite obtener la distribución asintótica, asumiendo la siguiente hipótesis adicional.

**Teorema 4.17.** *Se asumen las hipótesis de la proposición 4.15 y la hipótesis*

**H 4.18.**  *$f_n(x)$  converge en probabilidad a  $f(x) > 0$  definida para  $x \in \mathcal{D}$  por*

$$f(x) = \sum_{k \in \Delta} p_k d_k(x) c_k(x),$$

donde  $\Delta$  está definido en la hipótesis H 4.3.

Entonces, cuando  $n$  tiende a  $\infty$

$$\sqrt{n\psi(h_n)} \left( \phi_n(x) - \frac{E(g_n(x))}{E(f_n(x))} \right) \xrightarrow{w} N_1(0, \sigma^2(x)),$$

donde

$$\sigma^2(x) = \frac{a_1(x) - 2a_2(x)\phi(x) + a_3(x)\phi^2(x)}{f^2(x)}. \quad (4.2)$$

*Demostración.* Se tiene

$$\phi_n(x) - \frac{E(g_n(x))}{E(f_n(x))} = \frac{Q_n(x) - B_n(x)(f_n(x) - E(f_n(x)))}{f_n(x)},$$

donde

$$B_n(x) = \frac{E(g_n(x))}{E(f_n(x))} - \phi(x)$$

y

$$Q_n(x) = (g_n(x) - E(g_n(x))) - \phi(x)(f_n(x) - E(f_n(x))).$$

Usando la proposición 4.1

$$\lim_{n \rightarrow \infty} B_n(x) = 0.$$

Además por la hipótesis H 4.18  $f_n(x)$  converge en probabilidad a  $f(x) > 0$ , de donde

$$\frac{B_n(x) [f_n(x) - E(f_n(x))]}{f_n(x)} \quad (4.3)$$

converge en probabilidad a 0 y por lo tanto en distribución, y se obtiene el resultado. Para el cálculo de la varianza hay que observar que de la descomposición de  $\phi_n(x)$  se obtiene que el límite en distribución es el límite de la variable  $\frac{Q_n(x)}{f_n(x)}$ , es decir la variable  $\frac{U - \phi(x)V}{f(x)}$  donde  $(U, V())$  es el límite en distribución de

$$\sqrt{n\psi(h_n)} \left( g_n(x) - E(g_n(x)), f_n(x) - E(f_n(x)) \right),$$

y tiene distribución  $N_2(0, A)$  según la proposición 4.15. Se tiene que, al ser las variables centradas

$$\begin{aligned} \text{Var} \left( \frac{U - \phi(x)V}{f(x)} \right) &= \frac{E(U^2 - 2\phi(x)UV + \phi^2(x)V^2)}{f^2(x)} \\ &= \frac{a_1(x) - 2a_2(x)\phi(x) + a_3(x)\phi^2(x)}{f^2(x)}. \end{aligned}$$

□

La hipótesis H 4.18 se verifica por ejemplo si se cumplen las hipótesis del teorema 4.9, donde se demuestra la convergencia completa de  $f_n(x)$  a  $f(x)$ .

La distribución asintótica de  $\phi_n(x) - \phi(x)$  utiliza hipótesis adicionales sobre la regularidad de  $\phi$  y el decaimiento de la ventana.

**Teorema 4.19.** *Se asumen las hipótesis del teorema 4.17 y las siguientes hipótesis*

**H 4.20.** *La función  $\phi$  cumple que, para todo  $u, v \in \mathcal{D}$ ,*

$$|\phi(u) - \phi(v)| \leq L\|u - v\|^r,$$

donde  $r > 0$  y  $L$  es una constante positiva.

**H 4.21.** *Existe  $M > 0$  y una función  $\alpha$  tal que para cada  $k \in \{1, \dots, m\}$  y  $h > 0$ , se tienen que*

$$\left| \frac{P[|\varphi(\xi_1, z_k) - x| \leq h]}{\psi_k(h, x)} - c_k(x) \right| \leq M\alpha(h) \quad \forall \alpha > 0.$$

**H 4.22.** *La ventana  $h_n$  cumple que*

$$\lim_{n \rightarrow \infty} \alpha(h_n) \sqrt{n\psi(h_n)} = 0,$$

$$\lim_{n \rightarrow \infty} h_n^r \sqrt{n\psi(h_n)} = 0$$

**H 4.23.** *Para cada  $k \in \{1, \dots, m\}$ ,*

$$\lim_{n \rightarrow \infty} \sqrt{n\psi(h_n)} \left( \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) - p_k \right) = 0.$$

Entonces, cuando  $n$  tiende a  $\infty$ ,

$$\sqrt{n\psi(h_n)} (\phi_n(x) - \phi(x)) \xrightarrow{w} N_1(0, \sigma^2(x)),$$

donde  $\sigma^2(x)$  está definida por la ecuación (4.2).

*Demostración.* Como se tiene

$$\begin{aligned}\sqrt{n\psi(h_n)}(\phi_n(x) - \phi(x)) &= \sqrt{n\psi(h_n)}\left(\phi_n(x) - \frac{E(g_n(x))}{E(f_n(x))}\right) \\ &\quad + \sqrt{n\psi(h_n)}\left(\frac{E(g_n(x))}{E(f_n(x))} - \phi(x)\right),\end{aligned}$$

por el teorema anterior basta probar que

$$\lim_{n \rightarrow \infty} \sqrt{n\psi(h_n)}\left(\frac{E(g_n(x))}{E(f_n(x))} - \phi(x)\right) = 0.$$

Se tiene la siguiente descomposición

$$\begin{aligned}\frac{E(g_n(x))}{E(f_n(x))} - \phi(x) &= \frac{1}{E(f_n(x))} [E(g_n(x)) - f(x)\phi(x)] \\ &\quad + \frac{\phi(x)}{E(f_n(x))} [f(x) - E(f_n(x))].\end{aligned}$$

Considerando  $E(f_n(x)) - f(x)$  se obtiene en la proposición 4.1 el siguiente resultado

$$E(f_n(x)) = \sum_{k=1}^m \left( \frac{1}{\psi(h_n)} E[K_n(\varphi(\xi, z_k))] \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \right)$$

y luego

$$\begin{aligned}E(f_n(x)) - f(x) &= \sum_{k=1}^m \frac{E[K_n(\varphi(\xi, z_k))] - d_k(x)c_k(x)1_{\{k \in \Delta\}}}{\psi(h_n)} \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \\ &\quad + \sum_{k=1}^m d_k(x)c_k(x)1_{\{k \in \Delta\}} \left( \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) - p_k \right),\end{aligned}$$

y aplicando las hipótesis H 4.21, H 4.22 y H 4.1 se tiene la convergencia en probabilidad a 0 de

$$\sqrt{n\psi(h_n)}(f_n(x) - f(x)).$$

Para  $E(g_n(x)) - f(x)\phi(x)$  se cumple

$$|E(g_n(x)) - f(x)\phi(x)| \leq |\phi(x)| |E(f_n(x)) - f(x)| + |E(g_n(x)) - E(f_n(x))\phi(x)|.$$

Entonces

$$\begin{aligned}&|E(g_n(x)) - E(f_n(x))\phi(x)| \\ &\left| \frac{1}{\psi(h_n)} \sum_{k=1}^m E[K_n(\varphi(\xi, z_k)) (\phi(\varphi(\xi, z_k)) - \phi(x))] \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k) \right| \\ &\leq Lh_n^r \sum_{k=1}^m \frac{1}{\psi(h_n)} E \left[ K_n(\varphi(\xi, z_k)) \left\| \frac{\varphi(\xi, z_k) - x}{h_n} \right\|^r \right] \frac{1}{n} \sum_{i=1}^n P(Z_i = z_k).\end{aligned}$$

Con cálculos similares a los de la proposición 4.1 se tiene que

$$E \left[ \left\| \frac{\varphi(\xi_1, z_k) - x}{h_n} \right\|^r K_n(\varphi(\xi_1, z_k)) \right] = h_n c_k(x) \int_0^1 K(u) u^r \psi'_k(uh_n, x) du$$

$$\leq d_k(x) c_k(x),$$

de donde, usando la hipótesis H 4.20 se obtiene que

$$\lim_{n \rightarrow \infty} \sqrt{n\psi(h_n)} (E(g_n(x)) - f(x)\phi(x)) = 0.$$

□

## Capítulo 5

# Aplicaciones

En este capítulo se describe con más detalle el problema de telecomunicaciones que motiva este trabajo, cuáles son los datos y cómo se aplican los resultados presentados en los capítulos anteriores. Se presentan resultados para datos simulados y para datos reales.

### 5.1. Descripción del problema

Algunas aplicaciones en redes de datos, como por ejemplo aplicaciones de voz y video en tiempo real, tienen requerimientos de calidad de servicio mayores que otras aplicaciones, como pueden ser el correo electrónico o la transferencia de archivos. Por ejemplo no es posible realizar una conversación o una video conferencia si el retardo de los paquetes supera determinado valor. Por otra parte las pérdidas de paquetes en el caso de transferencia de archivos u otras aplicaciones similares se resuelven volviendo a enviar los paquetes que se perdieron, y esto se traduce en mayores retardos. Por este motivo este tipo de retransmisiones no es adecuado para las aplicaciones en tiempo real. Entonces, como no es fácil atenuar el efecto de las pérdidas o de los retardos, es imprescindible conocer el estado de la red en cuanto a la calidad de servicio que recibe el tráfico de voz o video y para esto se realizan medidas en la red.

El objetivo por lo tanto es medir parámetros de calidad de servicio de extremo a extremo tomando en cuenta las particularidades de las aplicaciones de tráfico de voz o video en tiempo real.

Estas medidas de calidad de servicio generalmente se realizan enviando tráfico de prueba. Una posibilidad es enviar tráfico de voz o video y calcular los parámetros de calidad de servicio de interés, por ejemplo pérdidas, retardo, variación del retardo. La dificultad para realizar estas medidas es que el tráfico, particularmente el video, sobrecarga la red, con lo cual no es posible realizar muchas medidas sin afectar la calidad de servicio de todos los usuarios.

Otras medidas en la red se realizan a partir de tráfico de prueba donde se generan paquetes livianos, y se envían por ejemplo paquetes periódicos con el

tamaño mínimo posible. Sin embargo no necesariamente el retardo o las pérdidas que sufrirá el tráfico de voz o video que se trasmite por ese mismo camino será similar al que se puede medir sobre el tráfico de prueba. Por otra parte, si bien las medidas sobre los paquetes de prueba no son iguales a las medidas sobre el tráfico de voz o video que interesa estudiar, existe una relación entre ambas. Esta es una de las hipótesis principales que se asume en este trabajo. Los parámetros de calidad de servicio de extremo a extremo para un tráfico de voz o video dependen del estado de la red, que está determinado por las características fijas de ésta, como la topología, las capacidades, o el tamaño máximo de las colas, y donde la variabilidad de estos estados en el tiempo depende del tráfico de todos los usuarios presentes. Se tienen estados diferentes según la cantidad de usuarios en la red y del tipo de tráfico que éstos generen. Se asume que el estado de la red se puede identificar mediante tráfico liviano, de modo que el problema consiste en caracterizar este estado de alguna manera para luego determinar la relación que existe entre esta caracterización y los parámetros de calidad de servicio de extremo a extremo para tráfico multimedia. Una vez hecha esta identificación se podrán obtener los parámetros de calidad de servicio para el tráfico de voz o video a partir del tráfico de prueba liviano, que caracteriza el estado de la red, y los parámetros de calidad de servicio están identificados previamente con los distintos estados.

Para identificar el estado de un único enlace una técnica usual consiste en enviar paquetes con el menor tamaño posible en forma periódica, o con determinada distribución de probabilidad para el tiempo entre envío de paquetes, y medir el tiempo entre paquetes a la llegada (Bolot 1993) (Dovrolis, Ramanathan & Moore 2001) (Lai & Baker 2001). En el caso de un enlace si el tiempo entre arribos de los paquetes es mayor que el tiempo entre salidas se asume que hay tráfico en la red de modo que los paquetes de prueba permanecen en una cola junto con paquetes correspondientes al tráfico de otro usuarios, lo que provoca esta separación entre los paquetes de prueba. De esta forma el tiempo entre llegadas de los paquetes brinda una medida del estado del enlace, y se puede asumir que a mayor tiempo entre paquetes existe más congestión en dicho enlace, donde esta congestión está provocada por el mayor volumen de tráfico en la red. Por otra parte si bien existen diversas situaciones que modifican esta suposición también existen diferentes técnicas para abordar esos inconvenientes que se encuentran detalladas en los trabajos citados. Para extender el argumento anterior de un enlace a una red, es decir cuando se considera un camino entre dos extremos con más de un enlace, muchas veces se asume que hay un único enlace que es el cuello de botella. En este caso el resto de los enlaces se comportan como si la red estuviera vacía, y las ideas anteriores son válidas en el enlace cuello de botella, de modo que el tiempo entre paquetes continúa siendo una caracterización del estado de la red. Sin embargo se puede suponer además que el tiempo entre paquetes de prueba siempre brinda una caracterización del estado de la red, aunque no esté relacionada directamente con el volumen de tráfico que existe en la red. Se puede probar que la estimación del tráfico en la red a partir de paquetes de prueba depende de las características de dicho tráfico y no siempre es posible tener una estimación de éste basada únicamente

en estos paquetes de prueba (Machiraju, Veitch, Baccelli, Nucci & Bolot 2005). En el trabajo citado los autores demuestran que no es posible obtener de manera general la distribución del proceso de tráfico de todos los usuarios a partir de paquetes de prueba, y que distintos tipos de tráfico conducen no sólo a las mismas observaciones sino a la misma distribución para el proceso estimado. De todos modos no es un objetivo de este trabajo tener una estimación precisa del tráfico en la red, sino tener alguna caracterización del estado de la red en base a paquetes de prueba y su correspondencia con parámetros de calidad de servicio para el tráfico de video o voz que se quiere analizar.

Para identificar medidas realizadas sobre tráfico de prueba liviano con parámetros de calidad de servicio para el tráfico de voz o video se considera el siguiente método de medición. Primero se envía una ráfaga de tráfico periódico formado por paquetes livianos, y luego una secuencia corta de video (o de voz) que sobrecarga un poco más la red, pero por un período de tiempo breve. Con varios de estos experimentos el objetivo es poder representar la relación que existe entre el estado de la red medido mediante el tráfico de prueba periódico y los parámetros de calidad de servicio del video. Es decir que se asume un modelo como el desarrollado en los capítulos anteriores, para modelar la relación entre el estado de la red y los parámetros de calidad de servicio, con la siguiente interpretación. Se considera

$$Y = \phi(X) + \varepsilon,$$

donde la variable  $Y$  es el parámetro de calidad de servicio para el tráfico multimedia,  $X$  es una variable que se obtiene a partir del tráfico de prueba periódico y  $\varepsilon$  es una variable centrada e independiente de  $X$  que representa el error. Una vez que se tiene una estimación  $\phi_n$  de  $\phi$ , obtenida a partir de un conjunto de observaciones,

$$\{(X_i, Y_i) : i = 1, \dots, n\},$$

para predecir calidad de servicio sólo será necesario enviar tráfico de prueba periódico, de modo de obtener un estado de la red  $x$  y estimar el parámetro de calidad de servicio mediante  $y = \phi_n(x)$ .

Por otra parte la intención es caracterizar de la manera más sencilla posible mediante el tiempo entre llegadas de paquetes el estado de la red. En los experimentos cada observación es una secuencia larga de tiempos entre paquetes junto con alguna medida de calidad de servicio para la secuencia de voz o video, por ejemplo retardo medio de esa secuencia. Sin embargo no es fácil caracterizar el estado de la red con unos pocos parámetros como podría ser la media del tiempo entre paquetes o media y varianza, etc.. La función de distribución del tiempo entre paquetes puede dar una buena caracterización, si bien se pierden las correlaciones temporales. Por otra parte tampoco parece una ventaja considerar directamente la serie temporal de los tiempos entre paquetes, pues series temporales muy distintas pueden representar estados de la red similares, y también en ese caso sería necesario un volumen de datos mucho mayor para estimar. Además, si bien es posible construir ejemplos donde la misma función

de distribución para el tiempo entre paquetes represente características muy diferentes de la red, esto no es lo que ocurre en los ejemplos considerados en redes reales. De todas maneras sigue siendo un tema para analizar cuáles son las características de la secuencias de tiempos entre paquetes que permiten representar adecuadamente el estado de la red.

Teniendo en cuenta lo anterior se consideran datos  $(X, Y)$  donde  $X$  es la distribución empírica del tiempo entre paquetes e  $Y$  es algún parámetro de calidad de servicio medido sobre la secuencia de video. Se asume que los datos están en las hipótesis de los resultados presentados, es decir que  $X$  es funcional y es una mezcla de un proceso estacionario débilmente dependiente y de otro que puede ser no estacionario y dependiente, según el modelo que se presentó en el capítulo 4. En este modelo se está asumiendo que  $X$  representa el estado de la red, y es de la forma  $X = \varphi(\xi, Z)$ , donde el proceso  $\xi$  es estacionario y débilmente dependiente pero  $Z$  no lo es. Se modela mediante  $Z$  la selección entre diferentes regímenes de tráfico, donde esto depende por ejemplo de la hora, del día, etc.. Por otra parte para cada tipo de tráfico, es decir condicionado a  $Z$ , se tendrá un comportamiento estacionario determinado por  $\xi$ .

En la próxima sección se presentará un ejemplo básico de datos con estas características, y en las secciones siguientes se presentan datos simulados en una red donde se considera un modelo de tráfico determinado y luego datos obtenidos en una red real.

La estimación puntual se realiza calculando el estimador

$$\phi_n(x) = \frac{\sum_{i=1}^n Y_i K_n(X_i)}{\sum_{i=1}^n K_n(X_i)}$$

definido en los capítulos anteriores. Para la construcción de intervalos de confianza es necesario estimar la varianza del estimador. En este trabajo se obtiene esta estimación mediante bootstrap. El bootstrap para variables funcionales en un modelo de regresión es usado en el trabajo de Ferraty, Mas & Vieu, (2007), donde se realiza bootstrap de los residuos

$$\varepsilon_i = Y_i - \phi_n(X_i),$$

siendo  $\phi_n$  un estimador de  $\phi$ . De esta forma obtienen secuencias bootstrap de  $n$  datos,

$$\{\varepsilon_1^j, \dots, \varepsilon_n^j\},$$

con  $j = 1, \dots, b$ . A partir de estas se obtienen secuencias

$$\{(X_1^j, Y_1^j), \dots, (X_n^j, Y_n^j)\},$$

con  $j = 1, \dots, b$ , donde  $Y_i^j = \phi_n^j(X_i) + \varepsilon_i^j$ , siendo  $\phi_n^j$  el estimador calculado para una ventana diferente a la utilizada para calcular  $\phi_n$ . Finalmente para la ventana original, con la que se calcula  $\phi_n$ , se calcula  $\tilde{\phi}_n^j$ , con la muestra bootstrap

$$\{(X_1^j, Y_1^j), \dots, (X_n^j, Y_n^j)\}.$$



En el trabajo citado los autores usan el bootstrap como método para calcular la ventana óptima, repitiendo lo anterior para las  $b$  réplicas y eligiendo la ventana que minimiza

$$\frac{1}{b} \sum_{j=1}^b \left( \tilde{\phi}_n^j(x) - \phi_n(x) \right)^2.$$

El resultado teórico sobre la validez del bootstrap en ese caso no está demostrado y los autores indican ese punto como uno de los temas para trabajo futuro.

En este trabajo, al igual que en el referido, tampoco se realiza el tratamiento teórico de la validez del bootstrap. Si bien no se estudia este tema, por ser el estimador  $\phi_n$  un estimador basado en promedios y donde no influye la correlación de las observaciones para calcularlo, se asume que el bootstrap para datos independientes e idénticamente distribuidos realizado directamente sobre las parejas  $(X, Y)$  basta para estimar la varianza en los datos simulados y en los datos reales.

## 5.2. Un ejemplo básico

En esta sección se construye un ejemplo de datos que siguen el modelo propuesto en los capítulos anteriores y se calcula el estimador y los intervalos de confianza correspondientes.

Se considera una variable  $U$  con distribución exponencial de parámetro  $\lambda$ . Para cada  $i$  se sortean  $m$  observaciones  $\{U_1^i, \dots, U_m^i\}$  y se tiene la distribución empírica  $\xi_i$ , es decir que cada  $\xi_i$  es una función definida por

$$\xi_i(t) = \frac{1}{m} \sum_{k=1}^m 1_{\{U_k^i \leq t\}}.$$

De esta manera se obtiene una sucesión  $(\xi_n)_{n \geq 1}$  estacionaria (en este caso es i.i.d.) en un espacio de funciones. Se considera además una variable  $Z$  periódica (con período aleatorio) tomando valores en  $\{0, 1\}$ , y la variable  $X_i$  es una función definida por  $X_i(t) = \xi_i(t - Z_i)$ . Se puede observar que  $X_i$  es la distribución empírica de  $U_1^i + Z_i, \dots, U_m^i + Z_i$ . La variable  $X$  corresponde al modelo  $X = \varphi(\xi, Z)$  definido en el capítulo 4.

Se considera además una variable  $Y = \phi(X) + \varepsilon$  definida como el percentil 0.9 de la variable  $X$ , es decir  $\phi(X) = X^{-1}(0,9)$  y con  $\varepsilon$  una variable gaussiana centrada e independiente de  $X$ .

Para una secuencia de 100 datos

$$\{(X_1^1, Y_1^1), \dots, (X_{100}^1, Y_{100}^1)\}$$

se calcula el estimador de  $Y_i^1$  para cada  $i = 1, \dots, 100$ , mediante  $\phi_n(X_i)$  a partir de otra secuencia

$$\{(X_1^2, Y_1^2), \dots, (X_{100}^2, Y_{100}^2)\}$$

con la misma distribución. La primera secuencia es la secuencia de validación y la segunda es la secuencia de entrenamiento o de estimación. Es decir que para

estimar cada valor de  $Y$  para la primera secuencia se utilizan los datos de la segunda secuencia. En la figura 5.1 se presenta  $X_1^1$ , es decir uno de los puntos donde se calcula una de las estimaciones, y 30 valores de la segunda secuencia,  $X_1^2, \dots, X_{30}^2$ , usados para estimar. Cada  $X$  es la distribución empírica para una muestra de 200 datos, es decir que se sortean 200 exponenciales para obtener cada valor de  $X$ . En la figura 5.2 se presentan los datos  $Y^1 = (Y_1^1, \dots, Y_{100}^1)$ , de la secuencia de validación, es decir que son los valores que se quiere estimar, y los valores de la secuencia de entrenamiento  $Y^2 = (Y_1^2, \dots, Y_{100}^2)$ , usados para construir el estimador.

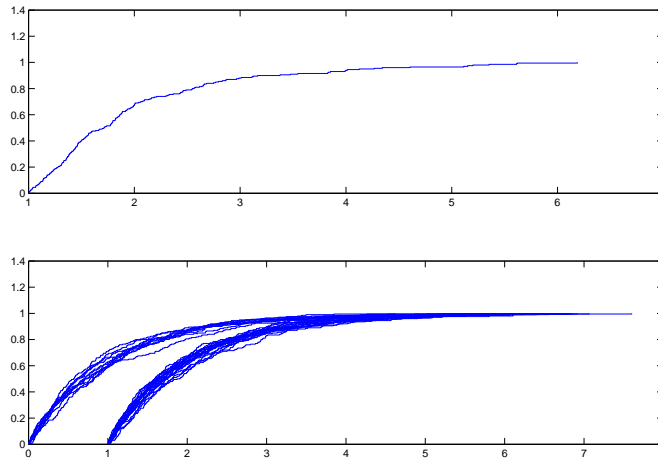


Figura 5.1: Las variables  $X$  son funciones de distribución empíricas. En la gráfica superior  $X_1^1$  es uno de los puntos donde se quiere estimar  $Y$ . En la gráfica inferior se presentan 30 observaciones  $X_1^2, \dots, X_{30}^2$  de  $X$  de la secuencia de entrenamiento. En ambos caso las exponenciales  $U$  tienen parámetro  $\lambda = 1$ .

Se considera la norma en  $L^1$  para medir la distancia entre las funciones de distribución empíricas, y el núcleo  $K(x) = (x^2 - 1)^2 1_{\{|x| < 1\}}$ . Se calcula la estimación puntual y el error relativo comparando con el valor real. Para estimar cada  $Y_i^1$ , con  $i = 1, \dots, 100$  se calculan las distancias de  $X_i^1$  a cada una de las observaciones  $X_1^2, \dots, X_{100}^2$  y se calcula el estimador

$$\phi_n(X_j^1) = \frac{\sum_{i=1}^n Y_i^2 K_n(X_i^2)}{\sum_{i=1}^n K_n(X_i^2)},$$

siendo  $K_n(X_i^2) = K\left(\frac{\|X_i^2 - X_j^1\|}{h_n}\right)$ . Previamente la ventana se puede elegir de modo de minimizar el error para la muestra de entrenamiento. Es decir que para esta muestra se calcula también la estimación de cada valor con las demás observaciones y se elige la ventana que minimice la suma de los errores. Se

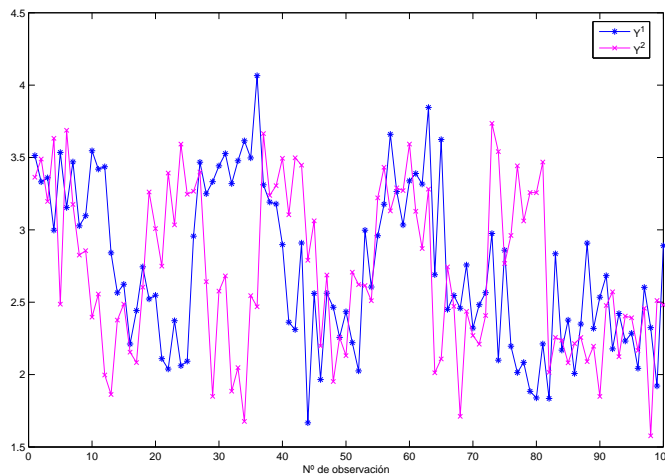


Figura 5.2:  $Y^1$  es la secuencia de valores que se quiere estimar,  $Y^2$  corresponde a la secuencia de estimación.  $\varepsilon$  tiene distribución  $N(0, (0, 1)^2)$ .

estima  $Y_j^2$  para cada  $j = 1, \dots, 100$ , a partir de la muestra de entrenamiento  $(X_1^2, Y_1^2), \dots, (X_{100}^2, Y_{100}^2)$ , que se obtiene excluyendo  $(X_j^2, Y_j^2)$  y se halla la ventana de modo de minimizar la suma de los errores

$$\sum_{i=1}^{100} (\phi_n(X_j^2) - Y_j^2)^2.$$

La estimación de cada valor de la secuencia de validación obtenidas según las consideraciones anteriores con la secuencia de estimación y el error relativo en cada punto se presentan en las figuras 5.3 y 5.4. Se tiene un error relativo medio de 0,0476.

Para la construcción de intervalos de confianza en cada punto se realizan 25 réplicas y se estima la varianza a partir de las réplicas, es decir que se estima cada uno de los valores de  $Y$  para  $X_1^1, \dots, X_{100}^1$  a partir de secuencias  $(X^3, Y^3), \dots, (X^{27}, Y^{27})$  y con estas estimaciones se construye la varianza y el intervalo de confianza. La estimación y el intervalo de confianza al 5% se presentan en la figura 5.5.

Los datos fueron generados en MATLAB, así como los programas para calcular las estimaciones. También se utilizaron programas realizados por Karine Bertin y Yoana Donoso.

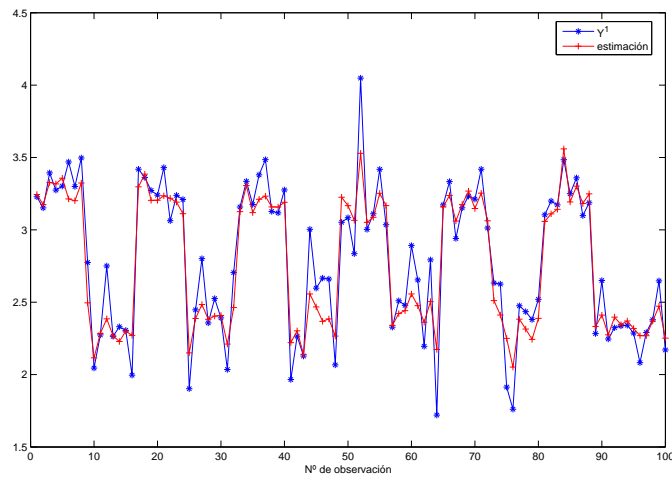


Figura 5.3: Valores estimados y valores de la muestra de validación.

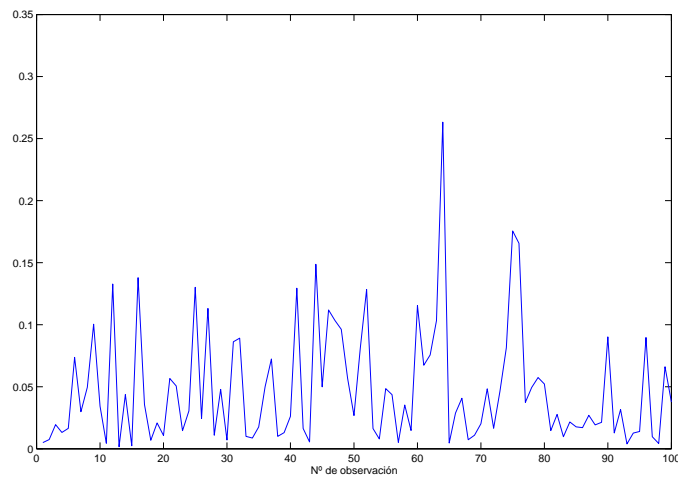


Figura 5.4: Error relativo en cada punto. Para cada estimación de  $Y_i^1$  el error relativo se calcula como  $\frac{|\phi_n(X_i^1) - Y_i^1|}{Y_i^1}$ .

### 5.3. Datos simulados

En esta sección se presentan los resultados para datos simulados en una red. Se simula el tráfico en un enlace según un proceso ON-OFF, que es un

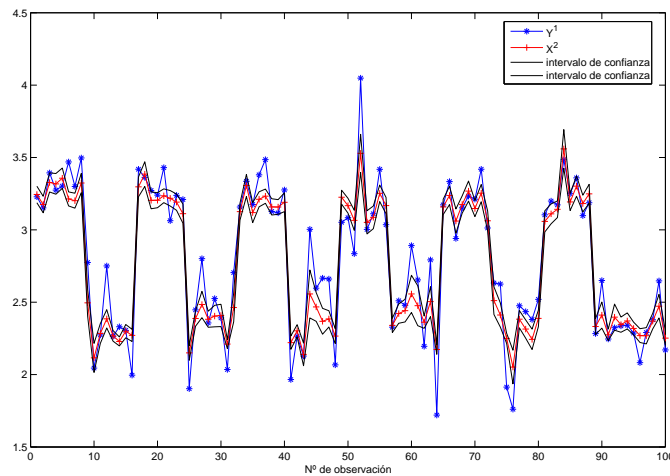


Figura 5.5: Estimación e intervalo de confianza en cada punto. La varianza está calculada con 25 réplicas.

modelo estándar para el tráfico en Internet. El proceso ON-OFF considerado tiene media variable en dos rangos diferentes, un rango correspondiente a un régimen de poco tráfico y otro correspondiente a un régimen de mucho tráfico.

El tráfico para un experimento se simula mediante un proceso ON-OFF de la siguiente manera. Los tiempos ON y OFF corresponden a variables exponenciales independientes. Luego se quiere simular el comportamiento no estacionario del tráfico, mediante un variable se selecciona periódicamente entre el régimen de poco tráfico y el de mucho tráfico. Para cada tipo de tráfico se sorteja la media de las exponenciales y la media de tráfico dentro de uno de dos rangos. Para el primer rango la tasa de bits del tráfico tiene media entre 150 Mb/s y 450 Mb/s y tiempo promedio en estado ON y OFF variando entre 100 y 300 ms. En el segundo rango la tasa media es entre 600 Mb/s y 900 Mb/s y el tiempo medio en el estado ON y OFF varía entre 200 500 ms. Con la media de tráfico sorteada y los tiempos ON y OFF se determina el tiempo entre paquetes en el período ON de modo de tener la media de tráfico sorteada para la realización correspondiente del proceso ON-OFF.

Asumiendo este tráfico en el enlace se desarrolla el procedimiento de medición descrito antes. Se realizan pruebas que consisten en envío de paquetes periódicos livianos para los que se calcula el tiempo entre llegadas y su función de distribución y luego de cada ráfaga de paquetes de prueba se envía una secuencia corta que simula el video, también mediante un proceso ON-OFF, para la cual se calcula el retardo medio. Es decir que cada realización del proceso ON-OFF del tráfico tiene asociada una función de distribución empírica del tiempo entre paquetes de prueba y un número correspondiente al retardo medio medido

sobre la secuencia de video simulado.

Los datos presentan periodicidad al alternarse los regímenes de mucho y poco tráfico y se puede considerar que la distribución empírica del tiempo entre paquetes responde al modelo  $X = \varphi(\xi, Z)$  estudiado en el capítulo 4 donde  $Z$  es una variable dependiente y no estacionaria que toma dos valores y decide entre los estados de mucho y poco tráfico, mientras que en cada estado puede asumirse que el tráfico es estacionario.

El modelo puede interpretarse de la siguiente manera. La variable  $\xi$  toma valores en un conjunto de funciones de distribución, que son las funciones de distribución del tiempo entre paquetes. Éstas varían según la media del tráfico, donde existen dos rangos para la media, correspondientes a mayor y menor tráfico presente en la red. La variable  $Z$  elige entre los dos grupos determinados por mayor o menor media, y luego dentro de cada grupo se sortea una función de distribución, correspondiente al sorteo de la media de tráfico de manera uniforme en el rango seleccionado.

Las simulaciones fueron realizadas en Ns-2 (McCanne & Floyd 2008) por Pablo Belzarena.

Para el tráfico simulado se tienen 360 observaciones  $(X, Y)$ . Se separa la muestra en 2,  $(X^1, Y^1)$  con los valores de las observaciones para  $i = 1, \dots, 60$  (muestra de validación) y  $(X^2, Y^2)$  con los valores de las observaciones para  $i = 61, \dots, 360$  (muestra de entrenamiento). Se calcula el estimador  $\phi_n(X_i^1)$ , para cada  $i = 1, \dots, 60$ , a partir de las 300 observaciones de la segunda muestra. El núcleo es  $K(x) = (x^2 - 1)^2 1_{\{|x| < 1\}}$  y el valor de la ventana se obtiene minimizando el error para la muestra de entrenamiento, estimando cada punto con los restantes 299 restantes. La varianza para el intervalo de confianza se calcula con bootstrap. En las figuras 5.6, 5.7 y 5.8 se presentan la estimaciones para los valores de la muestra de validación, el error relativo y el intervalo de confianza en cada punto. El error relativo es mayor en los períodos OFF donde hay poco tráfico en la red y los valores de retardo a estimar son muy pequeños.

## 5.4. Datos reales

Estos datos se obtuvieron mediante un software desarrollado en el Instituto de Ingeniería Eléctrica de la Facultad de Ingeniería (Belzarena, González Barbone, Larroca & Casas 2006) mediante el cual un usuario se conecta a través de una página web que envía paquetes de prueba, y el software del usuario envía esos paquetes de prueba nuevamente a su origen. Inmediatamente después de cada ráfaga de paquetes de prueba se envía una secuencia de video que el software del usuario envía nuevamente a su origen. Se mide el retardo en el trayecto de ida y vuelta (RTT por round trip time) tanto para los paquetes de prueba periódicos como para las secuencias de video. El objetivo de este software es obtener suficientes datos para tener una estimación de la función  $\phi$  y luego enviar únicamente paquetes de prueba livianos para caracterizar el estado de la red y estimar la calidad de servicio para tráfico multimedia, sin tener que en-

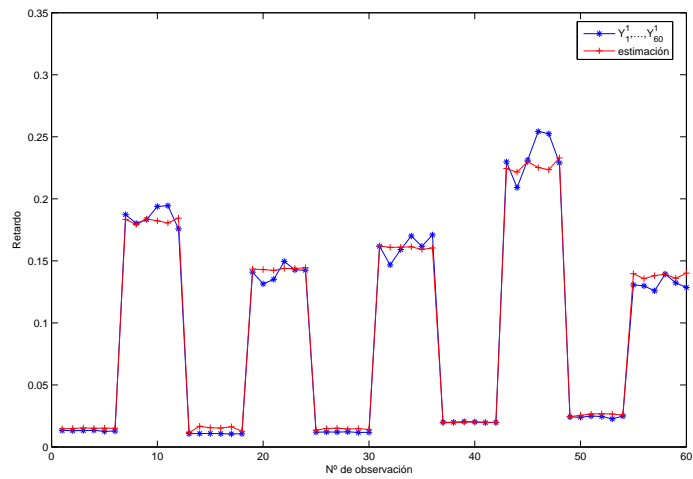


Figura 5.6: Retardo estimado para datos simulados.

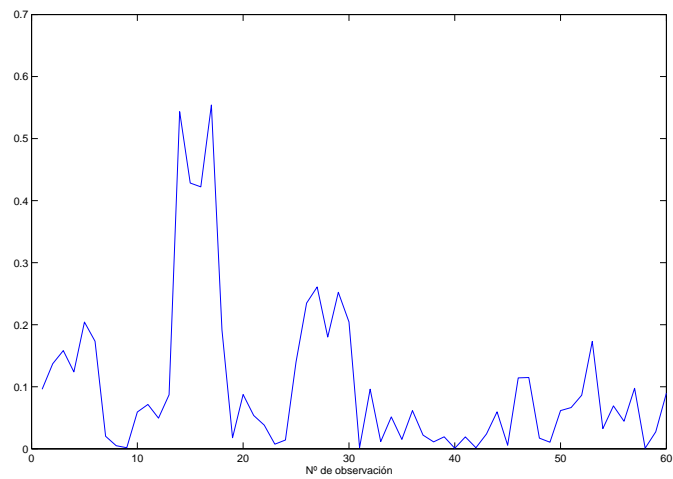


Figura 5.7: Error relativo para datos simulados.

viarlo, en base a la estimación previa. Las pruebas se realizaron entre el servidor arbolita.fing.edu.uy y un usuario conectado mediante un ADSL domiciliario. Se tiene una muestra de 50 observaciones correspondientes al tiempo entre paquetes para las ráfagas de paquetes de prueba de donde se obtiene la distribución empírica, junto con el RTT para la secuencia de video. En la figura 5.9 se

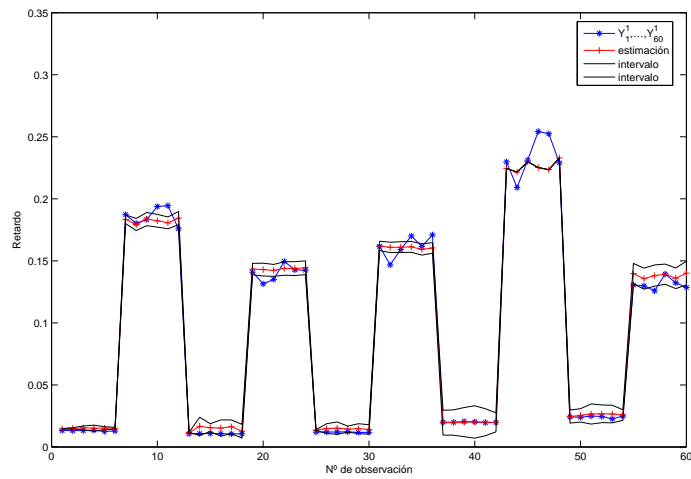


Figura 5.8: Intervalo de confianza para datos simulados, calculado con 25 réplicas.

muestran 30 datos  $X$ . Estos datos se usarán al mismo tiempo como muestra

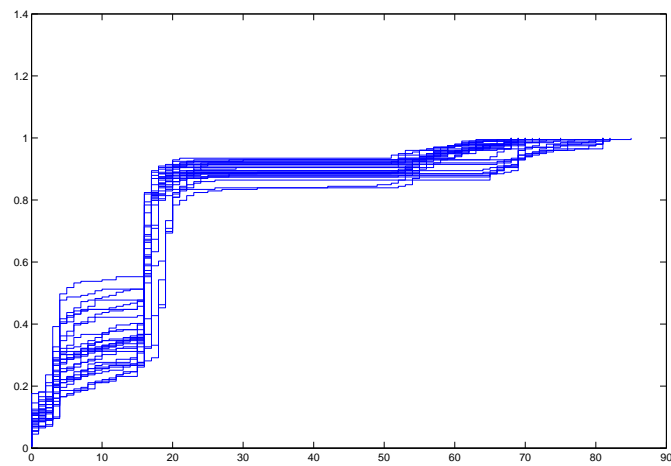


Figura 5.9: Las variables  $X$  son las funciones de distribución empíricas. Se presentan 30 observaciones  $X_1, \dots, X_{30}$ .

de entrenamiento y de validación. Para cada observación se estima  $Y_i$  mediante  $\phi_n(X_i)$  donde  $\phi_n$  es la estimación obtenida con todos los demás datos, usando



el mismo núcleo y procedimiento para elegir la ventana que en los ejemplos anteriores. Los ejemplos anteriores. La estimación y el error relativo se presentan en las figuras 5.10 y 5.11. Los intervalos de confianza se obtienen calculando la varianza mediante bootstrap para 25 réplicas y se presentan en la figura 5.12.

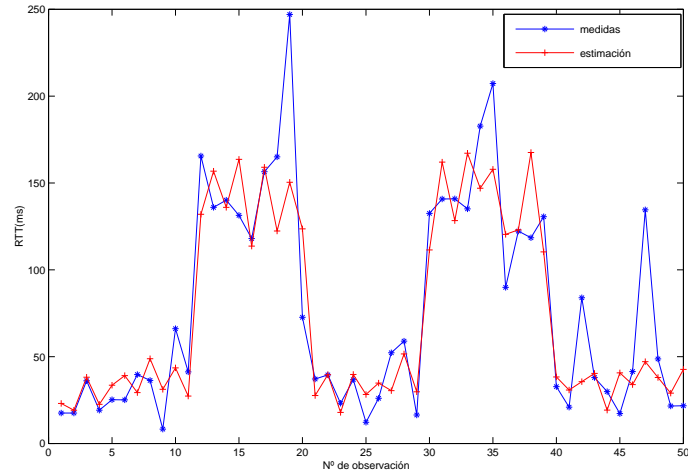


Figura 5.10: RTT real y estimado para datos reales.

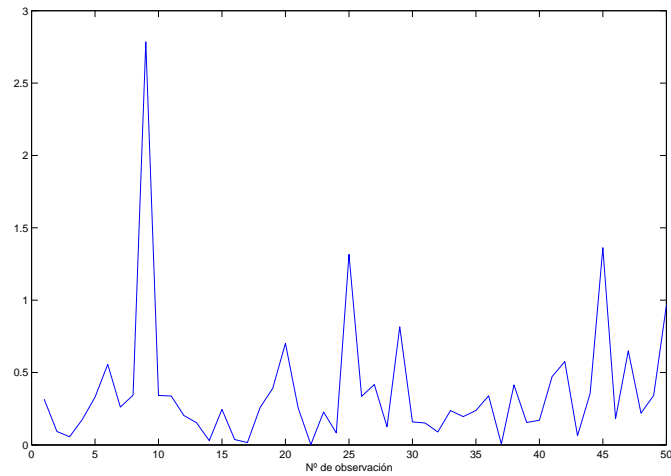


Figura 5.11: Error relativo para datos reales.

Los resultados para datos reales tienen menor precisión que en el caso de datos

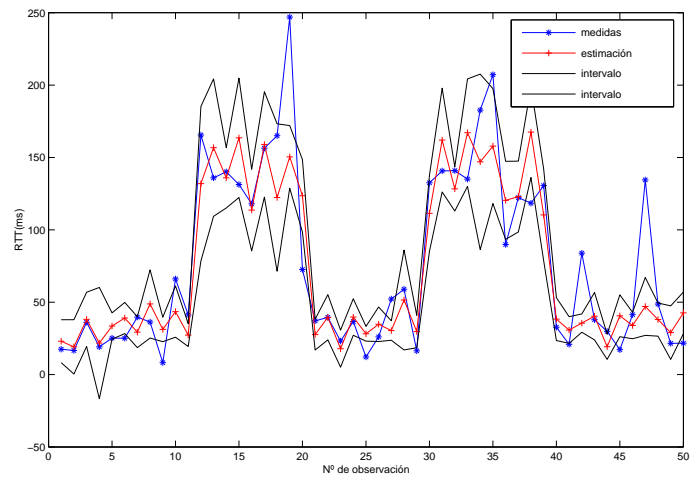


Figura 5.12: Intervalo de confianza para datos reales. La varianza está calculada con 25 réplicas.

simulados, pero se obtienen de todas formas buenas aproximaciones. Existen diversos temas a analizar cuando se desea tener una mejor caracterización del estado de la red, como son la escala de tiempo para el envío de paquetes de prueba, el tamaño de los paquetes así como el largo de la secuencia de video. Sin embargo se obtienen estimaciones bastante precisas para un procedimiento de medición y de estimación simple.

# Bibliografía

- Baccelli, F. & Bremaud, P. (2003), *Elements of Queueing Theory. Palm Martingale Calculus and Stochastic Recurrences.*, Vol. 26 of *Applications of Mathematics*, Springer.
- Belzarena, P., González Barbone, V., Larroca, F. & Casas, P. (2006), Metronet: software para medición de calidad de servicio en voz y video, in ‘Congreso Iberoamericano de Telemática, CITA’.
- Billingsley, P. (1995), *Probability and measure*, Wiley Series in Probability and Mathematical Statistics, third edn, John Wiley & Sons Inc., New York. A Wiley-Interscience Publication.
- Bolot, J.-C. (1993), ‘Characterizing end-to-end packet delay and loss in the internet’, *Journal of High Speed Networks* **2**, 305–323.
- CAIDA (2008), *Cooperative Association for Internet Data Analysis*.  
\*<http://www.caida.org/home/> [Consulta 05/2008]
- Dabo-Niang, S., Ferraty, F. & Vieu, P. (2004), ‘Estimation du mode dans un espace vectoriel semi-normé’, *C. R. Math. Acad. Sci. Paris* **339**(9), 659–662.
- De Boer, P. T. (2000), Analysis and efficient simulation of queueing models of telecommunication systems, PhD thesis, University of Twente.
- Dovrolis, C., Ramanathan, P. & Moore, D. (2001), What do packet dispersion techniques measure, in ‘IEEE INFOCOM’, pp. 905–914.
- Ferraty, F., Goia, A. & Vieu, P. (2002), ‘Functional nonparametric model for time series: a fractal approach for dimension reduction’, *Test* **11**(2), 317–344.
- Ferraty, F., Mas, A. & Vieu, P. (2007), ‘Nonparametric regression on functional data: inference and practical aspects’, *Aust. N. Z. J. Stat.* **49**(3).
- Ferraty, F. & Vieu, P. (2004), ‘Nonparametric models for functional data, with application in regression, time-series prediction and curve discrimination’, *J. Nonparametr. Stat.* **16**(1-2), 111–125.

- Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional data analysis: Theory and Practice*, Springer Series in Statistics, Springer, New York.
- Gasser, T., Hall, P. & Presnell, B. (1998), ‘Nonparametric estimation of the mode of a distribution of random curves’, *J. R. Stat. Soc. Ser. B Stat. Methodol.* **60**(4), 681–691.
- Greblicki, W., Krzyzak, A. & Pawlak, M. (1984), ‘Distribution-free pointwise consistency of kernel regression estimate’, *Ann. Statist.* **12**, 1570–1575.
- IPPM (2008), *IP Performance Metrics*.  
\*<http://www.ietf.org/html.charters/ippm-charter.html> [Consulta 05/2008]
- Karagiannis, T., Molle, M., Faloutsos, M. & Broido, A. (2004), A nonstationary poisson view of internet traffic, in ‘INFOCOM 2004. Twenty-third Annual Joint Conference of the IEEE Computer and Communications Societies’.
- Kelly, F. (1996), Notes on effective bandwidths, in F. P. Kelly, S. Zachary & I. Zeidins, eds, ‘Stochastic Networks: Theory and Applications’, Oxford University Press, pp. 141–168.
- Kleinrock, L. & Gail, R. (1996), *Queueing systems*, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York. Problems and solutions.
- Lai, K. & Baker, M. (2001), Nettimer: a tool for measuring bottleneck link, bandwidth, in ‘USITS’01: Proceedings of the 3rd conference on USENIX Symposium on Internet Technologies and Systems’, USENIX Association, Berkeley, CA, USA, pp. 11–11.
- Machiraju, S., Veitch, D., Baccelli, F., Nucci, A. & Bolot, J. C. (2005), Theory and practice of cross-traffic estimation, in ‘SIGMETRICS ’05: Proceedings of the 2005 ACM SIGMETRICS international conference on Measurement and modeling of computer systems’, ACM, New York, NY, USA, pp. 400–401.
- Masry, E. (2005), ‘Nonparametric regression estimation for dependent functional data: asymptotic normality’, *Stochastic Process. Appl.* **115**(1), 155–177.
- McCanne, S. & Floyd, S. (2008), *ns Network Simulator*.  
\*<http://www.isi.edu/nsnam/ns/> [Consulta 08/2008]
- Nadaraya, È. A. (1964), ‘On estimating regression’, *Theory of Probability and its Applications* **9**(1), 141–142.
- Nadaraya, È. A. (1989), *Nonparametric estimation of probability densities and regression curves*, Vol. 20 of *Mathematics and its Applications (Soviet Series)*, Kluwer Academic Publishers Group, Dordrecht.
- Oliveira, P. (2005), Nonparametric density and regression estimation functional data., Prepublication 05-09, Departamento de Matemática, Universidade de Coimbra.

- Ozturk, O., Mazumdar, R. R. & Likhanov, N. (2004), ‘Many sources asymptotics for networks with small buffers’, *Queueing Syst. Theory Appl.* **46**(1-2), 129–147.
- Perera, G. (1997), ‘Geometry of  $\mathbb{Z}^d$  and the central limit theorem for weakly dependent random fields’, *J. Theoret. Probab.* **10**(3), 581–603.
- Perera, G. (2002), ‘Irregular sets and central limit theorems’, *Bernoulli* **8**(5), 627–642.
- Ramsay, J. O. & Silverman, B. W. (2005), *Functional data analysis*, Springer Series in Statistics, second edn, Springer, New York.
- Robinson, P. M. (1983), ‘Nonparametric estimators for time series’, *J. Time Ser. Anal.* **4**(3), 185–207.
- Tablar, A. (2007), Estadística de procesos estocásticos y aplicaciones a modelos epidémicos, PhD thesis, Universidad del Sur, Bahía Blanca, Argentina.
- Watson, G. S. (1964), ‘Smooth regression analysis’, *Sankhyā Ser. A* **26**, 359–372.
- Willinger, W., Paxon, V., Reidi, R. & Taqqu, M. (2001), Long-range dependence and data network traffic, in G. O. P. Doukhan & M. S. Taqqu, eds, ‘Long-range Dependence: Theory and Applications’, Birkhäuser.
- Zhang, Y. & Duffield, N. (2001), On the constancy of internet path properties, in ‘IMW ’01: Proceedings of the 1st ACM SIGCOMM Workshop on Internet Measurement’, ACM, New York, NY, USA, pp. 197–211.
- Zhang, Y., Paxson, V. & Shenker, S. (2000), The stationarity of internet path properties: Routing, loss, and throughput, Technical report, ACIRI.

## Agradecimientos

A Gonzalo Perera, Pablo Belzarena, Karine Bertin y Paola Bermolen. A mi familia, a mis amigos y a mis compañeros de trabajo.