

**ANÁLISIS NO PARAMÉTRICO PARA LA PREDICCIÓN DE  
DATOS FUNCIONALES**

Por

Daiver De Jesús Vélez Ramos

Tesis sometida en cumplimiento parcial de los requerimientos para el grado de

**MAESTRÍA EN CIENCIAS**

en

**MATEMÁTICAS(ESTADÍSTICAS)**

**UNIVERSIDAD DE PUERTO RICO  
RECINTO UNIVERSITARIO DE MAYAGÜEZ**

2013

Aprobada por:

---

Wolfgang Rolke, Ph.D  
Miembro, Comité Graduado

---

Fecha

---

Edgardo Lorenzo González, Ph.D  
Miembro, Comité Graduado

---

Fecha

---

Edgar Acuña Fernández, Ph.D  
Presidente, Comité Graduado

---

Fecha

---

Rafael A. Rodríguez Solís, Ph.D  
Representante de Estudios Graduados

---

Fecha

---

Omar Colón Reyes, Ph.D  
Director del Departamento

---

Fecha

Resumen de Disertación Presentado a Escuela Graduada  
de la Universidad de Puerto Rico como requisito parcial de los  
Requerimientos para el grado de Maestría en Ciencias

## **ANÁLISIS NO PARAMÉTRICO PARA LA PREDICCIÓN DE DATOS FUNCIONALES**

Por

Daiver De Jesús Vélez Ramos

2013

Consejero: Edgar Acuña Fernández Ph.D

Departamento: Ciencias Matemáticas

Los datos de alta dimensión y funcionales están ganando importancia en problemas de predicción debido a los avances técnicos que permiten su captura y tratamiento. Este tipo de datos aparecen frecuentemente en la Medicina o Quimiometría. Todo esto hace que su estudio esté en auge. Tradicionalmente, la solución de estos problemas de predicción se ha llevado a cabo usando modelos paramétricos de regresión. Mientras, los tradicionales métodos matemáticos estudian las funciones como objetos matemáticos, teniendo en cuenta sus propiedades y fundamentos teóricos en un sentido paramétrico, Ferraty y Vieu [7] estudian toda esta teoría en un contexto no paramétrico, creando sus propios modelos no paramétricos y extendiendo los métodos tradicionales a un sentido no paramétrico. Así, es importante abordar una amplia revisión bibliográfica del (Análisis de Datos Funcionales)ADF hasta la actualidad y profundizar en el estudio de estos distintos enfoques viendo sus semejanzas, diferencias y vislumbrando posibles interacciones. En este último punto se pondrá especial atención a la utilización de métodos no paramétricos con datos funcionales, todo ello en el marco del problema de predicción.

Abstract of Dissertation Presented to the Graduate School  
of the University of Puerto Rico in Partial Fulfillment of the  
Requirements for the Degree of Master of Sciences

**NONPARAMETRIC ANALYSIS FOR PREDICTION OF  
FUNCTIONAL DATA**

By

Daiver De Jesús Vélez Ramos

UNDEFINED MONTH 2013

Chair: Edgar Acuña Fernández Ph.D  
Major Department: Mathematical Sciences

Data from high-dimensional and functional are becoming more important in prediction's problem nowadays, due to technical advances that allow their gathering and handling. This type of data appears frequently in Medicine and Chemometrics. For this reason its study is increasing. Traditionally, the solution of these prediction problem was carried out using parametric regression models. Whereas, the traditional mathematical methods study functions as mathematical objects, taking into account their properties and theoretical foundations on a parametric sense, Ferraty and Vieu [7] studied all this theory in a non-parametric context, creating their own non-parametric models and extending traditional methods but in a non-parametric way. It is of interest to make a broad literature review of (Functional Data Analysis) FDA until the present and to study further these different approaches seeing their similarities and differences and envisioning possible interactions. In this last part we will put special attention to the use of non-parametric methods with functional data in the context of the prediction problem.

Copyright © 2013

por

Daiver De Jesús Vélez Ramos

A mis padres, hermanos, novia, familiares, y amigos...

Cimiento de mi vida.

## AGRADECIMIENTOS

Al Dr. Edgar Acuña por su gran ayuda, interés y sugerencias en el desarrollo del presente documento.

Al departamento de Ciencias Matemáticas del Recinto de Mayagüez por darme esta oportunidad.

A los profesores del recinto por brindarme su conocimiento.

A los compañeros del programa Roberto, Fabian, Einstein, Carlos, Gustavo, Sindy y Shirley por ofrecerme su apoyo.

A mis amigos Charlie, Lucho, Roxana.

## Índice general

	<u>página</u>
RESUMEN EN ESPAÑOL . . . . .	II
ABSTRACT ENGLISH . . . . .	III
AGRADECIMIENTOS . . . . .	VI
Índice de figuras . . . . .	IX
LISTA DE SÍMBOLOS Y ABREVIACIONES . . . . .	X
1. INTRODUCCIÓN . . . . .	1
1.1. Objetivos . . . . .	2
1.2. Resumen de los Capítulos . . . . .	3
2. ASPECTOS TEÓRICOS . . . . .	4
2.1. Definiciones básicas . . . . .	4
2.2. Estadísticos no paramétricos para datos funcionales . . . . .	7
2.3. Datos funcionales Quimiométricos . . . . .	11
2.3.1. Descripción de datos espectrométricos . . . . .	12
2.3.2. Primeros estudios y problemática . . . . .	13
2.4. Espacio bien adaptado para datos funcionales . . . . .	14
2.4.1. Nociones de cercanía . . . . .	14
2.4.2. La maldición de dimensionalidad . . . . .	16
2.5. Semi-métricas en la práctica . . . . .	20
2.5.1. Semi-métrica basada en FPCA . . . . .	20
2.5.2. Semi-métrica basada en PLS . . . . .	22
2.5.3. Semi-métrica basada en las derivadas . . . . .	23
3. PONDERACIÓN LOCAL KERNEL PARA VARIABLES FUNCIONALES . . . . .	27
3.1. Estimación de Densidades . . . . .	28
3.1.1. Histograma . . . . .	28
3.1.2. Histograma Móvil . . . . .	29
3.1.3. Estimación por Kernel . . . . .	31
3.2. Ponderación Local y Probabilidad en Vecindades . . . . .	36
3.3. Métodos y Modelos de Predicción . . . . .	40
3.3.1. Métodos de Predicción . . . . .	40
3.3.2. Modelos no Paramétricos . . . . .	43

4.	RESULTADOS Y VALIDACIÓN . . . . .	47
4.1.	Estimadores de los Modelos No Paramétricos . . . . .	47
4.1.1.	Estimador del operador regresión . . . . .	47
4.1.2.	Estimador del operador f.d.a condicional . . . . .	50
4.1.3.	Estimador del operador densidad condicional . . . . .	52
4.2.	Convergencia de los Estimadores . . . . .	53
4.2.1.	Estimador de Regresión . . . . .	53
4.2.2.	Estimador Mediana condicional . . . . .	60
4.2.3.	Estimador Moda condicional . . . . .	65
4.2.4.	Estimador Cuantil condicional . . . . .	71
4.3.	Reimplementación y Aplicación de los Estimadores . . . . .	72
4.3.1.	Estimación por Regresión . . . . .	73
4.3.2.	Estimación de la Mediana Condicional . . . . .	75
4.3.3.	Estimación de la Moda Condicional . . . . .	76
4.3.4.	Prediciendo el contenido de grasa de las curvas de espec- trometría . . . . .	77
5.	CONCLUSIONES Y TRABAJOS FUTUROS . . . . .	81
5.1.	Conclusiones . . . . .	81
5.2.	Trabajos futuros . . . . .	83
	APÉNDICES . . . . .	84
A.	Algunos resultados probabilísticos e implementación . . . . .	85
A.1.	Definiciones y Teoremas Clásicos . . . . .	85
A.2.	Convergencia casi Completa . . . . .	86
A.3.	Desigualdad exponencial para <i>v.a.r</i> independientes. . . . .	92
A.4.	Algunas Rutinas . . . . .	94

## Índice de figuras

Figura	página
2-1. Datos Quimiométricos de 15 piezas . . . . .	12
2-2. Las Curvas de Espectrometría . . . . .	13
2-3. Forma de las derivadas de las curvas de Espectrometría . . . . .	17
2-4. $N(p)$ vs $p$ . . . . .	18
2-5. Funciones Base para $k=0,1,2,3$ . . . . .	24
3-1. Estimación por histograma de la densidad de una variable aleatoria $N(0, 1)$ (100 particiones). . . . .	29
3-2. Estimación por histograma móvil de la densidad de una variable aleatoria $N(0, 1)$ ( $h=0.5$ ) . . . . .	30
3-3. Kernel simétricos usuales . . . . .	32
3-4. Estimación por Kernel gaussiano con $h = 0.05; h = 0.5; h = 5$ de la densidad $N(0, 1)$ . . . . .	33
3-5. Kernel asimétricos usuales . . . . .	34
4-1. Varios ejemplos de Kernel integrados simétricos . . . . .	51
4-2. Rendimiento de los tres métodos de predicción . . . . .	79
4-3. Comparando los tres métodos de predicción y el método multimétodo . . . . .	80

## LISTA DE SÍMBOLOS Y ABREVIACIONES

$v.a.r$	Variable aleatoria real.
$(a, b)$	Intervalo abierto de $\mathbb{R}$ .
$[a, b]$	Intervalo cerrado de $\mathbb{R}$ .
$E$ o $F$	Espacios funcionales genéricos.
$d(., .)$	Semi-métrica en algún espacio funcional $E$ .
$(E, d)$	Espacio funcional genérico y su Semi-métrica.
$v.a.f$	Variable aleatoria funcional.
$a.co.$	Convergencia casi completa para sucesiones de $v.a.r$ o $v.a.f$ .
$B(\chi, h)$	Bola abierta con centro en $\chi$ y radio $h$ , en el espacio $(E, d)$ .
$f.d.a$	Función de distribución acumulativa.
$C$ o $C'$	Constantes positivas reales finitas genéricas.
$C_G^0$	Conjunto de las funciones continuas de valor real en $G$ .
$\mathbb{E}(Y)$ o $\mathbb{E}Y$	Esperanza para alguna $v.a.r$ $Y$ .
$\mathbb{E}(Y \mid \mathcal{X} = \chi)$	Espranza condicional de alguna $v.a.r$ $Y$ dada la $v.a.f$ $\mathcal{X}$ .
$f$	Densidad marginal de la $v.a.f$ $\mathcal{X}$ .
$F_Y^{\mathcal{X}}(\chi, y)$ o $F_Y^{\mathcal{X}}(y)$	Distribución condicional de alguna $v.a.r$ $Y$ dada la $v.a.f$ $\mathcal{X}$ .
$f_Y^{\mathcal{X}}(\chi, y)$ o $f_Y^{\mathcal{X}}(y)$	Densidad condicional de alguna $v.a.r$ $Y$ dada la $v.a.f$ $\mathcal{X}$ .
$\phi_{\chi}(h)$	Medida de $B(\chi, h)$ respecto a la ley de probabilidad de $\mathcal{X}$ .
$1_I(\cdot)$	Función indicador en algún conjunto $I$ .
$h$	Notación genérica para un ancho de banda $h = h(n)$ .

$H$	Notación genérica para el integrado de la función kernel.
$K$	Notación genérica para la función kernel asimétrica.
$K_0$	Notación genérica para una función kernel simétrica estandar.
$O_{a.co.}$	Razón de convergencia casi completa.
$O_{a.s.}$	Razón de convergencia casi segura.
$O_p$	Razón de convergencia en probabilidad.
$(\Omega, \mathcal{A}, P)$	Espacio de probabilidad donde las <i>v.a.f</i> están definidas.
$P(  \mathcal{X} = \chi)$	Probabilidad condicional (en la <i>v.a.f</i> $\chi$ ) medido en $\Omega$ .
$r$	Operador de regresión no lineal.
$S$	Subconjunto compacto de $\mathbb{R}$ genérico.
$\mathcal{S}$	Muestra funcional estadística $\mathcal{S} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$ .
$\mathcal{X}$	Variable aleatoria funcional genérica.
$\chi$	Elemento funcional (no aleatorio) genérico.
$\mathcal{X}_i, i = 1, \dots, n$	Muestra de <i>v.a.f</i> .
$\chi_i, i = 1, \dots, n$	Observaciones estadísticas de la <i>v.a.f</i> $\mathcal{X}_i$ .
$\mathbb{R}_*$	Conjunto de los números reales sin el 0.
$\mathbb{R}^+, \mathbb{R}_*^+$	Reales positivos con (respectivamente sin) 0.

# Capítulo 1

## INTRODUCCIÓN

Existe actualmente un número creciente de situaciones provenientes de diferentes campos de las ciencias aplicadas (Quimiometría, Biometría, Medicina, Econometría, etc) en la que los datos recolectados son curvas. De hecho, el progreso de las herramientas computacionales, tanto en términos de memoria como capacidad de cálculos, nos permiten hacer frente a grandes conjuntos de datos. En particular, para un único fenómeno, se puede observar un conjunto muy grande de variables. Por ejemplo, es común observar el comportamiento de una variable aleatoria en varios momentos distintos en el intervalo de tiempo  $(t_{min}, t_{max})$ .

Un problema estadístico consiste en estudiar la relación entre dos variables con el fin de predecir una de ellas (variable de respuesta), dado un nuevo valor (variable explicativa). Este problema ha sido estudiado ampliamente para una o más variables predictoras en el espacio, pero escasamente con las variables predictoras funcionales. Esta tesis está dedicada a este problema cuando la variable de respuesta es real y la variable explicativa es funcional.

Hay varias maneras de abordar el problema de predicción y uno de los más populares es sin duda el método de regresión que se basa en la esperanza condicional. Para los propósitos de robustez (de acuerdo con el comportamiento de la distribución condicional) hay dos técnicas alternativas, moda condicional y mediana condicional. Estos tres métodos predictivos han sido ampliamente estudiados en

situaciones reales multivariadas, y nuestro objetivo es de adaptarlas a la situación en el que las variables explicativas son de dimensión infinita. En el marco funcional, cada uno de estos tres enfoques de predicción funcional conduce a la estimación de algún operador (no lineal), la esperanza condicional (por regresión funcional), la función de distribución acumulativa condicional (para la mediana condicional) y la función de densidad (para la moda condicional). Para evitar el uso de modelos con supuestos demasiado restrictivos, analizaremos estos problemas por medio de modelos no paramétricos para estos operadores (no lineales). Además, queremos resaltar que es esta tesis se da a conocer un nuevo modo de convergencia (convergencia casi completa) que se demostro es mejor en términos matemáticos que los ya existentes en la literatura. Este modo de convergencia nos permitió demostrar que nuestros métodos de predicción son consistentes.

### 1.1. Objetivos

Esta tesis tiene como propósito general desarrollar y dar a conocer una forma tradicional de predecir en el contexto de datos funcionales.

#### **Objetivos específicos**

- Comparar tres métodos de predicción: Esperanza condicional, Mediana condicional y Moda condicional.
- Proponer un estudio sistemático de cada método no paramétrico funcional mediante el desarrollo de aspectos teóricos y prácticos.
- Promover el uso de estos tres métodos de predicción y su importancia en el campo de las ciencias aplicadas: Medicina, Biometría, etc.
- Aplicar los métodos a conjuntos de datos reales.

## 1.2. Resumen de los Capítulos

En este trabajo de tesis se desarrollarán técnicas generales de predicción en el marco del análisis de datos funcionales.

En el capítulo 2, se darán las definiciones básicas de datos funcionales y modelos no paramétricos.

En el capítulo 3, se definirá la ponderación local para datos funcionales, técnicas muy populares en la comunidad no paramétrica, porque están muy bien adaptadas a sus modelos. También se hará una descripción de los métodos y modelos de predicción.

En el capítulo 4, se describirán los estimadores de los tres operadores funcionales no lineales asociados a los modelos de predicción. Además, se dará soporte matemático dando resultados de consistencia para cada estimador no paramétrico funcional. Por otra parte, se harán aplicaciones usando el conjunto de datos de espectrometría que se detalla en el capítulo 2. Para cada método de predicción se darán rutinas en R.

Por último, en el capítulo 5, se darán conclusiones y trabajos futuros

## Capítulo 2

# ASPECTOS TEÓRICOS

El objetivo principal de este capítulo es dar conceptos estadísticos tanto funcionales como no paramétrico. En primer lugar y debido a la novedad de este campo de investigación, se darán algunas definiciones básicas con el fin de aclarar la terminología en datos funcionales /variables y modelos no paramétricos. Para ello se ha tomado como referencia el trabajo de Ferraty y Vieu [7]. Algunos de estos conceptos son muy básicos pero se ha considerado oportuno introducirlos por razones de completitud.

### 2.1. Definiciones básicas

**Definición 2.1.** *Una variable aleatoria  $\mathcal{X}$  se dice que es una variable aleatoria funcional (v.a.f) si esta toma valores en un espacio de dimensión infinita (o espacio funcional). Una observación  $\chi$  de  $\mathcal{X}$  se llama dato funcional, el cual puede ser una curva, superficie, o cualquier otro objeto matemático infinito dimensional.*

Cuando  $\mathcal{X}$  (resp.  $\chi$ ) denota una curva aleatoria, implícitamente se tomará la siguiente identificación  $\mathcal{X} = \{\mathcal{X}(t); t \in T\}$ . En esta situación la característica funcional viene directamente de las observaciones. La situación cuando la variable es una curva está asociada con la unidimensionalidad de  $T \subset \mathbb{R}$ . Aquí, es importante señalar que la noción de variable funcional cubre un área más grande que el análisis de curvas. En particular, una variable funcional puede ser una superficie aleatoria, como por ejemplo los niveles gris de una imagen o un vector de curvas (y en estos

casos  $T$  es un conjunto bidimensional  $T \subset \mathbb{R}^2$ ), o cualquier otro objeto matemático más complicado infinito dimensional.

**Definición 2.2.** *Una muestra o conjunto de datos funcional  $\chi_1, \dots, \chi_n$  de dimensión  $n$  es el conjunto de las observaciones de  $n$  variables funcionales  $\mathcal{X}_1, \dots, \mathcal{X}_n$  idénticamente distribuidas como  $\mathcal{X}$ .*

A partir de ahora, se asumirá que se tiene una muestra de datos funcionales, además todas las variables aleatorias funcionales consideradas se supondrán definidas en un mismo espacio de probabilidad  $(\Omega, \mathcal{A}, P)$ .

Es bien conocido en el análisis de varianza y en la construcción de modelos de regresión lineal, hacer uso de estadísticos tales como la media muestral, varianza muestral, mínimo, máximo, etc. En la siguiente sección se definirán estadísticos no paramétrico para datos funcionales. Antes se darán algunas definiciones para muestras de datos funcionales.

Sea  $\mathcal{X}$  una variable aleatoria funcional valorada en un espacio semi-métrico (ver sección 2.4) infinito dimensional  $(E, d)$ , sea  $\mathcal{S} = \{\mathcal{X}_1, \dots, \mathcal{X}_n\}$  una muestra de  $n$  variables funcionales idénticamente distribuidas como  $\mathcal{X}$ , sea  $\chi$  un elemento fijo de  $E$  y sea  $\chi_1, \dots, \chi_n$  una observación de  $\mathcal{S}$ . Se comenzará describiendo estadísticos muy conocidos como la media, la mediana y la moda, pero para la distribución de una variable funcional  $\mathcal{X}$ .

**Definición 2.3. Media**

La media de  $\mathcal{X}$  se define por:

$$\mathbb{E}(\mathcal{X}) = \int_{\Omega} \mathcal{X}(w) dP(w),$$

Un universal y siempre computable estimador de la media es su versión empírica, el cual llamaremos curva media:

$$\forall t \in \mathbb{R}, \bar{\mathcal{X}}(t) = \frac{1}{n} \sum_{i=1}^n \mathcal{X}_i(t).$$

Otra manera de definir la media es considerarla como solución (bajo existencia y unicidad) al problema de minimización:

$$\inf_{\chi \in E} \mathbb{E}(d(\chi, \mathcal{X})^2),$$

siendo  $d$  una semi-métrica. Similarmente, podemos definir la mediana de  $\mathcal{X}$

**Definición 2.4. Mediana**

Una solución (bajo suposición de existencia) del siguiente problema de minimización:

$$\inf_{\chi \in E} \mathbb{E}(d(\chi, \mathcal{X}))$$

será llamado mediana funcional asociado a la semi-métrica  $d$ .

Un estimador empírico para la mediana funcional es dado por:

$$\mathcal{X}_{med} = \inf_{\chi \in \mathcal{S}} \sum_{i=1}^n d(\chi, \mathcal{X}_i).$$

Por otra parte, en el contexto de dimensión finita, la moda es muy popular en la clasificación, ya que es una herramienta útil para representar grupos y también más robusta que la media (al igual que la mediana, la moda es menos sensible a los valores atípicos que la media). Formalmente, una moda de la distribución de probabilidad

de  $\mathcal{X}$  es un valor que maximiza localmente su densidad  $f$ . Por lo tanto, se propone la siguiente definición.

**Definición 2.5. Moda**

*Una solución (bajo suposición de existencia) del siguiente problema de maximización:*

$$\sup_{\chi \in \mathcal{D}} f(\chi),$$

*donde  $\mathcal{D}$  es un subconjunto no vacío de  $E$ , será llamado moda funcional (implícitamente asociada a la semi-métrica  $d$ ).*

**Definición 2.6.** *Definimos la función Varianza funcional como:*

$$\forall t \in \mathbb{R}, \text{Var}_{\mathcal{X}(t)} = \frac{1}{n-1} \sum_{i=1}^n (\mathcal{X}_i(t) - \bar{\mathcal{X}}(t))^2$$

*y definimos la desviación estándar funcional como:*

$$\forall t \in \mathbb{R}, \text{Stdesv}_{\mathcal{X}(t)} = \sqrt{\text{Var}_{\mathcal{X}(t)}}.$$

De la misma manera que la media funcional nos indica la tendencia central de las funciones en un  $t$  dado, la varianza funcional y la desviación estándar funcional nos cuantifican el valor medio al cuadrado y el valor medio respectivamente de las desviaciones respecto a la media en  $t$ .

## 2.2. Estadísticos no paramétricos para datos funcionales

Los métodos estadísticos tradicionales fallan en cuanto nos ocupamos de datos funcionales [7]. Por ejemplo, si consideramos una muestra de curvas finamente discretizadas, surgen dos problemas estadísticos cruciales. El primero viene de la relación entre el tamaño de la muestra y el número de variables (cada variable real corresponde a un punto de discretización). El segundo se debe a la existencia de fuertes correlaciones entre las variables y se convierte en un problema mal condicionado en

el contexto de un modelo lineal multivariado. Por lo tanto, existe la necesidad real de desarrollar métodos/modelos estadísticos a fin de tener en cuenta la estructura funcional de este tipo de datos. La mayoría de los actuales métodos estadísticos que tratan con datos funcionales utilizan modelos lineales para el objeto que se estima.

Hay muchas maneras de definir un modelo estadístico no paramétrico en el contexto de dimensión finita, y la distinción entre paramétrico y no paramétrico puede ser poco clara (ver la introducción de [5] para más detalles). La siguiente es la definición de un modelo no paramétrico en el contexto de dimensión finita.

**Definición 2.7.** *Sea  $X$  un vector aleatorio en  $\mathbb{R}^p$  y sea  $\phi$  una función definida en  $\mathbb{R}^p$  y en función de la distribución de  $X$ . Un modelo para la estimación de  $\phi$  consiste en introducir alguna restricción de la forma:*

$$\phi \in \mathcal{C}$$

*El modelo es llamado un modelo paramétrico para estimación de  $\phi$ , si  $\mathcal{C}$  es indexado por un número finito de elementos de  $\mathbb{R}$ . Caso contrario, se dice que el modelo es no paramétrico.*

Esta definición hace clara la distinción que hay entre un modelo paramétrico y no paramétrico, además puede ser extendida a la estructura funcional.

**Definición 2.8.** Sea  $\mathcal{X}$  una variable aleatoria funcional valorada en algún espacio de dimensión infinita  $E$  y sea  $\phi$  un mapa definido en  $E$  y en función de la distribución de  $\mathcal{X}$ . Un modelo para la estimación de  $\phi$  consiste en introducir alguna restricción de la forma:

$$\phi \in \mathcal{C}$$

El modelo es llamado un modelo paramétrico para la estimación de  $\phi$ , si  $\mathcal{C}$  es indexado por un número finito de elementos de  $E$ . Caso contrario, se dice que el modelo es no paramétrico.

La denominación **estadístico no paramétrico funcional** abarca todos los antecedentes estadísticos que involucran un modelo funcional no paramétrico. En la terminología **estadístico no paramétrico funcional** el adjetivo **no paramétrico** se refiere a la forma del conjunto de restricción, mientras la palabra **funcional** está relacionada con la naturaleza de los datos.

Para ilustrar el propósito sobre estos aspectos de modelado, se dará una idea en un modelo de regresión

$$Y = \phi(X) + error \tag{2.1}$$

donde  $Y$  es una variable aleatoria real pero condicionada a varias situaciones: Modelos de regresión lineal (paramétrico) o no paramétrico con curvas (*i.e.*  $\mathcal{X} = X = X(t); t \in (0, 1)$ ) o datos multivariados (*i.e.*  $\mathbf{X} = X = (X_1, \dots, X_p)$ ):

**Ejemplo 2.2.1. Paramétrico multivariado**

Corresponde al llamado modelo de regresión lineal multivariado

$$\begin{aligned} \phi : \mathbb{R}^p &\longrightarrow \mathbb{R} \\ X &\longrightarrow \phi(X) = a_0 + \sum_{j=1}^p a_j X_j \\ \text{entonces } Y &= a_0 + \sum_{j=1}^p a_j X_j + \text{error} \end{aligned}$$

donde  $\phi \in \mathcal{C} = \{r : \mathbb{R}^p \longrightarrow \mathbb{R} / r \text{ es lineal}\}$ .

El cual es un modelo paramétrico, con  $p+1$  parámetros desconocidos  $a_0, a_1, \dots, a_p \in \mathbb{R}$ .

**Ejemplo 2.2.2. No paramétrico multivariado**

$$\begin{aligned} \phi : \mathbb{R}^p &\longrightarrow \mathbb{R} \\ X &\longrightarrow \phi(X) \in C_{\mathbb{R}^p}^0 \\ \text{entonces } Y &= \phi(X) + \text{error} \end{aligned}$$

donde  $\phi \in \mathcal{C} = \{r : \mathbb{R}^p \longrightarrow \mathbb{R} / r \text{ es continua}\}$ .

**Ejemplo 2.2.3. Paramétrico funcional**

$$\begin{aligned} \phi : L_{(0,1)}^p &\longrightarrow \mathbb{R} \\ \chi &\longrightarrow \phi(\chi) = \int_0^1 \chi(t)\rho(t) dt, \quad \rho \in L_{(0,1)}^p \\ \text{entonces } Y &= \int_0^1 \chi(t)\rho(t) dt + \text{error} \end{aligned}$$

donde  $\phi \in \mathcal{C} = \{\chi \longmapsto \int_0^1 \chi(t)\rho(t) dt / \rho \in L_{(0,1)}^p\}$ .

Recuerde que para nuestra definición, este es un modelo de regresión paramétrico funcional, donde  $\rho(\cdot)$  es el único parámetro (funcional).

*Ejemplo 2.2.4. No paramétrico funcional*

$$\begin{aligned} \phi : L^p_{(0,1)} &\longrightarrow \mathbb{R} \\ \chi &\longrightarrow \phi(\chi) \\ \text{entonces } Y &= \phi(\chi) + \text{error} \end{aligned}$$

donde  $\phi \in \mathcal{C} = \{r/r \text{ es un operador continuo}\}$ .

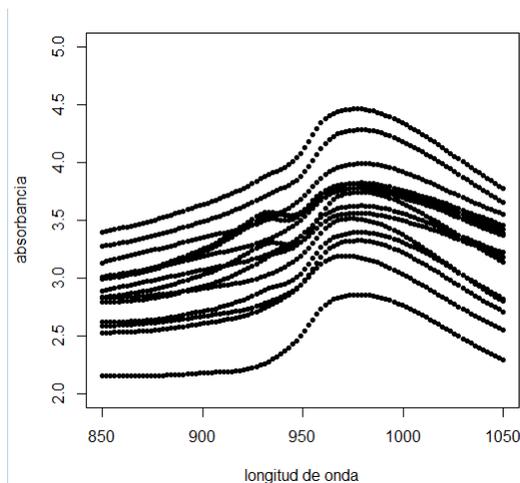
Este último ejemplo es una ilustración del modelo no paramétrico que se desarrollará en el presente documento, pero por ahora veamos un ejemplo de datos funcionales.

### 2.3. Datos funcionales Quimiométricos

La espectrometría es una herramienta moderna y de completo uso para analizar composiciones químicas de cualquier sustancia, como es señalado por [13]. A fin de analizar este tipo de datos "los quimiométricos han inventado sus propias técnicas basadas en razonamientos heurísticos e ideas intuitivas". Los dos métodos más populares son mínimos cuadrados parciales (ver [12] y [11]) y regresión de componentes principales ([20]). Como la Quimiometría fue un punto de partida para el desarrollo de la metodología no paramétrica funcional, desempeñan un papel importante, por eso se ha decidido analizar esta teoría usando estos datos, y es natural comenzar la presentación de ese conjunto de datos.

### 2.3.1. Descripción de datos espectrométricos

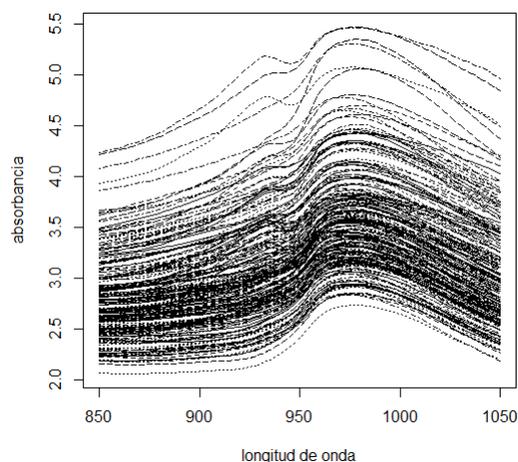
Los datos originales provienen de un problema de control de calidad en la industria alimentaria y pueden encontrarse en <http://lib.stat.cmu.edu/datasets/tecator>. Estos datos fueron estudiados primero por [1] usando un enfoque de redes neuronales. Los datos se refieren a una muestra de carne picada (piezas o trozos). La Figura 2–1 muestra algunas unidades entre los datos espectrométricos originales.



**Figura 2–1:** Datos Quimiométricos de 15 piezas

Esta es la gráfica de (absorbancia vs longitud de onda) para 15 trozos de carne seleccionados al azar. Más precisamente, para cada muestra de carne los datos consisten de 100 canales de absorbancia. La absorbancia es el  $-\log_{10}$  de la transmitancia medida por el espectrómetro y los datos se registraron en el Analizador Tecator Infractec Food trabajando en un rango de longitud de onda de  $(850 - 1050) \text{ nm}$  por el principio de la transmisión en el infrarrojo cercano. Una unidad aparece claramente como una curva de datos discretos. Debido a la finura de la malla, podemos considerar cada sujeto como una curva continua. Esto fue señalado por Silverman en [18] quién dijo que los espectros observados son a todos los efectos de las observaciones funcionales. Así, los análisis espectrométricos pueden resumirse por alguna

curva continua dando la absorbancia observada como función de la longitud de onda. La muestra de curvas, que es el conjunto de datos (continuos), se presenta en la Figura 2-2 a continuación.



**Figura 2-2:** Las Curvas de Espectrometría

Como podemos ver en la figura, la forma de estas curvas es muy suave y muy similar, excepto por un evidente desplazamiento vertical.

### 2.3.2. Primeros estudios y problemática

Teniendo en cuenta estos datos, pueden aparecer muchas preguntas. En particular, la primera sería saber si el desplazamiento vertical que se observa en la Figura 2-2 es realmente informativo o no. En otras palabras, si el desplazamiento vertical se debe a los componentes químicos de la carne o es solo debido a algún fenómeno adicional. Esta pregunta es fundamental porque el desplazamiento puede ocultar otras características importantes y actuar como algo ficticio. Otra pregunta viene del hecho de que el objetivo principal del análisis espectrométrico es permitir el descubrimiento de la proporción de algún contenido químico específico. Por ejemplo, si  $Y$  es la proporción de algún componente (para el conjunto de datos quimiométricos, se puede tomar  $Y$  como el porcentaje del contenido de grasa en la pieza de carne), nos gustaría utilizar las curvas de espectrometría para predecir  $Y$ .

Esta última observación será a lo que nos dedicaremos en adelante, tomando como ejemplo las curvas de espectrometría.

## 2.4. Espacio bien adaptado para datos funcionales

Utilizando datos funcionales surgen preguntas importantes estadísticas. Como por ejemplo, el espacio  $E$  en el que la variable toma sus valores. En el caso de datos funcionales, sabemos (por la naturaleza misma de los datos) que  $E$  es un espacio de dimensiones infinita. Por lo tanto, esta sección se centra en este problema esencial de alta (i.e., infinito) dimensionalidad de datos. Evidentemente, la noción de poca densidad está fuertemente relacionada con la forma utilizada para medir la cercanía entre los datos y se mostrará una original forma de abordar esta pregunta por medio de consideración de semi-métrica.

### 2.4.1. Nociones de cercanía

Medidas de cercanía entre objetos matemáticos desempeñan un papel importante en todos los métodos estadísticos. En muchas situaciones, puede utilizarse una norma clásica para medir la cercanía entre dos objetos. Debido a que en un espacio euclidiano finito dimensional (normalmente  $\mathbb{R}$ ) existe una equivalencia entre todas las normas. La elección en el sentido matemático de este tipo de medida no es crucial aparte de algunas limitaciones prácticas (como por ejemplo, facilidad computacional). Una vez que se fija una norma preliminar, está claro que podemos deducir una familia de normas y desde un punto de vista estadístico, queda una pregunta esencial: a saber, la elección entre estas diferentes métricas. Por ejemplo, una de las más popular en  $\mathbb{R}^p$  es la norma habitual euclidiana  $\| \cdot \|$  que se basa en la suma de los cuadrados de los componentes de cualquier vector. Mas precisamente, sea  $\mathbf{x} = \langle x_1, \dots, x_p \rangle$  un vector de  $\mathbb{R}^p$ ; a continuación, se define la norma euclidiana clásica por:

$$\|x\|^2 = \sum_{i=1}^p (x_i)^2 = \mathbf{x} \cdot \mathbf{x}^t.$$

Por supuesto, podemos deducir una familia de normas basado en la norma euclidiana mediante el uso de diferentes matrices  $\mathbf{M}$  definidas positivas, de la siguiente manera

$$\|x\|_{\mathbf{M}}^2 = \mathbf{x} \mathbf{M} \mathbf{x}^t.$$

La elección de la norma radica en la elección de  $\mathbf{M}$ . Ahora, considerando un espacio de dimensión infinita, falla la equivalencia entre las normas y el problema tiene que ser atacado de una manera diferente. En otras palabras, en el contexto funcional, la elección de la norma preliminar se vuelve crucial. Aún más, teniendo en cuenta los espacios métricos o normados puede convertirse en demasiado restrictiva. En algunas situaciones, (y este es el caso de nuestro conjunto de datos) parece que los espacios semi-métricos se adaptan mejor que los espacios métricos. Como se verá más adelante en la sección 2.5, la forma de los datos e información eventualmente exógena o el objetivo del estudio estadístico pueden ayudar a la elección de una semi-métrica. En lo que sigue, se muestra el beneficio de considerar semi-métricas como una medida de cercanía. Antes, se recordarán algunas definiciones básicas.

**Definición 2.9.**  $\|\cdot\|$  es una semi-norma en algún espacio  $E$  si y solo si:

1.  $\forall (\lambda, x) \in \mathbb{R} \times E, \|\lambda x\| = |\lambda| \|x\|$
2.  $\forall (y, x) \in E \times E, \|x + y\| \leq \|x\| + \|y\|$ .

**Definición 2.10.**  $d$  es una semi-métrica en algún espacio  $E$  si y solo si:

1.  $\forall x \in E, d(x, x) = 0,$
2.  $\forall (y, x, z) \in E \times E \times E, d(x, y) \leq d(x, z) + d(z, y).$

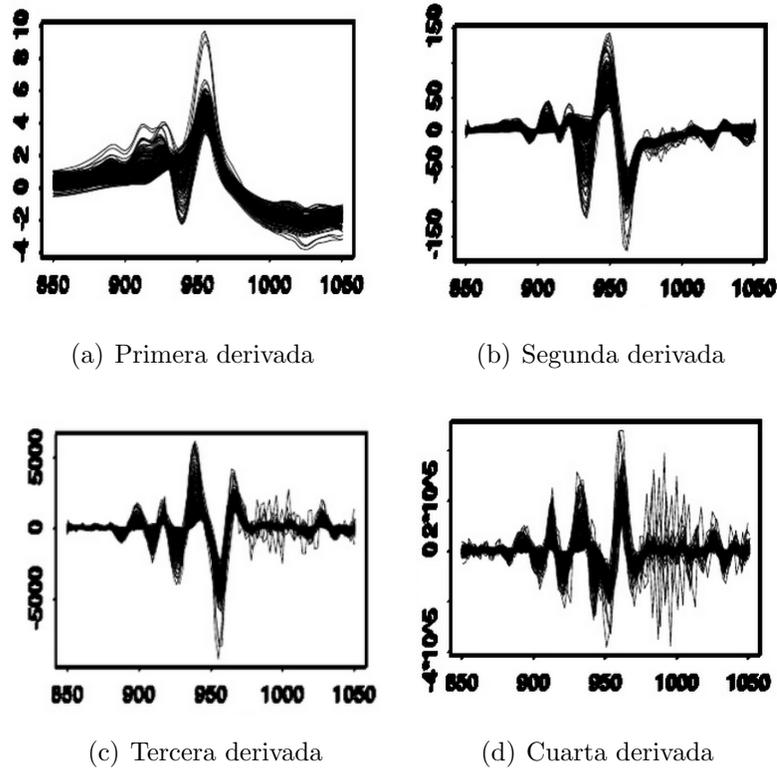
Es evidente, que las herramientas matemáticas para calcular las proximidades entre las curvas desempeñan un papel importante. Por eso la pregunta: ¿Cómo decidir cual es la mejor presentación gráfica sobre las curvas espectrométricas? o equivalentemente: ¿Cuál es el análisis que revela la información más pertinente? Una forma de dar una respuesta razonable es permitir una elección dentro de una gran familia de semi-métricas. Una familia será construida de acuerdo a cada problema estadístico especificado y el conjunto de datos. Por ejemplo, en el contexto de datos quimiométricos se centrará en la familia de semi-métricas.

$$\sqrt{\int (\chi_i^{(m)}(t) - \chi_{i'}^{(m)}(t))^2 dt} \quad m = 0, 1, 2, \dots$$

Donde, para cualquier función real  $\chi$ ,  $m$  – veces diferenciable,  $\chi^{(m)}$  denota la derivada  $m$  – esima de  $\chi$  (con  $\chi^0 = \chi$ ). Para dar una idea del interés de este enfoque para las curvas espectrométricas, en la Figura 2–3 se muestran sus diferentes derivadas. El principal efecto producido por el operador diferencial es poner de relieve algunos intervalos de longitudes de onda con grandes variaciones. Finalmente, se puede decir que las semi-métricas actúan como un filtro, y una ”buena semi-métrica” *a priori* puede seleccionar toda la información pertinente.

#### 2.4.2. La maldición de dimensionalidad

La maldición de dimensionalidad es un concepto muy conocido por los no-paramétricos. Esta idea está estrechamente vinculada con la escasez de datos en un espacio de alta dimensión. Es una pregunta interesante: ¿Qué pasa con la maldición de dimensionalidad cuando trabajamos con datos funcionales? Si tenemos observaciones en  $\mathbb{R}^p$ , una forma de ilustrar la maldición de dimensionalidad es contar el número  $N(p)$  de unidades que caen en un subconjunto (de tamaño fijo) de  $\mathbb{R}^p$  cuando  $p$  toma valores sucesivos  $(1, 2, \dots)$ . En particular, esta situación corresponde a datos funcionales vistos a través de su versión discretizada. Siguiendo la misma



**Figura 2–3:** Forma de las derivadas de las curvas de Espectrometría

idea, si tenemos  $n$  observaciones funcionales en un espacio semi-métrico  $(E, d)$ , contaremos el número  $N_d$  de unidades que caen en un subconjunto (de tamaño fijo) de  $E$ .

Esto es, sea  $\{\chi_i = \{\chi_i(t); t \in (t_{min}, t_{max})\}\}_{i=1, \dots, n}$  una muestra funcional y consideremos su versión discretizada  $\{\mathbf{x}_i = (\chi_i(t_1), \chi_i(t_2), \dots, \chi_i(t_J))\}_{i=1, \dots, n}$  el cual puede ser vista como una matriz de datos. En la siguiente matriz

$\chi_1(t_1)$	$\chi_1(t_2)$	$\cdots$	$\chi_1(t_J)$
$\chi_2(t_1)$	$\chi_2(t_2)$	$\cdots$	$\chi_2(t_J)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\chi_n(t_1)$	$\chi_n(t_2)$	$\cdots$	$\chi_n(t_J)$

tenemos  $J$  mediciones  $t_1, \dots, t_j$  de  $n$  curvas observadas (por ejemplo, para las curvas espectrométricas,  $J = 100$ ,  $n = 215$ ,  $t_{min} = 850$ ,  $t_{max} = 1050$ ). De este conjunto de datos, debemos extraer una subsecuencia de puntos  $t_{j_1}, \dots, t_{j_p}$  equiespaciados de  $t_1, \dots, t_j$  y considerar sólo los datos correspondientes a esta subsecuencia:

$\chi_1(t_{j_1})$	$\chi_1(t_{j_2})$	$\dots$	$\chi_1(t_{j_p})$
$\chi_2(t_{j_1})$	$\chi_2(t_{j_2})$	$\dots$	$\chi_2(t_{j_p})$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\chi_n(t_{j_1})$	$\chi_n(t_{j_2})$	$\dots$	$\chi_n(t_{j_p})$

Ahora podemos calcular para  $i = 1, \dots, n$  la siguiente métrica euclidiana:

$$\delta_p(\mathbf{x}_i, 0) = \sum_{k=1}^p (\chi(t_{j_k}))^2$$

y definir  $N(p)$  por:

$$N(p) = \sum_{i=1}^n 1 \left\{ \frac{\delta_p(\mathbf{x}_i, 0)}{\max_i \delta_p(\mathbf{x}_i, 0)} < 0,1 \right\}$$

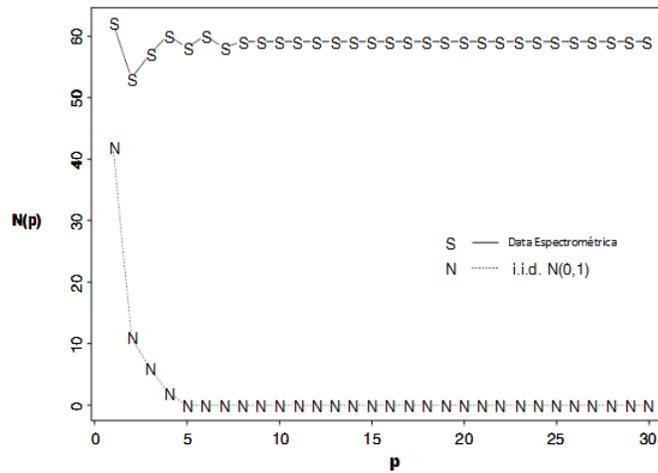


Figura 2-4:  $N(p)$  vs  $p$

La Figura 2-4 muestra  $N(p)$  vs  $p$  para los datos espectrométricos centrados. Además, reemplazamos cada columna de la muestra funcional por  $p$  variables independientes idénticamente distribuidas como una gaussiana estándar y también calculamos las cantidades correspondientes  $N(p)$ . La maldición de la dimensionalidad aparece claramente para los conjuntos de datos gaussianos (los datos son escasos para la cota superior a tres y la poca densidad aumenta exponencialmente con  $p$ ) mientras que los datos espectrométricos no parecen afectados por la dimensión.

Ahora, si tenemos en cuenta la característica funcional de los datos, es decir si reemplazamos  $\mathbf{x}_i = (\chi_i(t_1), \chi_i(t_2), \dots, \chi_i(t_J))$  por  $\chi_i = \{\chi_i(t); t \in (t_{min}, t_{max})\}$ , podemos calcular la cantidad de  $N_d$  definida por:

$$N_d = \sum_{i=1}^n 1 \left\{ \frac{d(\chi_i, 0)}{\max_i d(\chi_i, 0)} < 0,1 \right\},$$

donde  $d$  denota una medida funcional de cercanía. Por ejemplo, si consideramos las curvas espectrométricas con  $d = d_2^{derive}$  como se definirá en la sección 2.5, obtenemos  $N_d = 24$ .

Finalmente, parece que la maldición de dimensionalidad no afecta a los datos funcionales con alta correlación como los datos espectrométricos pero es dramática para los no correlacionados. No obstante, teniendo en cuenta características funcionales, incluso si los datos no están correlacionados, parcialmente cancelamos la maldición de dimensionalidad. Esto lo podemos lograr jugando con las derivadas de las curvas, como se vió en la Figura 2-3, para el caso de las curvas de espectrometría que la correlación aumenta, por supuesto, un desafío crucial sería la elección de la medida de cercanía  $d$  para evitar la maldición de la dimensionalidad, pero esto se sale un poco del objetivo de este trabajo, para más detalles vea [7].

## 2.5. Semi-métricas en la práctica

Debido a que las muestras funcionales más disponibles son curvas, describimos aquí semi-métricas bien adaptadas para este tipo de datos. Consideremos una muestra de curvas  $\mathcal{X}_1, \dots, \mathcal{X}_n$  idénticamente distribuidas como (*v.a.f*)  $\mathcal{X} = \{\mathcal{X}(t); t \in T\}$ . Hay que elegir entre diferentes tipos de semi-métricas la que puede ser adaptada a la forma de la curva (por ejemplo, curvas suaves tendrían que ser tratadas con semi-métricas que no sean ásperas). Aquí, presentamos tres familias de semi-métricas, aunque muchas otras pueden construirse. Las dos primeras están bien adaptadas para las curvas ásperas mientras que la tercera se adapta a datos bastante suaves.

### 2.5.1. Semi-métrica basada en FPCA

En muchas situaciones hablando del contexto multivariado, el análisis de componentes principales (PCA) clásico es considerado como una herramienta útil para mostrar datos en un espacio de dimensiones reducido. Muy recientemente, los métodos PCA fueron extendidos a datos funcionales y utilizados para diferentes fines estadísticos. Aquí, vamos a ver que el FPCA también es una buena herramienta para calcular las aproximaciones entre curvas en un espacio de dimensión reducido.

El FPCA de la *v.a.f*  $\mathcal{X}$  nos permite obtener la siguiente expansión de  $\mathcal{X}$  (ver [14]):

$$\mathcal{X} = \sum_{k=1}^{\infty} \left( \int \mathcal{X}(t)v_k(t) dt \right) v_k,$$

donde  $v_1, v_2, \dots$ , son las funciones propias ortonormales del operador covarianza

$$\Gamma_{\mathcal{X}}(s, t) = E(\mathcal{X}(s)\mathcal{X}(t)),$$

asociadas con los valores propios  $\lambda_1 \geq \lambda_2 \geq \dots$ . Ahora sea

$$\tilde{\mathcal{X}}^{(q)} = \sum_{k=1}^q \left( \int \mathcal{X}(t) v_k(t) dt \right) v_k$$

una versión truncada de la previa expansión de  $\mathcal{X}$ . El principal interés de tal descomposición es que esta versión truncada es minimizada  $E(\int (\mathcal{X}(t) - P_q \mathcal{X}(t))^2 dt)$  sobre todas las proyecciones  $P_q$  de  $\mathcal{X}$  en espacios  $q$ -dimensionales. Por lo tanto, podemos definir una clase parametrizada de semi-normas de la clásica *norma*  $-L_2$  de la siguiente manera:

$$\| \chi \|_q^{PCA} = \sqrt{\int (\tilde{\chi}^{(q)}(t))^2 dt} = \sqrt{\sum_{k=1}^q \left( \int \chi(t) v_k(t) dt \right)^2}.$$

Esto conduce a la familia parametrizada de semi-métricas:

$$d_q^{PCA}(\mathcal{X}_i, \chi) = \sqrt{\sum_{k=1}^q \left( \int [\mathcal{X}_i(t) - \chi(t)] v_k(t) dt \right)^2}.$$

Estas semi-métricas sólo pueden utilizarse con datos medidos en los mismos puntos y tomados de una partición suficientemente fina para que los estimadores empíricos sean consistentes. De hecho nunca se observa exactamente  $\{\chi_i = \{\chi_i(t); t \in T\}\}_{i=1, \dots, n}$  solo una versión discretizada  $\{\mathbf{x}_i = (\chi_i(t_1), \dots, \chi_i(t_J))\}_{i=1, \dots, n}$ . Así, desde un punto de vista práctico, se puede aproximar la integral de la siguiente manera:

$$\int [\mathcal{X}_i(t) - \chi(t)] v_k(t) dt \approx \sum_{j=1}^J w_j (\mathcal{X}_i(t_j) - \chi(t_j)) v_k(t_j),$$

donde  $w_1, \dots, w_J$  son pesos de cuadratura que definen la integración aproximada. Para fijar ideas, tomar en cuenta que una opción estándar podría ser  $w_j = t_j - t_{j-1}$ . Si se toman dos curvas discretizadas  $\mathbf{x}_i$  y  $\mathbf{x}_{i'}$ , la cantidad  $d_q^{PCA}(\chi_i, \chi_{i'})$  se aproxima por su versión empírica:

$$\mathbf{d}_q^{PCA}(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\sum_{k=1}^q \left( \sum_{j=1}^J w_j (\chi_i(t_j) - \chi_{i'}(t_j)) [v_k]_j \right)^2},$$

donde  $\mathbf{v}_1, \mathbf{v}_2, \dots$ , son los vectores propios  $\mathbf{W}$ -ortonormal de la matriz covarianza ( $\mathbf{W} = \text{diag}(w_1, \dots, w_J)$ )

$$\mathbf{\Gamma}^n \mathbf{W} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i^t \mathbf{x}_i \mathbf{W},$$

asociado con los valores propios  $\lambda_{1,n} \geq \lambda_{2,n} \geq \dots$ . Note que por supuesto  $\mathbf{d}_q^{PCA}(\mathbf{x}_i, \mathbf{x}_{i'})$  está cerca de  $d_q^{PCA}(\chi_i, \chi_{i'})$  tan pronto como la partición  $(t_1, \dots, t_J)$  sea lo suficientemente fina.

### 2.5.2. Semi-métrica basada en PLS

El principal objetivo es construir una nueva familia de semi-métricas para situaciones cuando observamos una respuesta adicional, adaptando el enfoque de regresión por mínimos cuadrado parcial multivariante (MPLSR). Al igual que el PCA, las ideas del método PLS pueden ser usadas para diferentes propósitos involucrando datos funcionales (ver por ejemplo, [4], y [3]). En particular, veremos como PLS nos permite construir una familia de semi-métricas.

Sea  $\mathbf{v}_1^q, \dots, \mathbf{v}_p^q$  vectores en  $\mathbb{R}^J$  usados por MPLSR donde  $q$  denota el número de los factores y  $p$  el número de las respuestas escalares. Entonces se define la semi-métrica basado en el MPLSR como sigue:

$$\mathbf{d}_q^{PLS}(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\sum_{k=1}^p \left( \sum_{j=1}^J w_j (\chi_i(t_j) - \chi_{i'}(t_j)) [v_k^q]_j \right)^2},$$

donde los pesos de cuadratura  $w_1, \dots, w_J$  son los descritos anteriormente. Esta semi-métrica es aplicada sólo a datos balanceados.

### 2.5.3. Semi-métrica basada en las derivadas

Otra forma de construir una familia parametrizada de semi-métricas entre curvas es considerar una distancia entre uno de sus derivados. Más precisamente, dados dos curvas observadas  $\chi_i$  y  $\chi_{i'}$ , consideramos las siguientes semi-métricas:

$$d_q^{deriv}(\chi_i, \chi_{i'})^2 = \int \left( \chi_i^{(q)}(t) - \chi_{i'}^{(q)}(t) \right)^2 dt$$

donde  $\chi^{(q)}$  denota la  $q$ -ésima derivada de  $\chi$ . Notar que  $d_0^{deriv}(\chi, 0)$  es la clásica *norma*  $-L^2$  de  $\chi$ . El cálculo de derivadas sucesivas numéricamente es muy sensible. Para anular este problema de estabilidad numérica, se puede utilizar una curva B-spline ([2] o [16]) de aproximación para las curvas. Demos una idea de B-spline.

#### **B-splines**

Método para construir curvas polinomiales por trozos.

#### Ventajas

- El grado de la curva se escoge a priori.
- Tienen carácter local: cada vértice de control va asociado a una única función de base, que tiene soporte local.

#### Parámetros

- Orden la curva está formada por polinomios por trozos de grado  $k$ .
- Vértices de control (puntos del plano o del espacio).
- Vector de nodos (knots)  $[x_0, \dots, x_{k+n+1}]$  tales que  $x_i \in \mathbb{R}$  y  $x_i \leq x_{i+1}$ .

**Definición 2.11.** Dadas el orden  $k$ , los vértices de control  $\beta_0, \dots, \beta_n$ , y el vector de nodos  $[x_0, \dots, x_{n+k+1}]$ , definimos la curva B-spline asociada a estos datos como:

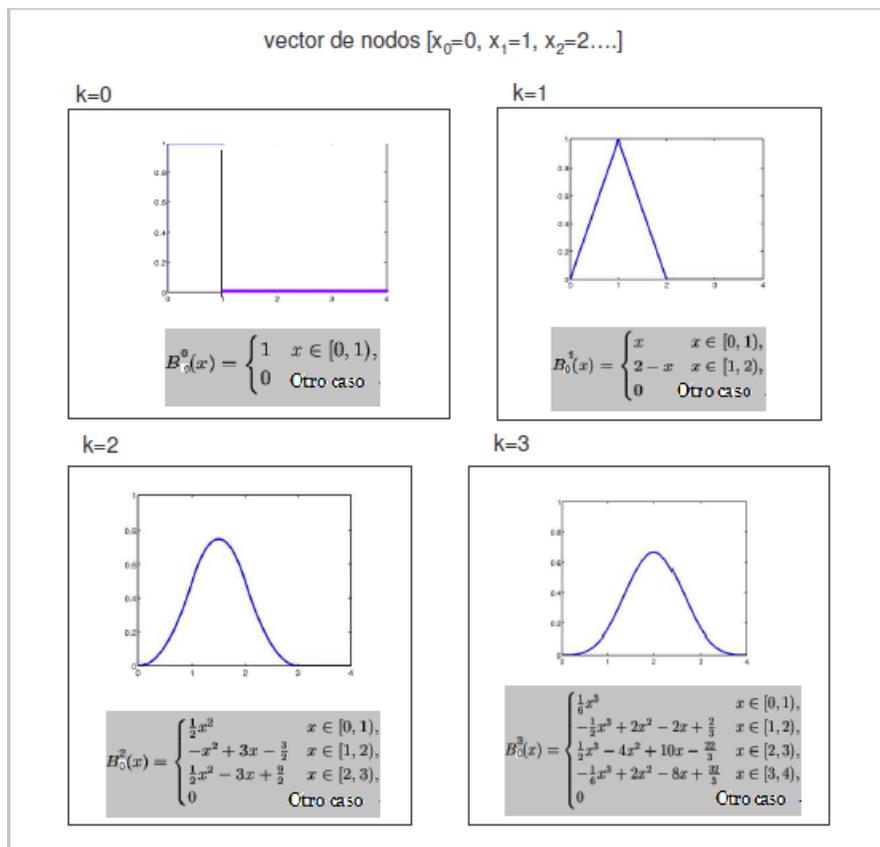
$$P(t) = \sum_{i=0}^n \beta_i B_i^k(t) \quad \text{con } t \in [x_k, x_{n+1}]$$

donde  $B_i^k(t)$  es una función base la cual fue definida por Cox y De Boor como:

$$B_i^0(t) = \begin{cases} 1 & \text{si } t \in [x_i, x_{i+1}] \\ 0 & \text{en otro caso} \end{cases} \quad \text{si } k = 0$$

$$B_i^k(t) = \frac{(t - x_i)B_i^{k-1}(t)}{x_{i+k} - x_i} + \frac{(x_{i+k+1} - t)B_{i+1}^{k-1}(t)}{x_{i+k+1} - x_{i+1}} \quad k \geq 1.$$

*Ejemplo 2.5.1. Algunas funciones bases*



**Figura 2-5:** Funciones Base para  $k=0,1,2,3$

Una vez que hemos obtenido una expansión B-spline analítica para cada curva, las derivadas sucesivas se calculan directamente diferenciando varias veces su forma analítica. Por lo tanto, al tener la siguiente aproximación para  $\mathbf{x}_i = (\chi_i(t_1), \dots, \chi_i(t_J))$  :

$$\widehat{\chi}_i(t) = \sum_{n=1}^J \beta_{in} B_{in}^k(t)$$

Podemos ahora derivar fácilmente las curvas aproximadas:

$$\widehat{\chi}_i^{(q)}(t) = \sum_{n=1}^J \beta_{in} B_{in}^{k(q)}(t)$$

Teniendo en cuenta dos curvas discretizada  $\mathbf{x}_i$  y  $\mathbf{x}_{i'}$ , tenemos que calcular las siguientes cantidades:

$$\mathbf{d}_q^{deriv}(\mathbf{x}_i, \mathbf{x}_{i'}) = \sqrt{\int \underbrace{\left( \widehat{\chi}_i^{(q)}(t) - \widehat{\chi}_{i'}^{(q)}(t) \right)^2}_{f(t)} dt.}$$

Queda por calcular la integral que puede hacerse mediante el método de Gauss. Este método se aplica si se puede evaluar el integrando de  $f$  en cualquier punto, y ese es el caso aquí. De hecho, el método de Gauss propone hacer la siguiente aproximación:

$$\int_a^b f(t) \sim \frac{b-a}{2} \sum_{k=1}^K w_k f\left(\frac{b-a}{2} + \frac{b-a}{2} \delta_k\right)$$

donde los pesos  $w_k$  y los reales  $\delta_k$  ("puntos de gauss") son tabulados. La precisión de este método numérico viene del hecho de que este método es exacto para cualquier polinomio de grado  $< 2K - 1$ .

Por lo tanto, en la práctica, esta clase de semi-métricas estarán bien adaptadas si son utilizados en la presencia de curvas suaves, tales como las espectrométricas descritas anteriormente.

Para concluir este capítulo se puede decir que la modelización de semi-métricas, a pesar de su aspecto teórico más complicado, es una manera razonable para modelar datos funcionales. En esta etapa, se puede decir que el propio conjunto de datos debe ser un elemento prominente para elegir la que será la semi-métrica a ser utilizada. Cada una de las tres familias discutidas anteriormente se adaptan a un tipo especial de datos: Semi-métricas tipo-PCA se espera que den resultados interesantes para los conjuntos de datos en bruto, tipo-PLS son recomendables cuando se tiene a mano una respuesta multivariante, mientras que las tipo-derivada se adaptan a los conjuntos de datos suaves.

## Capítulo 3

# PONDERACIÓN LOCAL KERNEL PARA VARIABLES FUNCIONALES

En muchos temas de estadística se deben enfrentar problemas que incluyen datos aleatorios en forma de curvas. Más precisamente, muestras discretas de curvas. Una primera idea para modelar esto, es tomarlo como un vector cuya dimensión es la cantidad de muestras que se cuenten de la misma. Pero esta no es la mejor solución pues el agregado de nuevas muestras haría aumentar la dimensión (la cual podría llegar a valores que harían el problema intratable). Además, no es natural tratar muestras tan correlacionadas entre sí como dimensiones distintas de un vector. Es así que surge la idea de *variables aleatorias funcionales*. Debido a que únicamente se cuenta con una cantidad finita de muestras de la curva, deberá aplicarse alguna técnica de interpolación. En el caso de dimensión finita, las técnicas de ponderación local son muy populares en la comunidad no paramétrica porque están muy bien adaptadas a los modelos no paramétricos. El objetivo de este capítulo es explicar como el concepto de suavizado local puede extenderse al caso de datos funcionales. Evidentemente, los enfoques locales deben tener a la mano algunas formas topológicas para medir la proximidad entre los datos funcionales, y por lo tanto, este capítulo estará directamente vinculado con el modelado semi-métrico en el capítulo 2. En el caso de dimensión finita, uno de los enfoques más comunes entre estos métodos de ponderación local es sin duda el Kernel. El capítulo está organizado de la siguiente manera. En la sección 3.1 se dará una discusión básica sobre el método Kernel, explicando cómo y por qué lo que clásicamente se hace para variables de dimensión

finita puede ser adaptado al contexto funcional. La sección 3.2 consiste en ver como la ponderación local se relaciona con la noción de las probabilidad de vecindades. Como veremos, las probabilidad de vecindades pueden ser vistas como una herramienta para describir algunos comportamientos locales de datos funcionales y el planteamiento Kernel nos permite tener en cuenta este tipo de propiedades locales, a la vez se propondrán algunos aspectos teóricos generales relativos a la ponderación local Kernel. La sección 3.3 describe los tres métodos de predicción, asociados a la definición 2.8.

### 3.1. Estimación de Densidades

Sean  $X_1, \dots, X_n$  *v.a.r* i.i.d. con densidad  $g(x)$ . El problema consiste en estimar  $g(x)$  en cualquier punto a partir de las observaciones. Se verán primero algunas técnicas que llevarán a un entendimiento más intuitivo de la estimación por Kernel.

#### 3.1.1. Histograma

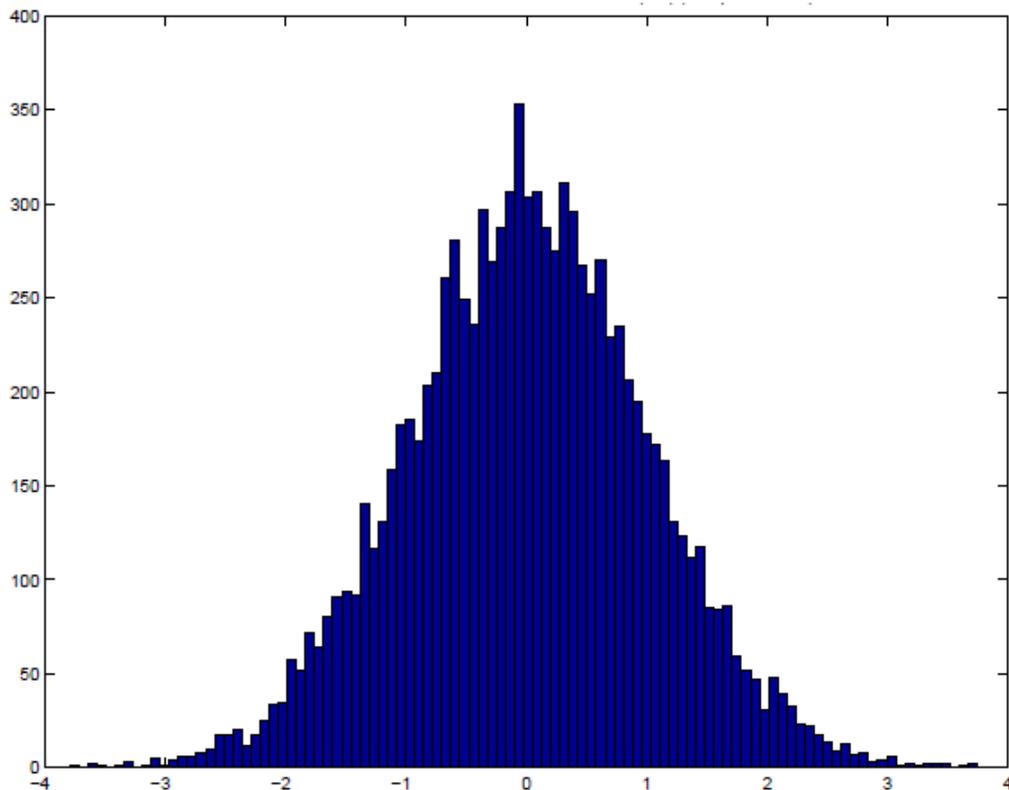
El método clásico para la estimación de la densidad a partir de un conjunto de observaciones es el histograma. Si se supone  $X \in \mathbb{R}$  y se desea estimar su densidad en el intervalo  $I = (a, b]$ , el método consiste en dividir el intervalo de interés en  $k$  subintervalos del mismo largo de forma tal que la unión de los mismos sea igual a  $I$ , y contar las ocurrencias de los  $X_i$  en los subintervalos. Sea el  $j$ -ésimo intervalo  $I_j$ :

$$I_j = \left( a + (j-1) \frac{b-a}{k}, a + j \frac{b-a}{k} \right], \quad j = 1, \dots, k$$

entonces el estimador de histograma es:

$$\hat{g}_n(x \in I_j) = \frac{1}{n} \sum_{j=1}^n 1_{I_j}(X_j) \quad (3.1)$$

Este estimador tiene una serie de problemas, el más notable es que  $\hat{g}_n$  es continua a trozos, (ver Figura 3-1) cuando  $g$  por lo general es continua.



**Figura 3-1:** Estimación por histograma de la densidad de una variable aleatoria  $N(0,1)$  (100 particiones).

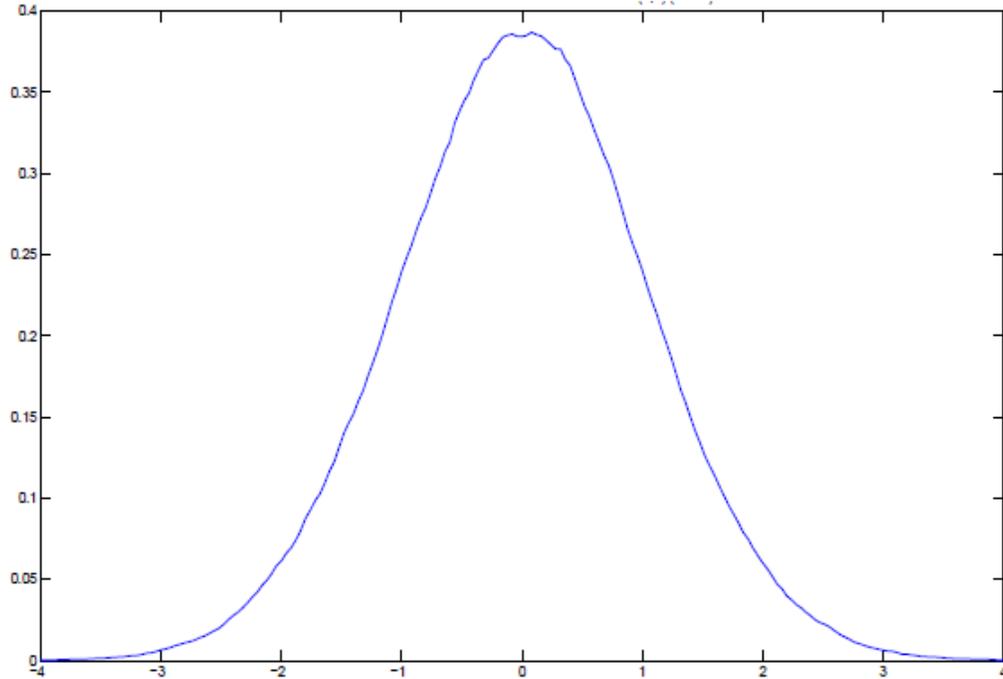
### 3.1.2. Histograma Móvil

Este método intenta suavizar la estimación obtenida por el histograma. En vez de partir el intervalo de interés en subintervalos fijos, se considera un intervalo o ventana de ancho fijo  $2h$  centrado en el punto en el que se desea realizar la estimación de la densidad. De esta forma, la estimación resulta:

$$\hat{g}_n(x) = \frac{1}{2nh} \sum_{j=1}^n 1_{(x-h, x+h]}(X_j) \quad (3.2)$$

Esta estimación, aunque también es continua a trozos, es más suave que el histograma (ver figura 3-2) y es el punto de partida para la estimación por Kernel.

F



**Figura 3-2:** Estimación por histograma móvil de la densidad de una variable aleatoria  $N(0,1)$  ( $h=0.5$ )

Para llegar a la idea intuitiva de la estimación por Kernel, resulta más claro reescribir esta estimación de la siguiente forma:

$$\begin{aligned} \hat{g}_n(x) &= \frac{1}{2nh} \sum_{j=1}^n 1_{(x-h, x+h]}(X_j) = \frac{1}{nh} \sum_{j=1}^n \frac{1}{2} 1_{(-1,1]} \left( \frac{x - X_j}{h} \right) = \frac{1}{nh} \sum_{j=1}^n K \left( \frac{x - X_j}{h} \right) \\ &= \frac{1}{n} \sum_{j=1}^n \frac{1}{h} K \left( \frac{x - X_j}{h} \right) \Rightarrow \hat{g}_n(x) = \frac{1}{n} \sum_{j=1}^n K_h(x - X_j) \end{aligned} \quad (3.3)$$

donde  $K(t) = \frac{1}{2} 1_{(-1,1]}(t)$  y  $K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right)$ . Antes de presentar el estimador por Kernel, se demostrará que el estimador de histograma móvil converge a su valor real:

$$\mathbb{E}(\hat{g}_n(x)) = \frac{1}{hn} \sum_{j=1}^n \mathbb{E} \left\{ K \left( \frac{x - X_j}{h} \right) \right\} = \frac{1}{h} \mathbb{E} \left\{ \frac{1}{2} 1_{(x-h, x+h]}(X_1) \right\}$$

Dado que:

$$\frac{1}{2} 1_{(x-h, x+h]}(X_1) = \begin{cases} 1 & : \text{ si } X_1 \in (x-h, x+h] \text{ (con probabilidad } p = \int_{x-h}^{x+h} g(t) dt) \\ 0 & : \text{ en otro caso (con probabilidad } 1-p) \end{cases}$$

entonces

$$\left. \begin{aligned} \mathbb{E}\{\hat{g}_n(x)\} &= \frac{1}{2h} \int_{x-h}^{x+h} g(t) dt \\ \text{si } g \text{ es continua en } x &\Rightarrow \frac{1}{2h} \int_{x-h}^{x+h} g(t) dt \xrightarrow{h \rightarrow 0} g(x) \end{aligned} \right\} \Rightarrow$$

$$\Rightarrow \mathbb{E}\{\hat{g}_n(x)\} \xrightarrow{h \rightarrow 0} g(x) \forall x \text{ punto continuo de } g. \quad (3.4)$$

### 3.1.3. Estimación por Kernel

La estimación dada en (3.3) sigue siendo continua a trozos al tratarse de una combinación lineal de funciones lineales a trozos. Surge entonces la idea de sustituir  $K(t)$  por una función continua, por lo que la fórmula del estimador por Kernel seguirá siendo la misma:

$$\hat{g}_n(x) = \frac{1}{nh} \sum_{j=1}^n K \left( \frac{x - X_j}{h} \right) \quad (3.5)$$

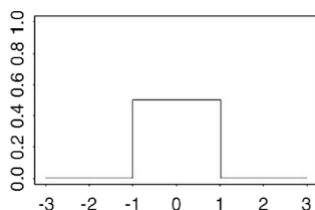
pero con un  $K(t)$  que será continuo. Esta función  $K$  recibe el nombre de *función Kernel*, más adelante en la definición 3.1 y definición 3.2 daremos las condiciones necesarias y suficientes para que una función sea una función Kernel. La estimación de densidad por Kernel no es mas que un promedio ponderado por la distancia de las observaciones al punto a ser estimado (es esta la razón por la que la llamaremos *ponderación local Kernel*). Cuanto mayor es la distancia del punto a un elemento de las observaciones, menor será su peso en la estimación. El peso lo determinarán el Kernel elegido y el valor de  $h$ . Cuanto mayor sea el valor de este

último, mayor será el peso de aquellos elementos de las observaciones que se encuentren alejados del punto, por lo que a  $h$  generalmente se lo denomina *ancho de banda*. Por último denotaremos la ponderación local Kernel por:

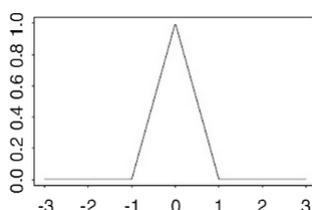
$$\Delta_i = \Delta_i(x, h, K) = \frac{1}{h} K\left(\frac{x - X_j}{h}\right)$$

para transformar las  $n$  v.a.r  $X_1, X_2, \dots, X_n$  en  $\Delta_1, \Delta_2, \dots, \Delta_n$ . Además, sustituimos  $1/nh$  por  $1/h$ , e implícitamente daremos por entendido que  $h$  depende de  $n$ . Aquí, se presentará, los Kernel más utilizados en su forma unidimensional. La Figura 3-3 muestra sus gráficas.

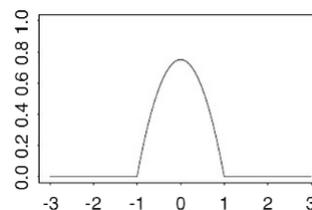
- **Kernel de Caja:**  $K(u) = \frac{1}{2}1_{[-1,1]}(u)$ ,
- **Kernel Triangular:**  $K(t) = \begin{cases} 1 - |u| & \text{si } |u| < 1 \\ 0 & \text{en otro caso} \end{cases}$ ,
- **Kernel Cuadrático:**  $K(u) = \frac{3}{4}(1 - u^2)1_{[-1,1]}(u)$ ,
- **Kernel Gaussiano:**  $K(u) = \frac{1}{\sqrt{2\pi}}\exp\{-\frac{u^2}{2}\}$ .



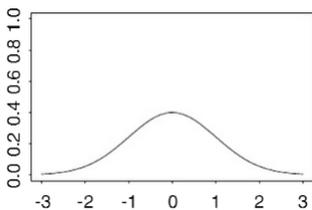
(a) Kernel de Caja



(b) Kernel Triangular



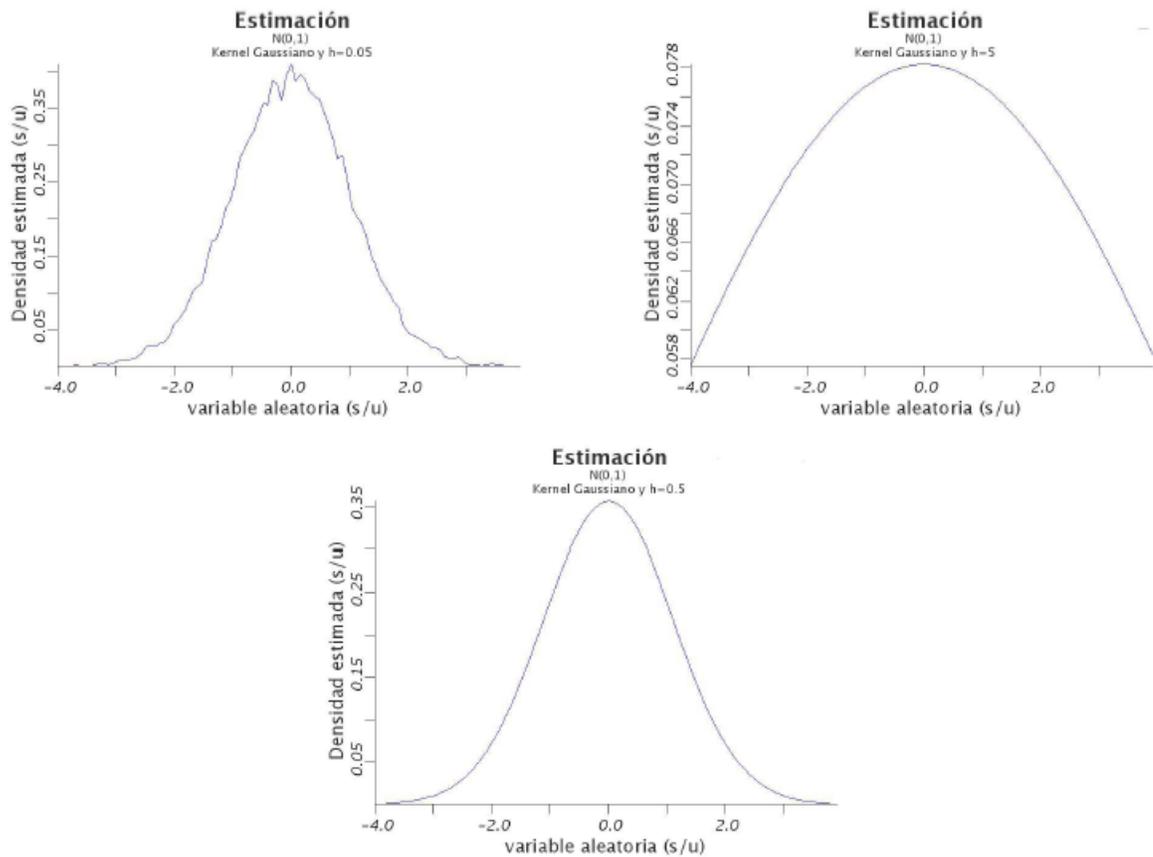
(c) Kernel Cuadrático



(d) Kernel Gaussiano

**Figura 3-3:** Kernel simétricos usuales

Luego de haber escogido el Kernel, resta tomar un valor de  $h$ . Este es un problema nada trivial, pues un  $h$  demasiado pequeño tiene como efecto que la varianza del estimador aumente demasiado (ver la primera gráfica de la Figura 3–4) pues son pocos los elementos de las observaciones considerados por punto, y un valor demasiado alto da resultados con un sesgo alto (ver la segunda gráfica de la Figura 3–4) pues se promedian demasiados elementos de la muestra. A este compromiso en el valor de  $h$  se le denomina compromiso *sesgo – varianza* y existen resultados para hallar el valor óptimo en algún sentido para el ancho de banda  $h$ .



**Figura 3–4:** Estimación por Kernel gaussiano con  $h = 0.05$ ;  $h = 0.5$ ;  $h = 5$  de la densidad  $N(0, 1)$

Métodos Kernel son conocidos e intensamente utilizados por la comunidad de no paramétricos porque son una forma útil para hacer la ponderación local. Vimos

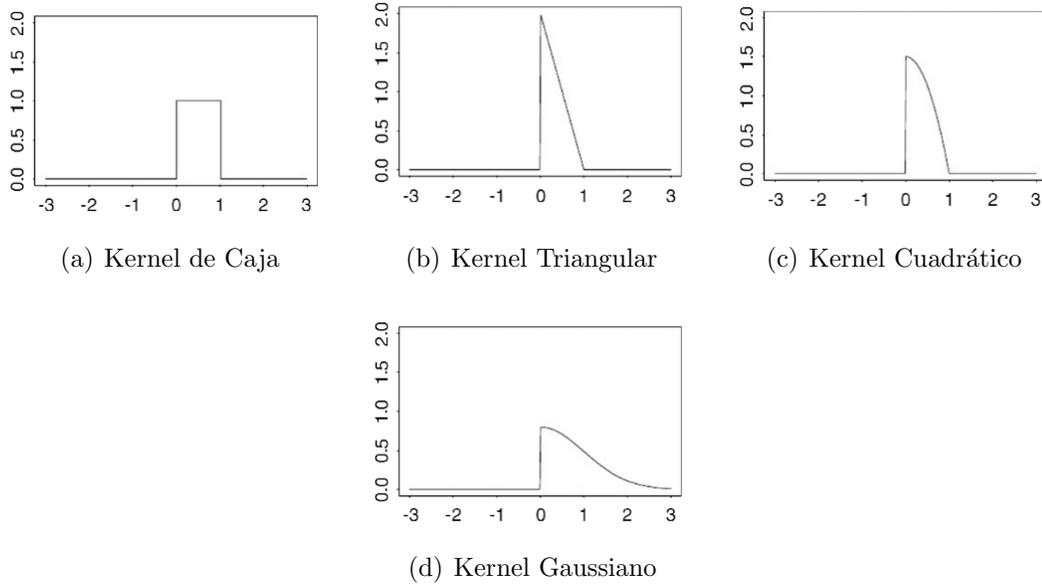
anteriormente un poco lo que es la ponderación local Kernel en el caso real, ahora la extenderemos al contexto funcional.

### Caso Funcional

El enfoque presentado anteriormente es suficiente para introducir la ponderación local Kernel en el caso funcional. Sean  $\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n$   $n$  *v.a.f* valoradas en  $E$  y sea  $\chi$  un elemento fijo de  $E$ . Una extensión funcional de las ideas de ponderación local Kernel anteriores sería transformar las  $n$  *v.a.f*  $\mathcal{X}_1, \dots, \mathcal{X}_n$  en las  $n$  cantidades:

$$\Delta_i = \frac{1}{V(h)} K \left( \frac{d(\chi, \mathcal{X}_i)}{h} \right), \quad (3.6)$$

donde  $d$  es una semi-métrica en  $E$  y  $K$  es un Kernel real asimétrico (lo de asimétrico, es debido a que  $K$  está valorada sobre  $d$ ) (la Figura 3–5 da ejemplos de Kernel asimétricos).



**Figura 3–5:** Kernel asimétricos usuales

En esta expresión  $V(h)$  es el volumen de:

$$B(\chi, h) = \{\chi' \in E, d(\chi, \chi') \leq h\}, \quad (3.7)$$

el cual es una bola, respecto a la topología inducida por  $d$ , centrada en  $\chi$  y de radio  $h$ . En otras palabras, se necesita tener a mano una medida en  $E$ . Esta es la diferencia principal con el caso real y multivariado para el cual implícitamente se utiliza la medida de Lebesgue, mientras que en el caso funcional el espacio  $E$  no tiene una medida universalmente aceptada de referencia (ver [17] para una discusión más profunda). Por lo tanto, con el fin de liberarse de una elección de una medida, se construye la normalización utilizando directamente la distribución de probabilidad de la *v.a.f.* Luego la ponderación local Kernel *funcional* de las variables se define por:

$$\Delta_i = \frac{K(h^{-1}d(\chi, \mathcal{X}_i))}{\mathbb{E}(K(h^{-1}d(\chi, \mathcal{X}_i)))}. \quad (3.8)$$

Así, es evidente ahora que (3.8) es una extensión de la ponderación local Kernel real al marco funcional. Recordemos que las funciones Kernel  $K$  a utilizarse aquí son necesariamente asimétricas, (ver Figura 3–4). En aras de simplicidad, en el resto de este documento, consideraremos solo dos tipos de Kernel de ponderación de variables funcionales, y los mostramos en las siguientes definiciones.

**Definición 3.1.** .

*I. Una función  $K : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  tal que  $\int K = 1$  es llamada un Kernel de tipo I, si existen dos constantes reales  $0 < C_1 < C_2$  tales que:*

$$C_1 1_{[0,1]} \leq K \leq C_2 1_{[0,1]}$$

*II. Una función  $K : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  tal que  $\int K = 1$  es llamada un Kernel de tipo II, si su soporte es  $[0, 1]$  y si su derivada  $K'$  existe en  $[0, 1]$  y satisface para dos constantes reales  $C_2 < C_1 < 0$ :*

$$C_2 \leq K' \leq C_1.$$

La primera familia de Kernel contiene los Kernel discontinuos habituales como la caja asimétrica, mientras que la segunda familia contiene los Kernel asimétricos

continuos (como el triangular, cuadrático,...). Por último, consideraremos este otro tipo de Kernel.

**Definición 3.2.** *Una función  $K : \mathbb{R} \rightarrow \mathbb{R}^+$  tal que  $\int K = 1$  con soporte compacto en  $[-1, 1]$  y tal que  $\forall u \in (0, 1)$ ,  $K(u) > 0$  es llamada un Kernel de tipo 0.*

### 3.2. Ponderación Local y Probabilidad en Vecindades

Ahora se podrá construir el puente entre la ponderación local y la noción de probabilidad en vecindades. Para fijar ideas, considere el Kernel más simple entre los de tipo I que es el Kernel de caja asimétrico. Sea  $\mathcal{X}$  una *v.a.f* valorada en  $E$  y  $\chi$  un elemento fijo de  $E$ . Podemos escribir:

$$\mathbb{E} \left( 1_{[0,1]} \left( \frac{d(\chi, \mathcal{X})}{h} \right) \right) = \mathbb{E} (1_{B(\chi, h)}(\mathcal{X})) = P(\mathcal{X} \in B(\chi, h)).$$

Teniendo en cuenta la ponderación local Kernel funcional de las variables (3.8), la probabilidad de la bola  $B(\chi, h)$  aparece claramente en la normalización. En donde el parámetro de suavizado  $h$  disminuye con el tamaño de la muestra de las variables funcionales (más precisamente,  $h$  tiende a cero cuando  $n$  tiende a  $\infty$ ). Así, cuando tomamos  $n$  muy grande,  $h$  es cercano a cero y entonces  $B(\chi, h)$  es considerada como una vecindad pequeña y  $P(\mathcal{X} \in B(\chi, h))$  como una probabilidad en vecindades pequeñas.

De ahora en adelante, para todo  $\chi$  en  $E$  y para todo  $h$  real positivo, utilizaremos la notación:

$$\varphi_\chi(h) = P(\mathcal{X} \in B(\chi, h)).$$

Esta noción de probabilidad en vecindades desempeña un papel importante tanto desde el punto de vista teórico como el práctico. Debido a que la noción de bola está fuertemente vinculada con la semi-métrica  $d$ , la elección de esta semi-métrica

se convertirá en una etapa importante. Esto se verá desde un punto de vista teórico a través de este trabajo, ya que la razón de convergencia de estas estimaciones funcionales no paramétricas se vinculan sistemáticamente con  $d$  a través del comportamiento, en torno a 0, de la función  $\varphi_\chi$  de probabilidad en la vecindad. Además, como se discutió el capítulo 2 la elección de la semi-métrica también se espera que sea un punto crucial, siempre y cuando nos centramos en los aspectos aplicados.

Debido a que las ideas de ponderación local Kernel funcional serán de vital importancia en todos los métodos estadísticos no paramétricos funcionales que se estudiarán más adelante en la sección 3.3, se discutirán dos resultados, de acuerdo con el hecho de que el Kernel es de tipo I o II. Esto puede ser visto como versiones funcionales de lo que se conoce como el teorema de Bochner en la literatura habitual no paramétrica de dimensión finita (ver [8] y la discusión). Los siguientes lemas son importantes en esta tesis.

Sea  $\mathcal{X}$  una *v.a.f* que toma sus valores en el espacio semi-métrico  $(E, d)$ , sea  $\chi$  un elemento fijo de  $E$ , sea  $h$  un número real positivo y sea  $K$  una función Kernel.

**Lema 1.** *Si  $K$  es un Kernel tipo I, entonces existen constantes reales finitas no negativas  $C$  y  $C'$  tales que:*

$$C\varphi_\chi(h) \leq \mathbb{E}K\left(\frac{d(\chi, \mathcal{X})}{h}\right) \leq C'\varphi_\chi(h).$$

**Demostración:** Debido a que  $K$  es un Kernel de tipo I, tenemos por definición 3.1i, que existen dos constantes  $C_1$  y  $C_2$  tal que

$$C_1 1_{[0,1]} \leq K \leq C_2 1_{[0,1]},$$

lo cual implica que

$$C_1 1_{[0,1]}\left(\frac{d(\chi, \mathcal{X})}{h}\right) \leq K\left(\frac{d(\chi, \mathcal{X})}{h}\right) \leq C_2 1_{[0,1]}\left(\frac{d(\chi, \mathcal{X})}{h}\right),$$

$$C_1 1_{B(\chi, h)}(\mathcal{X}) \leq K \left( \frac{d(\chi, \mathcal{X})}{h} \right) \leq C_2 1_{B(\chi, h)}(\mathcal{X}).$$

y por consiguiente

$$\mathbb{E} (C_1 1_{B(\chi, h)}(\mathcal{X})) \leq \mathbb{E} \left( K \left( \frac{d(\chi, \mathcal{X})}{h} \right) \right) \leq \mathbb{E} (C_2 1_{B(\chi, h)}(\mathcal{X})),$$

$$C_1 \mathbb{E} (1_{B(\chi, h)}(\mathcal{X})) \leq \mathbb{E} \left( K \left( \frac{d(\chi, \mathcal{X})}{h} \right) \right) \leq C_2 \mathbb{E} (1_{B(\chi, h)}(\mathcal{X})),$$

$$C \varphi_\chi(h) \leq \mathbb{E} K \left( \frac{d(\chi, \mathcal{X})}{h} \right) \leq C' \varphi_\chi(h).$$

Donde  $C = C_1$ ,  $C' = C_2$  y  $\varphi_\chi(h) = \mathbb{E} (1_{B(\chi, h)}(\mathcal{X}))$

□

**Lema 2.** Si  $K$  es un Kernel de tipo II y si  $\varphi_\chi(\cdot)$  satisfase

$$\exists C_3 > 0, \exists \epsilon_0, \forall \epsilon < \epsilon_0, \int_0^\epsilon \varphi_\chi(u) du > C_3 \epsilon \varphi_\chi(\epsilon), \quad (3.9)$$

entonces existen constantes reales finitas no negativas  $C$  y  $C'$  tal que, para  $h$  lo suficientemente pequeño:

$$C \varphi_\chi(h) \leq \mathbb{E} K \left( \frac{d(\chi, \mathcal{X})}{h} \right) \leq C' \varphi_\chi(h).$$

**Demostración:** como  $K$  es de soporte en  $[0, 1]$  tenemos que

$$\mathbb{E} K \left( \frac{d(\chi, \mathcal{X})}{h} \right) = \int_0^1 K(t) dP^{\frac{d(\chi, \mathcal{X})}{h}}(t),$$

y porque  $K'$  existe,  $K(t) = K(0) + \int_0^t K'(u) du$ , lo cual implica que

$$\begin{aligned} \mathbb{E} K \left( \frac{d(\chi, \mathcal{X})}{h} \right) &= \int_0^1 K(0) dP^{\frac{d(\chi, \mathcal{X})}{h}}(t) + \int_0^1 \left( \int_0^t K'(u) du \right) dP^{\frac{d(\chi, \mathcal{X})}{h}}(t), \\ &= K(0) \int_0^1 dP^{\frac{d(\chi, \mathcal{X})}{h}}(t) + \int_0^1 \left( \int_0^1 K'(u) 1_{[u, 1]}(t) du \right) dP^{\frac{d(\chi, \mathcal{X})}{h}}(t), \\ &= K(0) P \left( 0 < \frac{d(\chi, \mathcal{X})}{h} < 1 \right) + \int_0^1 K'(u) \left( \int_0^1 1_{[u, 1]}(t) dP^{\frac{d(\chi, \mathcal{X})}{h}}(t) \right) du, \end{aligned}$$

esta última ecuación se obtiene aplicando el *teorema de Fubini*, por consiguiente

$$\begin{aligned}
&= K(0)\varphi_X(h) + \int_0^1 K'(u)P\left(u < \frac{d(\mathcal{X}, \mathcal{X})}{h} < 1\right) du, \\
&= K(0)\varphi_X(h) + \int_0^1 K'(u)(\varphi_X(h) - \varphi_X(hu)) du, \\
&= K(0)\varphi_X(h) + \varphi_X(h) \int_0^1 K'(u) du - \int_0^1 K'(u)\varphi_X(hu) du, \\
&= K(0)\varphi_X(h) + \varphi_X(h)K(1) - \varphi_X(h)K(0) - \int_0^1 K'(u)\varphi_X(hu) du,
\end{aligned}$$

usando el hecho de que  $K(1) = 0$ , resulta que

$$\mathbb{E}K\left(\frac{d(\mathcal{X}, \mathcal{X})}{h}\right) = - \int_0^1 K'(u)\varphi_X(hu) du.$$

Ahora como  $K$  es de tipo II, existen  $C_1 < 0$  tal que  $K' \leq C_1$ , así que

$$-K'(u)\varphi_X(hu) \geq -C_1\varphi_X(hu)$$

por lo tanto

$$- \int_0^1 K'(u)\varphi_X(hu) du \geq -C_1 \int_0^1 \varphi_X(hu) du$$

haciendo el cambio de variable  $t = hu$  obtenemos

$$- \int_0^1 K'(u)\varphi_X(hu) du \geq -\frac{C_1}{h} \int_0^h \varphi_X(t) dt$$

haciendo uso de (3.9), para  $h < \epsilon_0$

$$\mathbb{E}K\left(\frac{d(\mathcal{X}, \mathcal{X})}{h}\right) \geq -\frac{C_1}{h}C_3h\varphi_X(h),$$

y si tomamos  $C = -C_3C_1$

$$\mathbb{E}K\left(\frac{d(\mathcal{X}, \mathcal{X})}{h}\right) \geq C\varphi_X(h).$$

con respecto a la cota superior, sabemos que  $K$  es continua con soporte  $[0, 1]$ , y por tanto es acotada, así que podemos elegir  $C' = \sup_{t \in [0,1]} K(t)$  como la cota superior y

seguir los mismos argumentos como para el lema 1 y obtener

$$\mathbb{E}K\left(\frac{d(\chi, \mathcal{X})}{h}\right) \leq C' \varphi_\chi(h)$$

□

### 3.3. Métodos y Modelos de Predicción

Como se discutió en la sección 3.1, la ponderación local Kernel se adapta muy bien para propósitos de predicción, y todas estas estimaciones no paramétricas se basarán en las ideas del Kernel.

Describiremos varios enfoques relativos a la predicción no paramétrica de alguna respuesta escalar. El contexto funcional aparece a través de la variable funcional explicativa. Nos centraremos en tres métodos de predicción complementarios, a saber, la esperanza condicional, la mediana condicional y la moda condicional. La esperanza condicional se refiere al conocido método de regresión, mientras que tanto la mediana condicional como la moda condicional están estrechamente vinculadas con la estimación de la distribución condicional.

#### 3.3.1. Métodos de Predicción

Sean  $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$ ,  $n$  pares independientes e idénticamente distribuidos como  $(\mathcal{X}, Y)$  y valorados en  $E \times \mathbb{R}$ , donde  $(E, d)$  es un espacio semi-métrico. Sea  $\chi$  un elemento fijo de  $E$  y  $y \in \mathbb{R}$ , sea  $\mathcal{N}_\chi \subset E$  una vecindad de  $\chi$  y  $S$  un subconjunto compacto fijo de  $\mathbb{R}$  tal que  $y \in S$ . Dado  $\chi$  se denotará por  $\hat{y}$  el valor de la predicción para la respuesta escalar.

Los métodos de predicción de la respuesta escalar  $Y$  de la predictora funcional  $\mathcal{X}$  son todos basados en la distribución condicional de  $Y$  dado  $\mathcal{X}$ .

### 3.3.1.1. Operador Regresión

El operador regresión (no lineal)  $r$  de  $Y$  en  $\mathcal{X}$  es definido por:

$$r(\chi) = \mathbb{E}(Y \mid \mathcal{X} = \chi), \quad (3.10)$$

y la función de distribución acumulativa condicional (*f.d.a*) de  $Y$  dado  $\mathcal{X}$  es definida por:

$$\forall y \in \mathbb{R}, F_Y^{\mathcal{X}}(\chi, y) = P(Y \leq y \mid \mathcal{X} = \chi). \quad (3.11)$$

Además, si la función de probabilidad de  $Y$  dado  $\mathcal{X}$  es absolutamente continua con respecto a la medida de Lebesgue, notaremos  $f_Y^{\mathcal{X}}(\chi, y)$  el valor de la correspondiente función de densidad condicional en  $(\chi, y)$ . Note que bajo suposición de diferenciación de  $F_Y^{\mathcal{X}}(\chi, \cdot)$ , esta función de densidad condicional puede ser escrita como:

$$\forall y \in \mathbb{R}, f_Y^{\mathcal{X}}(\chi, y) = \frac{\partial}{\partial y} F_Y^{\mathcal{X}}(\chi, y).$$

Está claro que cada operador (no lineal) da información acerca de la relación entre  $\mathcal{X}$  y  $Y$  y cualquiera puede ser de completo uso para la predicción de  $y$  dado  $\chi$ . Además, cada uno de ellos dará algún método de predicción específico.

La primera forma para la construcción de esta predicción se obtiene directamente del operador regresión tomando:

$$\hat{y} = \hat{r}(\chi), \quad (3.12)$$

donde  $\hat{r}$  es una estimación de  $r$ .

### 3.3.1.2. Operador Mediana Condicional

Este segundo operador consiste en considerar la mediana  $m(\chi)$  de la *f.d.a* condicional  $F_Y^{\mathcal{X}}$ :

$$m(\chi) = \inf\{y \in \mathbb{R}, F_Y^{\mathcal{X}}(\chi, y) \geq 1/2\}, \quad (3.13)$$

y se usa como predictor:

$$\hat{y} = \hat{m}(\chi), \quad (3.14)$$

donde  $\hat{m}(\chi)$  es una estimación de esta mediana condicional funcional  $m(\chi)$ . Notar, que tal estimador dependerá de alguna estimación previa del operador no lineal  $F_Y^{\mathcal{X}}$ .

### 3.3.1.3. Operador Moda Condicional

Finalmente, el tercer predictor es basado en la moda  $\theta(\chi)$  de la densidad condicional de  $Y$  dado  $\mathcal{X}$ :

$$\theta(\chi) = \arg \sup_{y \in S} f_Y^{\mathcal{X}}(\chi, y). \quad (3.15)$$

Esta definición asume implícitamente que  $\theta(\chi)$  existe en  $S$ . El predictor es denotado por:

$$\hat{y} = \hat{\theta}(\chi), \quad (3.16)$$

donde  $\hat{\theta}(\chi)$  es una estimación de  $\theta(\chi)$ . Una vez más notar que este estimador dependerá directamente de alguna estimación previa del operador no lineal  $f_Y^{\mathcal{X}}$ .

### 3.3.1.4. Banda de Confianza predictiva

El hecho de que los tres métodos presentados anteriormente son sobre predicción puntual, cabe señalar que el segundo método también se puede utilizar para la construcción de la banda de confianza de predicción, ya que se puede utilizar para estimar cualquier cuartil de la *f.d.a*  $F_Y^{\mathcal{X}}$ . Precisamente, estos cuartiles son definidos para  $\alpha \in (0, 1)$  por:

$$t_\alpha(\chi) = \inf\{y \in \mathbb{R}, F_Y^{\mathcal{X}}(\chi, y) \geq \alpha\}. \quad (3.17)$$

Por lo tanto, a partir de una estimación  $\hat{t}_\alpha(\chi)$  de  $t_\alpha(\chi)$ , el siguiente intervalo:

$$(\hat{t}_\alpha(\chi), \hat{t}_{1-\alpha}(\chi))$$

es una manera de construir, para  $\alpha \in (0, 1)$ , un  $(1 - \alpha)$  banda de confianza de predicción.

### 3.3.2. Modelos no Paramétricos

Conociendo ya los métodos estadísticos, queda por construir de forma explícita algunas estimaciones de los predictores teóricos que se han introducido (Es decir, la regresión, mediana condicional, moda condicional y cuantiles condicional). Sin embargo, esto no se puede hacer antes de tener precisamente el tipo de modelo estadístico que deseamos introducir. Ese es el objetivo de esta sección.

Cada uno de los tres predictores se basa en la estimación de algún operador (no lineal). Por lo tanto, la primera etapa de la modelización estadística consiste en introducir algunos conjuntos de limitaciones que actúan bien en  $r$ ,  $F_Y^{\mathcal{X}}$  o  $f_Y^{\mathcal{X}}$ . Teniendo en cuenta que se desea hallar modelos libre de parámetros, esto conduce a la introducción de modelos no paramétricos (de acuerdo con la Definición 2.8).

#### Modelo para la Esperanza Condicional

Este método es referido al operador regresión  $r$  (un operador no lineal de  $E$  en  $\mathbb{R}$ ) y se considerará el siguiente modelo:

$$r \in C_E^0, \tag{3.18}$$

donde

$$C_E^0 = \left\{ f : E \rightarrow \mathbb{R} / \lim_{d(\chi, \chi') \rightarrow 0} f(\chi') = f(\chi) \right\}.$$

### Modelo para la Mediana Condicional Funcional

Este método se refiere al operador  $F_Y^{\mathcal{X}}$  (un operador no lineal de  $E \times \mathbb{R}$  en  $\mathbb{R}$ ). Antes de escribir el modelo, con el fin de simplificar la representación y no enmascarar nuestro objetivo principal, se asumiremos que el operador mediana condicional pertenece el siguiente conjunto:

$$\mathcal{S}_{f.d.a}^{\mathcal{X}} = \{g : E \times \mathbb{R} \rightarrow \mathbb{R}, g(\chi, \cdot) \text{ es una f.d.a estrictamente creciente}\}.$$

De hecho,  $F_Y^{\mathcal{X}} \in \mathcal{S}_{f.d.a}^{\mathcal{X}}$  garantiza la existencia y unicidad de la mediana condicional, que puede ser definida como:

$$m(\chi) = F_Y^{\mathcal{X}-1}(1/2) \text{ donde } F_Y^{\mathcal{X}} = \begin{cases} \mathbb{R} \rightarrow [0, 1] \\ y \mapsto F_Y^{\mathcal{X}}(y) = F_Y^{\mathcal{X}}(\chi, y). \end{cases}$$

y luego consideramos el siguiente modelo:

$$F_Y^{\mathcal{X}} \in C_0^{E \times \mathbb{R}} \cap \mathcal{S}_{f.d.a}^{\mathcal{X}}, \quad (3.19)$$

donde

$$C_0^{E \times \mathbb{R}} = \left\{ \begin{array}{l} f : E \times \mathbb{R} \rightarrow \mathbb{R} / \forall \chi' \in \mathcal{N}_{\chi}, \lim_{d(\chi, \chi') \rightarrow 0} f(\chi', y) = f(\chi, y) \\ \text{y } \forall y' \in \mathbb{R} \lim_{|y-y'| \rightarrow 0} f(\chi, y') = f(\chi, y). \end{array} \right\}$$

### Modelo para la Moda Condicional funcional

Este método se refiere al operador  $f_Y^{\mathcal{X}}$  (un operador no lineal de  $E \times \mathbb{R}$  en  $\mathbb{R}$ ). Recordemos que  $S$  es un subconjunto compacto de  $\mathbb{R}$ . Empecemos introduciendo el siguiente conjunto de restricción:

$$\mathcal{S}_{dens}^x = \left\{ \begin{array}{l} f : E \times \mathbb{R} \rightarrow \mathbb{R} / \\ \exists \xi > 0, \exists y_0 \in S, f(\chi, \cdot) \text{ es estrictamente creciente en } (y_0 - \xi, y_0) \\ \text{y estrictamente decreciente en } (y_0, y_0 + \xi). \end{array} \right\}$$

Es claro que si  $f_Y^{\mathcal{X}}$  cae en  $\mathcal{S}_{dens}^x$ , el problema de maximizar  $f_Y^{\mathcal{X}}(\chi, y)$  sobre  $S$  tiene una única solución la cual es exactamente  $y_0$ . Por lo tanto, bajo esta restricción, la moda condicional  $\theta(\chi)$  puede ser definida como:

$$\theta(\chi) = \arg \sup_{y \in S} f_Y^{\mathcal{X}}(\chi, y).$$

El conjunto de restricción  $\mathcal{S}_{dens}^x$  definido así, en una manera de garantizar la unicidad de la moda condicional  $\theta(\chi)$ . Ahora podemos definir el modelo no paramétrico funcional para la predicción mediante la moda condicional como:

$$f_Y^{\mathcal{X}} \in C_0^{E \times \mathbb{R}} \cap \mathcal{S}_{dens}^x. \quad (3.20)$$

Los modelos no paramétricos funcionales (3.18), (3.19), (3.20) serán llamados modelos no paramétricos funcionales tipo-continuidad porque la propiedad de continuidad es su principal característica funcional en común. De hecho, veremos que los modelos del tipo-continuidad, nos permiten obtener resultados de convergencia para las estimaciones no paramétricas de los mismos. Otro tipo de modelo considerado por Ferraty y Vieu son los de tipo-Lipschitz que nos permiten obtener de forma precisa la razón de convergencia de las estimaciones de estos modelos (ver [7]).

Estas consideraciones teóricas no son sorprendentes, ya que obedecen al siguiente principio estadístico general: mientras más sea restrictivo el modelo más precisamente se puede describir el comportamiento teórico. En otras palabras, existe una compensación entre el tamaño del conjunto de restricción (que producen los modelos) y la precisión de la razón de convergencia que podamos esperar. Otro punto

importante acerca de tal modelo no paramétrico funcional es que las dos principales dificultades que aparecen son: la dimensión infinita del espacio de restricción (i.e. el modelo no paramétrico) y el espacio infinito  $E$  (la característica funcional de la variable explicativa). Por lo tanto, un gran desafío teórico consiste en la provisión de propiedades asintóticas de este doble contexto de dimensión infinita, razón principal que justifica mi interés en este trabajo y esto se hará en el Capítulo 4.

## Capítulo 4

# RESULTADOS Y VALIDACIÓN

Una vez que el modelado no paramétrico se ha introducido, se tiene que encontrar la manera de estimar los diversos objetos matemáticos expuestos en los modelos anteriores, a saber, los operadores (no lineales)  $r$ ,  $F_Y^{\mathcal{X}}$ ,  $f_Y^{\mathcal{X}}$ . De acuerdo con la discusión en el capítulo 3, los estimadores Kernel son buenos candidatos para lograr una ponderación local enfocada en el ámbito funcional.

### 4.1. Estimadores de los Modelos No Paramétricos

Sea  $(\mathcal{X}_i, Y_i)_{i=1, \dots, n}$  una muestra funcional de tamaño  $n$ , queremos estimar  $r(\chi)$ ,  $F_Y^{\mathcal{X}}(\chi, y)$ ,  $f_Y^{\mathcal{X}}(\chi, y)$  para un  $(\chi, y)$  no necesariamente incluido en la muestra. Supongamos que  $\Phi(\chi)$  es la densidad marginal de  $\chi$ , que  $f(\chi, y)$  es la densidad conjunta de  $(\chi, y)$  y que  $\Psi(\chi) = \int y f(\chi, y) dy$ . Veremos a continuación la construcción de los estimadores modelo por modelo.

#### 4.1.1. Estimador del operador regresión

Como se vió en el capítulo 3 nuestro operador no lineal de regresión está asociado a la esperanza condicional, entonces:

$$r(\chi) = \mathbb{E}(Y \mid \mathcal{X} = \chi) = \int y f_Y^{\mathcal{X}}(\chi, y) dy = \int y \frac{f(\chi, y)}{\Phi(\chi)} dy = \frac{1}{\Phi(\chi)} \int y f(\chi, y) dy = \frac{\Psi(\chi)}{\Phi(\chi)},$$

así que, una idea para la estimación del operador no lineal de regresión  $r$  sería el siguiente estimador de regresión kernel funcional:

$$\hat{r}(\chi) = \frac{\sum_{i=1}^n Y_i \frac{K(h^{-1}d(\chi, \mathcal{X}_i))}{\mathbb{E}(K(h^{-1}d(\chi, \mathcal{X}_i)))}}{\sum_{i=1}^n \frac{K(h^{-1}d(\chi, \mathcal{X}_i))}{\mathbb{E}(K(h^{-1}d(\chi, \mathcal{X}_i)))}} = \frac{\hat{\Psi}(\chi)}{\hat{\Phi}(\chi)}.$$

El numerador no es más que una estimación por Kernel de  $\Psi(\chi) = \int yf(\chi, y) dy$  y el denominador es la estimación de la densidad  $\Phi(\chi)$  de  $\mathcal{X}_i$  tal como se vio en la sección 3.1 del capítulo 3, conociendo que  $K$  es un Kernel asimétrico y que  $h$  (que depende de  $n$ ) es el ancho de banda. Esta es una extensión funcional del conocido estimador Nadayara-Watson ([6] y [10]) que fue previamente introducido para regresión no paramétrica finita dimensional (ver [19]). El estimador acarrea los mismos problemas vistos en la estimación de densidad por Kernel en cuanto a la dificultad en la elección del valor del ancho de banda. El compromiso sesgo-varianza continúa manteniéndose en este caso, por las mismas razones que en la estimación de la densidad por Kernel. De todas formas, también existen resultados para la elección del ancho de banda óptimo en algún sentido [21], pero esto se aleja un poco de nuestro interés.

Finalmente, podemos definir un estimador para el operador no lineal de regresión como:

$$\hat{r}(\chi) = \frac{\sum_{i=1}^n Y_i K(h^{-1}d(\chi, \mathcal{X}_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i))}. \quad (4.1)$$

El principal cambio en esta extensión funcional viene de la semi-métrica  $d$  que mide la proximidad entre objetos funcionales. Para ver como tal estimador funciona, vamos a considerar las siguientes cantidades:

$$w_{i,h}(\chi) = \frac{K(h^{-1}d(\chi, \mathcal{X}_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathcal{X}_i))}.$$

Por lo tanto, es fácil escribir el estimador Kernel como:

$$\hat{r}(\chi) = \sum_{i=1}^n w_{i,h} Y_i,$$

el cual es en realidad un promedio ponderado porque:

$$\sum_{i=1}^n w_{i,h} = 1.$$

El comportamiento de los  $w_{i,h}(\chi)$ 's se puede deducir de la forma de la función Kernel asimétrica  $K$  (ver la Figura 3-5 del capítulo 4). Para fijar ideas, consideremos Kernel positivos con soporte en  $[0, 1]$ . En este caso, es claro que cuando más pequeño es  $d(\chi, \mathbf{x}_i)$ , más grande es  $K(h^{-1}d(\chi, \mathbf{x}_i))$ . En otras palabras, entre más cercano es  $\mathbf{x}_i$  de  $\chi$ , más grande es el peso asignado a  $Y_i$ . Los aspectos locales de tal método aparecen a través de comportamiento local de los pesos alrededor de  $\chi$ , el punto de  $E$  en el que está valorado el estimador. Debemos tener en cuenta también que, tan pronto como  $d(\chi, \mathbf{x}_i) > h$ , tenemos  $w_{i,h}(\chi) = 0$ . Lo que quiere decir que el estimador  $\hat{r}(\chi)$  solo tiene en cuenta los  $Y_i$ 's para los que los correspondiente  $\mathbf{x}_i$ 's están distantes de  $\chi$  a lo más  $h$ . Así, el parámetro  $h$  juega el mejor rol porque controla el número de términos en el promedio ponderado. Además, entre más pequeño es  $h$  más pequeño es el número de  $Y_i$ 's a tener en cuenta en el promedio. En otras palabras, entre más pequeño es  $h$ , más  $r(\chi)$  es sensible a pequeñas variaciones de los  $Y_i$ . En el caso contrario, entre más grande es  $h$ , más grande es el número de términos en la suma y menos sensible es  $r(\chi)$  con respecto a las pequeñas variaciones de los  $Y_i$ . Podemos decir que  $h$  tiene un efecto de suavización, y en este sentido  $h$  es un **parámetro de suavizado** como se vió en el capítulo 3.

### 4.1.2. Estimador del operador f.d.a condicional

Nuestro propósito ahora es la estimación  $\hat{F}_Y^{\mathbf{x}}$  de la condicional  $F_Y^{\mathbf{x}}$ , pero primero vamos a explicar como podemos extender la idea previamente utilizada para la construcción del estimador de regresión Kernel. Claramente,  $F_Y^{\mathbf{x}}(\chi, y) = P(Y \leq y \mid \mathbf{x} = \chi)$ , puede ser expresado en términos de la esperanza condicional:

$$F_Y^{\mathbf{x}}(\chi, y) = \mathbb{E}(1_{(-\infty, y]}(Y) \mid \mathbf{x} = \chi),$$

por lo tanto una buena idea sería definir un estimador f.d.a condicional Kernel como:

$$\tilde{F}_Y^{\mathbf{x}}(\chi, y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i))1_{(-\infty, y]}(Y_i)}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i))}.$$

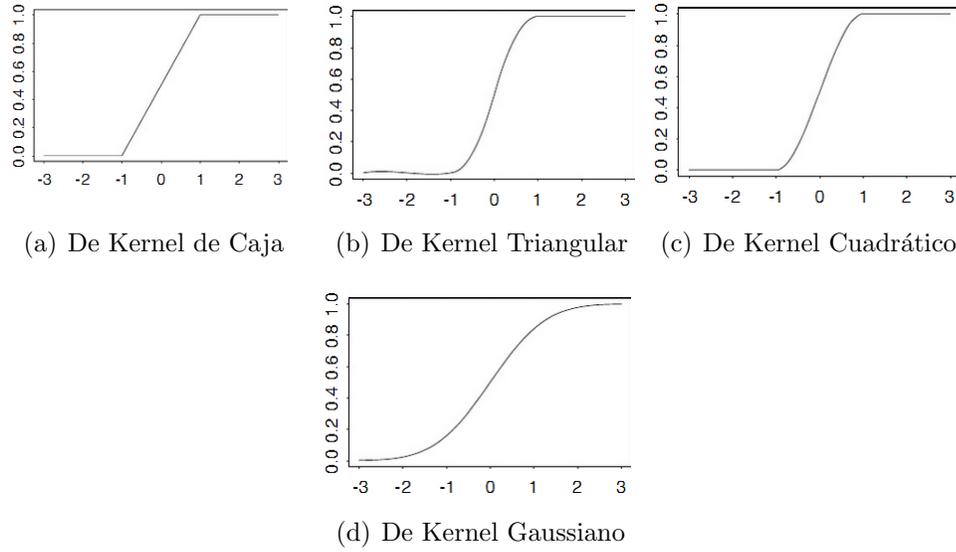
Al seguir las ideas previas desarrolladas por [9] y [15] en el caso finito dimensional, es fácil construir una versión suave de este estimador. Para ello, es suficiente cambiar la función de indicador básico por una función de distribución acumulativa (f.d.a). Sea  $K_0$  un Kernel simétrico usual, sea  $H$  definido como:

$$\forall u \in \mathbb{R}, \quad H(u) = \int_{-\infty}^u K_0(v) dv,$$

y definimos el estimador f.d.a condicional Kernel como sigue:

$$\hat{F}_Y^{\mathbf{x}}(\chi, y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i))H(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i))}, \quad (4.2)$$

donde  $g$  es un número real estrictamente positivo (que depende de  $n$ ). La Figura 4-1 muestra varios ejemplos de Kernel integrados basados en los Kernel simétricos vistos anteriormente.



**Figura 4–1:** Varios ejemplos de Kernel integrados simétricos

De este estimador f.d.a condicional (4.2), uno puede atacar el problema de predicción mediante la definición de un estimador Kernel de la mediana condicional funcional  $m(\chi)$  como sigue:

$$\hat{m}(\chi) = \inf\{y \in \mathbb{R}, \hat{F}_Y^{\mathcal{X}}(\chi, y) \geq 1/2\}. \quad (4.3)$$

Del mismo modo, podemos también definir por (4.2) un estimador Kernel del cuantil condicional funcional  $t_\alpha(\chi)$ , para cualquier  $\alpha \in (0, 1)$ , como sigue:

$$\hat{t}_\alpha(\chi) = \inf\{y \in \mathbb{R}, \hat{F}_Y^{\mathcal{X}}(\chi, y) \geq \alpha\}. \quad (4.4)$$

### 4.1.3. Estimador del operador densidad condicional

Se sabe que bajo cualquier suposición de diferenciabilidad, la función de densidad condicional puede ser obtenida por la diferenciación de la f.d.a condicional. Ya que tenemos a la mano un estimador  $\hat{F}_Y^{\mathcal{X}}$  de  $F_Y^{\mathcal{X}}$  es natural proponer el siguiente estimador:

$$\hat{f}_Y^{\mathcal{X}}(\chi, y) = \frac{\partial}{\partial y} \hat{F}_Y^{\mathcal{X}}(\chi, y).$$

Asumiendo la diferenciabilidad de  $H$ , tenemos

$$\frac{\partial}{\partial y} \hat{F}_Y^{\mathcal{X}}(\chi, y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i)) \frac{\partial}{\partial y} H(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i))},$$

y esta es la motivación a la siguiente expresión para la estimación de la densidad condicional Kernel:

$$\hat{f}_Y^{\mathcal{X}}(\chi, y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i)) \frac{1}{g} H'(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i))}.$$

Más generalmente, podemos afirmar para cualquier Kernel  $K_0$  la siguiente definición:

$$\hat{f}_Y^{\mathcal{X}}(\chi, y) = \frac{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i)) \frac{1}{g} K_0(g^{-1}(y - Y_i))}{\sum_{i=1}^n K(h^{-1}d(\chi, \mathbf{x}_i))}. \quad (4.5)$$

Este tipo de estimador ha sido ampliamente estudiado en el entorno no-funcional, es decir, en la condición cuando  $\mathcal{X}$  se cambia a una variable de dimensión finita.

Se puede fácilmente definir el siguiente estimador moda condicional funcional Kernel de  $\theta(\chi)$ :

$$\hat{\theta}(\chi) = \arg \sup_{y \in S} \hat{f}_Y^{\mathcal{X}}(\chi, y). \quad (4.6)$$

## 4.2. Convergencia de los Estimadores

El objetivo de esta sección es presentar los resultados asintóticos vinculados a los estimadores no paramétricos de los tres predictores funcionales ya definidos en la Sección 4.1, esperanza condicional, mediana condicional y moda condicional. Los predictores no paramétricos que son utilizados se basan en las ideas de un Kernel suavizado y son los definidos anteriormente en la sección 3.1. Todos los resultados son presentados en términos de convergencia casi completa. Este modo de convergencia ha sido seleccionado porque tiene dos ventajas importantes, en primer lugar es más fuerte que la convergencia casi segura y la convergencia en probabilidad, y segundo es más fácil probar. Para más detalles acerca de este tipo de convergencia vea el apéndice A, allí encontrará las definiciones básicas provenientes con convergencia casi completa y mostramos enlaces con otros modos usuales de convergencia como lo es la convergencia casi segura y convergencia en probabilidad.

### 4.2.1. Estimador de Regresión

Aquí nos centraremos en la convergencia puntual casi completa del estimador Kernel funcional de regresión  $r(\boldsymbol{\mathcal{X}}) = \mathbb{E}(Y \mid \boldsymbol{\mathcal{X}})$  el cual fue definido en (4.1), y lo podemos escribir como

$$Y = r(\boldsymbol{\mathcal{X}}) + \epsilon \quad \text{con } \mathbb{E}(\epsilon \mid \boldsymbol{\mathcal{X}}) = 0.$$

Antes de dar el resultado asintótico principal, tenemos que hacer algunas suposiciones. La primera es acerca de la probabilidad de bolas pequeñas de la variable funcional  $\boldsymbol{\mathcal{X}}$ , que es quien nos permite decir que la probabilidad de observar la *v.a.f* alrededor de  $\chi$  (el elemento funcional en el que se evalúa el operador de regresión) es no nula:

$$\forall \epsilon > 0, P(\boldsymbol{\mathcal{X}} \in B(\chi, \epsilon)) = \varphi(\epsilon) > 0. \quad (4.7)$$

Es clásico en el contexto no paramétrico multivariado suponer que la densidad de la variable explicativa multivariante es estrictamente positiva, y la hipótesis (4.7)

es una extensión de esa noción. Además, los parámetros que intervienen en el estimador, que es el ancho de banda y la función Kernel, tienen que satisfacer:

$$\left\{ \begin{array}{l} h \text{ es una sucesión positiva tal que} \\ \lim_{n \rightarrow \infty} h = 0 \text{ y } \lim_{n \rightarrow \infty} \frac{\log n}{n\varphi_\chi(h)} = 0, \\ K \text{ es un Kernel de tipo I} \\ o \\ K \text{ es un Kernel de tipo II y (3.9) se cumple.} \end{array} \right. \quad (4.8)$$

Finalmente, consideraremos una variable de respuesta escalar  $Y$  tal que:

$$\forall m \geq 2, \mathbb{E}(|Y^m| | \mathcal{X} = \chi) < \sigma_m(\chi) < \infty \text{ con } \sigma_m(\cdot) \text{ continua en } \chi \quad (4.9)$$

Este supuesto nos permite hacer frente a las variables no acotadas.

Usando la notación introducida por (3.8), para  $i = 1, 2, 3, \dots, n$ , recordamos que  $\Delta_i$  es la cantidad definida como

$$\Delta_i = \frac{K(h^{-1}d(\chi, \mathcal{X}_i))}{\mathbb{E}K(h^{-1}d(\chi, \mathcal{X}_1))}.$$

Notar que el Lema 1 y el Lema 2, junto con (4.7), aseguran que  $\mathbb{E}K(h^{-1}d(\chi, \mathcal{X}_1)) > 0$ .

Sea  $\hat{r}_1(\chi)$  y sea  $\hat{r}_2(\chi)$  las siguientes cantidades:

$$\hat{r}_1(\chi) = \frac{1}{n} \sum_{i=1}^n \Delta_i \quad (4.10)$$

y

$$\hat{r}_2(\chi) = \frac{1}{n} \sum_{i=1}^n Y_i \Delta_i.$$

Veamos los siguientes lemas.

**Lema 3.** *Bajo (3.18) y (4.8) tenemos que:*

$$\lim_{n \rightarrow \infty} \mathbb{E} \hat{r}_2(\chi) = r(\chi).$$

**Demostración.** Tenemos que:

$$\begin{aligned} r(\chi) - \mathbb{E} \hat{r}_2(\chi) &= r(\chi) - \mathbb{E}(Y_1 \Delta_1), \\ &= r(\chi) - \mathbb{E}(\mathbb{E}(Y_1 \Delta_1 \mid \boldsymbol{\mathcal{X}}_1)), \\ &= r(\chi) - \mathbb{E}(r(\boldsymbol{\mathcal{X}}_1) \Delta_1), \\ &= \mathbb{E}((r(\chi) - r(\boldsymbol{\mathcal{X}}_1)) \Delta_1). \end{aligned}$$

Debido a que el soporte de la función Kernel  $K$  es  $[0, 1]$ , tenemos:

$$|r(\chi) - r(\boldsymbol{\mathcal{X}}_1)| \Delta_1 \leq \sup_{\chi' \in B(\chi, h)} |r(\chi) - r(\chi')| \Delta_1,$$

y la suposición de continuidad en  $r$  permite obtener el resultado.  $\square$

Como podemos ver el Lema 3 en una extensión de lo hecho en (3.4).

**Lema 4.** *Tenemos que:*

I. *Bajo las suposiciones (4.7), (4.8), (4.9), tenemos:*

$$\hat{r}_2(\chi) - \mathbb{E} \hat{r}_2 = O_{a.co} \left( \sqrt{\frac{\log n}{n \varphi_\chi(h)}} \right)$$

II. *Bajo las suposiciones (4.7) y (4.8), se cumple:*

$$\hat{r}_1(\chi) - 1 = O_{a.co} \left( \sqrt{\frac{\log n}{n \varphi_\chi(h)}} \right).$$

**Demostración.** .

I. para  $i = 1, 2, \dots, n$ ,  $K_i = K(h^{-1}d(\chi, \boldsymbol{\mathcal{X}}_i))$ . Para demostrar este lema haremos uso de la desigualdad de Bernsteins (ver Corolario A.1II). En efecto

$$\begin{aligned} P(|\hat{r}_2(\chi) - \mathbb{E}\hat{r}_2| > \epsilon) &= P\left|\frac{1}{n} \sum_{i=1}^n Y_i \Delta_i - \frac{1}{n} \sum_{i=1}^n \mathbb{E}Y_i \Delta_i\right| > \epsilon) \\ &= P\left(\frac{1}{n} \left| \sum_{i=1}^n (Y_i \Delta_i - \mathbb{E}Y_i \Delta_i) \right| > \epsilon\right) \end{aligned}$$

sólo tenemos que mostrar que  $\exists \epsilon_0 > 0$  tal que

$$\sum_{n \in \mathbb{N}} P\left(\frac{1}{n} \left| \sum_{i=1}^n (Y_i \Delta_i - \mathbb{E}Y_i \Delta_i) \right| > \epsilon_0 \sqrt{\frac{\log n}{n \varphi_\chi(h)}}\right) < \infty.$$

Por lo tanto, aplicamos el corolario A.1II) con  $Z_i = Y_i \Delta_i - \mathbb{E}Y_i \Delta_i$ , que son variables aleatorias reales centradas y con  $a^2 = \frac{1}{\varphi_\chi(h)}$ . Para hacer esto tenemos que probar que

$$\exists c > 0 \text{ tal que } \forall m \geq 2 \quad \mathbb{E}|(Y_1 \Delta_1 - \mathbb{E}Y_1 \Delta_1)^m| < c \varphi_\chi^{-m+1}(h)$$

en efecto

$$(Y_1 \Delta_1 - \mathbb{E}Y_1 \Delta_1)^m = \sum_{k=0}^m c_{m,k} (Y_1 \Delta_1)^k (\mathbb{E}Y_1 \Delta_1)^{m-k} (-1)^{m-k}$$

por otro lado

$$\mathbb{E}Y_1 \Delta_1 = \mathbb{E}(\mathbb{E}(Y_1 \Delta_1 \mid \boldsymbol{\mathcal{X}} = \chi)) = \mathbb{E}(\mathbb{E}(Y_1 \mid \boldsymbol{\mathcal{X}} = \chi) \Delta_1) = \mathbb{E}(r(\chi) \Delta_1) = r(\chi)$$

por lo tanto

$$|(Y_1 \Delta_1 - \mathbb{E}Y_1 \Delta_1)^m| \leq \sum_{k=0}^m c_{m,k} |Y_1 \Delta_1|^k |r(\chi)|^{m-k}$$

así que para  $c > 0$

$$\begin{aligned}
\mathbb{E}|(Y_1\Delta_1 - \mathbb{E}Y_1\Delta_1)^m| &\leq c \sum_{k=0}^m c_{k,m} \mathbb{E}|Y_1\Delta_1|^k |r(\chi)|^{m-k} \\
&\leq c \sum_{k=0}^m c_{k,m} \left( \max_{k=0,1,\dots,m} \mathbb{E}|Y_1\Delta_1|^k \right) |r(\chi)|^{m-k} \\
&= c \left( \max_{k=0,1,\dots,m} \mathbb{E}|Y_1\Delta_1|^k \right) \sum_{k=0}^m c_{k,m} |r(\chi)|^{m-k} \\
&= c \left( \max_{k=0,1,\dots,m} \mathbb{E}|Y_1\Delta_1|^k \right) (1 + |r(\chi)|)^m \\
&= C \left( \max_{k=0,1,\dots,m} \mathbb{E}|Y_1\Delta_1|^k \right).
\end{aligned}$$

Ahora veamos que para  $k \geq 2$

$$\mathbb{E}|Y_1\Delta_1|^k = O(\varphi_h(\chi)^{-k+1}). \quad (4.11)$$

En efecto,

$$\begin{aligned}
\mathbb{E}|Y_1|^k \Delta_1^k &= \mathbb{E}|Y_1|^k \frac{K_1^k}{(\mathbb{E}K_1)^k} = \frac{1}{(\mathbb{E}K_1)^k} \cdot \{\mathbb{E}|Y_1|^k K_1^k\} \\
&= \frac{1}{(\mathbb{E}K_1)^k} \cdot \mathbb{E}(\mathbb{E}(|Y_1|^k K_1^k \mid \mathcal{X})) \\
&= \frac{1}{(\mathbb{E}K_1)^k} \cdot \mathbb{E}(\mathbb{E}(|Y_1|^k \mid \mathcal{X}) K_1^k) \\
&< \frac{1}{(\mathbb{E}K_1)^k} \cdot \{\mathbb{E}(\sigma_m(\mathcal{X}) K_1^k)\} \\
&= \frac{1}{(\mathbb{E}K_1)^k} \cdot \{\mathbb{E}((\sigma_m(\mathcal{X}) - \sigma_m(\chi)) K_1^k) + \sigma_m(\chi) \mathbb{E}K_1^k\} \\
&= \mathbb{E}((\sigma_m(\mathcal{X}) - \sigma_m(\chi)) \Delta_1^k) + \sigma_m(\chi) \mathbb{E}\Delta_1^k
\end{aligned}$$

lo cual implica que

$$\begin{aligned}
|\mathbb{E}|Y_1|^k \Delta_1^k| &\leq \mathbb{E}|\sigma_m(\mathcal{X}) - \sigma_m(\chi)| \Delta_1^m + \sigma_m(\chi) \mathbb{E}\Delta_1^k \\
&\leq \left( \sup_{\chi' \in B(\chi, h)} |\sigma_m(\chi') - \sigma_m(\chi)| \right) \mathbb{E}\Delta_1^m + \sigma_m(\chi) \mathbb{E}\Delta_1^k.
\end{aligned}$$

Ahora como  $\int K_1^m < \infty$ , y como  $K$  es de tipo I (resp. tipo II) entonces  $\frac{K_1^m}{\int K_1^m}$  es de tipo I (resp. de tipo II). Luego por el Lema 1 y el Lema 2,  $\exists c_1, c_2 > 0$  tal que

$$\begin{aligned} \int K_1^m c_1 \varphi_\chi(h) &\leq \mathbb{E}K_1^m \leq \int K_1^m c_2 \varphi_\chi(h) \\ C_1 \varphi_\chi(h) &\leq \mathbb{E}K_1^m \leq C_2 \varphi_\chi(h). \end{aligned}$$

Ahora como  $K_1$  es de tipo I (resp. de tipo II)  $\exists c'_1, c'_2 > 0$  tal que

$$c'_1 \varphi_\chi(h) \leq \mathbb{E}K_1 \leq c'_2 \varphi_\chi(h)$$

así que

$$c_1^m \varphi_\chi^m(h) \leq (\mathbb{E}K_1)^m \leq c_2^m \varphi_\chi^m(h)$$

por lo tanto

$$\frac{1}{c_2^m \varphi_\chi^m(h)} \leq \frac{1}{(\mathbb{E}K_1)^m} \leq \frac{1}{c_1^m \varphi_\chi^m(h)}$$

por consiguiente

$$\begin{aligned} \frac{C_1 \varphi_\chi^{-m+1}(h)}{c_2^m} &\leq \frac{\mathbb{E}K_1^m}{(\mathbb{E}K_1)^m} \leq \frac{C_2 \varphi_\chi^{-m+1}(h)}{c_1^m} \\ b_1 \varphi_\chi^{-m+1}(h) &\leq \mathbb{E}\Delta_1^m \leq b_2 \varphi_\chi^{-m+1}(h) \end{aligned}$$

lo cual implica que

$$\begin{aligned} |\mathbb{E}|Y_1^k| \Delta_1^k| &\leq \left( \sup_{\chi' \in B(\chi, h)} |\sigma_m(\chi') - \sigma^m(\chi)| + \sigma^m(\chi) \right) b_2 \varphi_\chi^{-m+1}(h) \\ &= C \varphi_\chi^{-m+1}(h) \end{aligned}$$

esto quiere decir que

$$\mathbb{E}|Y_1|^k \Delta_1^k = O(\varphi_\chi^{-m+1}(h)).$$

Por consiguiente

$$\mathbb{E}|(Y_1 \Delta_1 - \mathbb{E}Y_1 \Delta_1)^m| \leq C \left( \max_{k=0,1,\dots,m} \varphi_\chi^{-m+1}(h) \right) \text{ para } k \geq 2$$

para  $k=1$  tenemos que

$$\begin{aligned}
 |\mathbb{E}|Y_1|\Delta_1| &\leq \mathbb{E}|Y_1|\Delta_1 = \mathbb{E}(\mathbb{E}(|Y_1|\Delta_1 | \boldsymbol{\mathcal{X}} = \chi)) = \mathbb{E}(\mathbb{E}(|Y_1| | \boldsymbol{\mathcal{X}} = \chi)\Delta_1) \\
 &= \mathbb{E}(|r(\chi)|\Delta_1) \\
 &= |r(\chi)| \\
 &= c
 \end{aligned}$$

así que  $\mathbb{E}|Y_1|\Delta_1 = O(1)$

entonces debido a que  $\varphi_\chi(h)$  tiende a cero con  $n$ , se sigue que

$$\mathbb{E}|(Y_1\Delta_1 - \overline{\mathbb{E}Y_1\Delta_1})^m| = O(\varphi_\chi^{-m+1}(h))$$

y esto finaliza la prueba.

II. El resultado puede ser obtenido directamente de I tomando  $Y_1 = 1$  y viendo que

$$\mathbb{E}\hat{r}_1(\chi) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\Delta_i = \frac{1}{n} \sum_{i=1}^n 1 = \frac{n}{n} = 1.$$

y esto culmina con la prueba del lema. □

**Teorema 4.1.** *Bajo la condición de tipo-continuida (3.18) con la condición de probabilidad (4.7), si el estimador verifica(4.8) y si la variable de respuesta  $Y$  satisface (4.9), entonces tenemos:*

$$\lim_{n \rightarrow \infty} \hat{r}(\chi) = r(\chi), a.co.$$

**Demostración.** Tenemos que  $\hat{r}(\chi) = \frac{\hat{r}_2(\chi)}{\hat{r}_1(\chi)}$ . Nuestra prueba es basada en la siguiente descomposición

$$\hat{r}(\chi) - r(\chi) = \frac{1}{\hat{r}_1(\chi)} \{(\hat{r}_2(\chi) - \mathbb{E}\hat{r}_2(\chi)) - (r(\chi) - \mathbb{E}\hat{r}_2(\chi))\} - \frac{r(\chi)}{\hat{r}_1(\chi)} \{\hat{r}_1(\chi) - 1\}.$$

Los numeradores de esta descomposición son tratados directamente usando el Lema 3 y el Lema 4, mientras que los denominadores son tratados usando la parte II) del Lema 4 junto con la parte I) de la Proposición 4. Esto finaliza la prueba.  $\square$

#### 4.2.2. Estimador Mediana condicional

Ahora daremos el mismo tipo de resultado de consistencia para la mediana condicional funcional  $m(\chi)$  y su estimación Kernel  $\hat{m}(\chi)$ , definidas respectivamente en (3.13) y (4.3). Recordemos que el estimador *f.d.a* condicional  $\hat{F}_Y^{\mathcal{X}}$  fue definido en (3.19). Debido a que los resultados asintóticos son basados en un  $\chi \in E$  fijo, simplificaremos algunas notaciones como sigue:

$$\forall \chi \in E, F_Y^{\mathcal{X}}(\cdot) = F_Y^{\mathcal{X}}(\chi, \cdot), \quad \hat{F}_Y^{\mathcal{X}}(\cdot) = \hat{F}_Y^{\mathcal{X}}(\chi, \cdot) \quad y \quad F_Y^{\mathcal{X}}(\cdot) = F_Y^{\mathcal{X}}(\mathcal{X}, \cdot).$$

Notar que la parte funcional de  $\hat{F}_Y^{\mathcal{X}}$  es la misma como en el caso de regresión. Por lo tanto, es natural esperar que los supuestos necesarios para hacer frente a esta parte funcional sean los mismos que en la sección anterior. En cuanto a la parte escalar las siguientes restricciones en la función Kernel  $K_0 = H'$  y su ancho de banda  $g$  asociado son introducidas:

$$\left\{ \begin{array}{l} g \text{ es una sucesión positiva tal que } \lim_{n \rightarrow \infty} g = 0 \quad , \\ K_0 \text{ es de tipo 0.} \end{array} \right. \quad (4.12)$$

Nótese que esta es una condición muy débil en el Kernel. Asegura, de acuerdo con la Definición 3.2, que la función  $H$  es continua y estrictamente creciente sobre el conjunto  $\{u, 0 < K(u) < 1\}$ . Esto tiene la ventaja principal que la estimación mediana condicional Kernel pueda ser definido como la única solución de la ecuación

$$\hat{m}(\chi) = \hat{F}_Y^{\mathcal{X}-1}(1/2).$$

**Lema 5.** *Bajo el modelo tipo-continuidad definido por (3.19) y (4.7), y si el estimador Kernel satisface (4.8) y (4.12), tenemos*

$$\lim_{n \rightarrow \infty} \hat{F}_Y^X(y) = F_Y^X(y), a.co.$$

**Demostración.**

$$\hat{F}_Y^X(y) - F_Y^X(y) = \frac{1}{\hat{r}_1(\chi)} \{(\hat{r}_3(\chi, y) - \mathbb{E}\hat{r}_3(\chi, y)) - (F_Y^X(y) - \mathbb{E}\hat{r}_3(\chi, y))\} - \frac{F_Y^X(y)}{\hat{r}_1(\chi)} \{\hat{r}_1(\chi) - 1\},$$

donde  $\hat{r}_1(\chi)$  es definido como en (4.10) y donde

$$\begin{aligned} \hat{r}_3(\chi, y) = \hat{r}_1(\chi) \hat{F}_Y^X(y) &= \frac{1}{n} \sum_{i=1}^n \frac{K_i}{\mathbb{E}K_1} \cdot \frac{\sum_{i=1}^n K_i H(g^{-1}(y - Y_i))}{\sum_{i=1}^n K_i} \\ &= \frac{1}{n \mathbb{E}K_1} \cdot \sum_{i=1}^n K_i H(g^{-1}(y - Y_i)) \\ &= \frac{1}{n} \cdot \sum_{i=1}^n \frac{K_i}{\mathbb{E}K_i} H(g^{-1}(y - Y_i)) \\ &= \frac{1}{n} \cdot \sum_{i=1}^n \Delta_i H(g^{-1}(y - Y_i)) \\ &= \frac{1}{n} \cdot \sum_{i=1}^n \Delta_i \Gamma_i(y) \end{aligned}$$

con  $\Gamma_i(y) = H(g^{-1}(y - Y_i))$ .

Note que los denominadores son tratados usando Lema 4-II junto con la Proposición 4-I. Note también que el último término es tratado usando el Lema 4-II.

Finalmente el lema es probado siempre y cuando se pruebe que

$$\lim_{n \rightarrow \infty} \mathbb{E}\hat{r}_3(\chi, y) = F_Y^X(y) \tag{4.13}$$

y

$$\lim_{n \rightarrow \infty} \hat{r}_3(\chi, y) - \mathbb{E}\hat{r}_3(\chi, y) = 0, a.co. \tag{4.14}$$

- El hecho que  $\mathbb{E}\hat{r}_1(\chi) = 1$  y que  $K$  sea de soporte compacto conduce directamente a

$$\begin{aligned}
\mathbb{E}\hat{r}_3(\chi, y) - F_Y^\chi(y) &= \mathbb{E}\Delta_1\Gamma_1(y) - F_Y^\chi(y) \\
&= \mathbb{E}\Delta_1\Gamma_1(y) - \mathbb{E}\Delta_1F_Y^\chi(y) \\
&= \mathbb{E}(\Delta_1(\mathbb{E}(\Gamma_1(y) | \mathcal{X}) - F_Y^\chi(y))) \\
&= \mathbb{E}(\Delta_1 1_{B(\chi, h)}(\mathcal{X})(\mathbb{E}(\Gamma_1(y) | \mathcal{X}) - F_Y^\chi(y))). \quad (4.15)
\end{aligned}$$

Esta última esperanza puede ser calculada fácilmente mediante el Teorema de Fubini y usando el hecho de que  $H' = K_0$ .

$$\begin{aligned}
\mathbb{E}(\Gamma_1(y) | \mathcal{X}) &= \int_{\mathbb{R}} H(g^{-1}(u - Y)) dP(u | \mathcal{X}) = \int_{\mathbb{R}} \int_{-\infty}^{\frac{y-u}{g}} K_0(v) dv dP(u | \mathcal{X}) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K_0(v) 1_{[v, \infty]}(g^{-h}(y - u)) dv dP(u | \mathcal{X}) \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K_0(v) 1_{[v, \infty]}(g^{-h}(y - u)) dP(u | \mathcal{X}) dv
\end{aligned}$$

como  $u \leq g^{-1}(y - u) \Rightarrow vg \leq y - u \Rightarrow u \leq y - vg$

$$\begin{aligned}
\mathbb{E}(\Gamma_1(y) | \mathcal{X}) &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_0(v) 1_{[u, \infty]}(y - vg) dP(u | \mathcal{X}) dv \\
&= \int_{\mathbb{R}} K_0(v) \int_{-\infty}^{y - vg} dP(u | \mathcal{X}) dv \\
&= \int_{\mathbb{R}} K_0(v) F_Y^\chi(y - vg) dv.
\end{aligned}$$

Debido a que  $K_0$  integra hasta 1

$$\mathbb{E}(\Gamma_1(y) | \mathcal{X}) - F_Y^\chi(y) = \int_{\mathbb{R}} K_0(v)(F_Y^\chi(y - vg) - F_Y^\chi(y)) dv$$

por otro lado tenemos que

$$|F_Y^\chi(y - vg) - F_Y^\chi(y)| \leq |F_Y^\chi(y - vg) - F_Y^\chi(y - vg)| + |F_Y^\chi(y - vg) - F_Y^\chi(y)|.$$

Porque  $K_0$  tiene soporte en  $[-1, 1]$  y porque  $h$  y  $g$  tienden a cero, la propiedad de continuidad de  $F_Y^{\mathcal{X}}$  nos permite escribir

$$\lim_{n \rightarrow \infty} \sup_{v \in [-1, 1]} 1_{B(\chi, h)}(\mathcal{X}) |F_Y^{\mathcal{X}}(y - vg) - F_Y^{\mathcal{X}}(y)| = 0$$

y

$$\lim_{n \rightarrow \infty} \sup_{v \in [-1, 1]} (\mathcal{X}) |F_Y^{\mathcal{X}}(y - vg) - F_Y^{\mathcal{X}}(y)| = 0,$$

de tal manera que podemos escribir finalmente que

$$\lim_{n \rightarrow \infty} 1_{B(\chi, h)}(\mathcal{X}) |\mathbb{E}(\Gamma_1(y) | \mathcal{X}) - F_Y^{\mathcal{X}}(y - vg)| = 0.$$

Este último resultado junto con (4.15) y el hecho de que  $\mathbb{E}\Delta_1 = 1$  prueban (4.13).

• Probemos ahora (4.14).

En efecto, podemos descomponer este término como una suma de variables aleatorias reales con media cero

$$\hat{r}_3(\chi, y) - \mathbb{E}\hat{r}_3(\chi, y) = \frac{1}{n} \sum_{i=1}^n (T_i - \mathbb{E}T_i)$$

donde  $T_i = \Delta_i \Gamma_i(y)$ .

Usando el Lema 3 o Lema 4 de acuerdo con el hecho de que  $K$  es de tipo I o II, y porque  $\Gamma_i(y)$  es acotada, tenemos:

$$T_i \leq C/\varphi_\chi(y)$$

por otro lado aplicando el resultado (4.11) para  $k = 2$ , tenemos

$$\mathbb{E}T_i \leq C\mathbb{E}\Delta_i^2 \leq C/\varphi_\chi(y).$$

Debido a que las variables  $T_i$  son acotadas, las dos desigualdades anteriores son suficientes para tratar el término  $T_i$ . Precisamente ahora estamos en condiciones de aplicar la desigualdad exponencial de tipo Bernstein dada por el corolario A.2-II. y

obtenemos

$$\hat{r}_3(\chi, y) - \mathbb{E}\hat{r}_3(\chi, y) = O_{a.co} \left( \frac{\log n}{n\varphi(h)} \right)^{1/2}.$$

□

**Teorema 4.2.** *Bajo las condiciones del Lema 5 tenemos que para cualquier punto real fijo  $y$ :*

$$\lim_{n \rightarrow \infty} \hat{m}(\chi) = m(\chi), a.co.$$

**Demostración.** La condición (4.12) asegura que el estimador *f.d.a* condicional  $\hat{F}_Y^\chi(\cdot)$  es continuo y estrictamente creciente. Por lo tanto, la función  $\hat{F}_Y^{\chi^{-1}}(\cdot)$  existe y es continua. La propiedad de continuidad de  $\hat{F}_Y^{\chi^{-1}}(\cdot)$  en el punto  $\hat{F}_Y^\chi(m(\chi))$  se puede escribir como:

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y, |\hat{F}_Y^\chi(y) - \hat{F}_Y^\chi(m(\chi))| \leq \delta(\epsilon) \Rightarrow |y - m(\chi)| \leq \epsilon,$$

y en el caso especial cuando  $y = \hat{m}(\chi)$ , tenemos

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, |\hat{F}_Y^\chi(\hat{m}(\chi)) - \hat{F}_Y^\chi(m(\chi))| \leq \delta(\epsilon) \Rightarrow |\hat{m}(\chi) - m(\chi)| \leq \epsilon,$$

así que

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, P(|\hat{F}_Y^\chi(\hat{m}(\chi)) - \hat{F}_Y^\chi(m(\chi))| \leq \delta(\epsilon)) \leq P(|\hat{m}(\chi) - m(\chi)| \leq \epsilon),$$

y finalmente obtenemos que

$$\begin{aligned} \forall \epsilon > 0, \exists \delta(\epsilon) > 0, P(|\hat{m}(\chi) - m(\chi)| > \epsilon) &\leq P(|\hat{F}_Y^\chi(\hat{m}(\chi)) - \hat{F}_Y^\chi(m(\chi))| > \delta(\epsilon)) \\ &= P(|F_Y^\chi(m(\chi)) - \hat{F}_Y^\chi(m(\chi))| > \delta(\epsilon)), \end{aligned}$$

la última desigualdad se obtiene siguiendo la observación

$$\hat{F}_Y^\chi(\hat{m}(\chi)) = F_Y^\chi(m(\chi)) = 1/2.$$

La convergencia puntual casi completa del estimador *f.d.a* condicional Kernel  $\hat{F}_Y^\chi$  (ver el Lema 5) conduce directamente a

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} P(|\hat{m}(\chi) - m(\chi)| > \epsilon) < \infty,$$

y esto culmina con la prueba.  $\square$

### 4.2.3. Estimador Moda condicional

Por último daremos resultados de consistente para la moda condicional  $\theta(\chi)$  y su estimación Kernel  $\hat{\theta}(\chi)$ , que se definen respectivamente en (3.15) y (4.6). Debido a que nos centramos en la convergencia puntual en un  $\chi$  fijo en  $E$ , se considera la siguiente notación simplificada:

$$\forall \chi \in E, f_Y^\chi(\cdot) = f_Y^{\mathbf{x}}(\chi, \cdot), \quad \hat{f}_Y^\chi(\cdot) = \hat{f}_Y^{\mathbf{x}}(\chi, \cdot) \quad y \quad f_Y^{\mathbf{x}}(\cdot) = f_Y^{\mathbf{x}}(\mathbf{x}, \cdot).$$

Las suposiciones necesarias para el estimador densidad condicional Kernel  $\hat{f}_Y^{\mathbf{x}}$  definido por (3.20) se cierran a las introducidas en la sección anterior para estimar la *f.d.a* condicional  $\hat{F}_Y^{\mathbf{x}}$ . Los siguientes supuestos no restrictivos tienen que ser añadidos:

$$\left\{ \begin{array}{l} \exists C < \infty, \forall (x, x') \in \mathbb{R} \times \mathbb{R}, |K_0(x) - K_0(x')| \leq C|x - x'|, \\ \lim_{n \rightarrow \infty} \frac{\log n}{ng\varphi_\chi(\chi)} = 0 \quad y \quad \exists \zeta > 0, \lim_{n \rightarrow \infty} gn^\zeta = \infty. \end{array} \right. \quad (4.16)$$

**Lema 6.** *Bajo el modelo tipo-continuidad definido por (3.20) y (4.7), y si el estimador Kernel satisface (4.8), (4.12) y (4.16), tenemos para cualquier subconjunto compacto  $S \subset \mathbb{R}$ :*

$$\lim_{n \rightarrow \infty} \sup_{y \in S} |f_Y^\chi - \hat{f}_Y^\chi| = 0, \text{ a.co.}$$

**Demostración.**

$$\hat{f}_Y^\chi(y) - f_Y^\chi(y) = \frac{1}{\hat{r}_1(\chi)} \{(\hat{r}_4(\chi, y) - \mathbb{E}\hat{r}_4(\chi, y)) - (f_Y^\chi(y) - \mathbb{E}\hat{r}_4(\chi, y))\} - \frac{f_Y^\chi(y)}{\hat{r}_1(\chi)} \{\hat{r}_1(\chi) - 1\},$$

donde  $\hat{r}_1(\chi)$  es definido como en (4.10) y donde

$$\begin{aligned}
\hat{r}_4(\chi, y) = \hat{r}_1(\chi) \hat{f}_Y^\chi(y) &= \frac{1}{n} \sum_{i=1}^n \frac{K_i}{\mathbb{E}K_1} \cdot \frac{\sum_{i=1}^n K_i g^{-1} K_0(g^{-1}(y - Y_i))}{\sum_{i=1}^n K_i} \\
&= \frac{1}{n \mathbb{E}K_1} \cdot \sum_{i=1}^n K_i g^{-1} K_0(g^{-1}(y - Y_i)) \\
&= \frac{1}{n} \cdot \sum_{i=1}^n \frac{K_i}{\mathbb{E}K_i} g^{-1} K_0(g^{-1}(y - Y_i)) \\
&= \frac{1}{n} \cdot \sum_{i=1}^n \Delta_i g^{-1} K_0(g^{-1}(y - Y_i)) \\
&= \frac{1}{n} \cdot \sum_{i=1}^n \Delta_i \Omega_i(y)
\end{aligned}$$

con  $\Omega_i(y) = g^{-1} K_0(g^{-1}(y - Y_i))$ .

Notar que los denominadores son tratados usando Lema 4-II junto con la Proposición 4-I. Notar también que el último término es tratado usando el Lema 4-II y que la función  $f_Y^\chi$  es uniformemente acotada sobre  $y \in S$  (puesto que es continua en un conjunto compacto  $S$ ). Por lo tanto, basta con probar:

$$\lim_{n \rightarrow \infty} \frac{1}{\hat{r}_1(\chi)} \sup_{y \in S} |\mathbb{E} \hat{r}_4(\chi, y) - f_Y^\chi(y)| = 0, a.co, \quad (4.17)$$

y

$$\lim_{n \rightarrow \infty} \frac{1}{\hat{r}_1(\chi)} \sup_{y \in S} |\hat{r}_4(\chi, y) - \mathbb{E} \hat{r}_4(\chi, y)| = 0, a.co. \quad (4.18)$$

- Primero probaremos (4.17).

Debido a que  $\mathbb{E} \Delta_1 = 1$  y que  $K_0$  es integrado hasta 1, tenemos después de integrar

por sustitución.

$$\begin{aligned}
\mathbb{E}\hat{r}_4(\chi, y) - f_Y^\chi(y) &= \mathbb{E}\Delta_1\Omega_1(y) - f_Y^\chi(y) \\
&= \mathbb{E}(\Delta_1(\mathbb{E}(\Omega_1(y) \mid \mathcal{X}) - f_Y^\chi(y))) \\
&= \mathbb{E}(\Delta_1(\int_{\mathbb{R}} g^{-1}K_0(g^{-1}(y-u))f_Y^\chi(y) du - f_Y^\chi(y))) \\
&= \mathbb{E}(\Delta_1 \int_{\mathbb{R}} g^{-1}K_0(g^{-1}(y-u))(f_Y^\chi(u) - f_Y^\chi(y) du) \\
&= \mathbb{E}(\Delta_1 \int_{\mathbb{R}} K_0(v)(f_Y^\chi(y-vg) - f_Y^\chi(y) du) \\
&= \mathbb{E}(1_{B(\chi,y)}(\mathcal{X})\Delta_1 \int_{\mathbb{R}} K_0(v)(f_Y^\chi(y-vg) - f_Y^\chi(y) du)(4.19)
\end{aligned}$$

Además

$$|f_Y^\chi(y-vg) - f_Y^\chi(y)| \leq |f_Y^\chi(y-vg) - f_Y^\chi(y-vg)| + |f_Y^\chi(y-vg) - f_Y^\chi(y)|.$$

Como  $S$  es compacto, la función  $f_Y^\chi$  es uniformemente continua sobre  $y \in S$ , además  $K_0$  es de soporte compacto en  $[-1, 1]$ , y el modelo (3.20) es de tipo-continuidad, entonces

$$\lim_{n \rightarrow \infty} \sup_{y \in S} 1_{B(\chi,y)}(\mathcal{X})|f_Y^\chi(y-vg) - f_Y^\chi(y-vg)| = 0,$$

y

$$\lim_{n \rightarrow \infty} \sup_{y \in S} |f_Y^\chi(y-vg) - f_Y^\chi(y)| = 0.$$

Lo cual implica que

$$\lim_{n \rightarrow \infty} \sup_{y \in S} 1_{B(\chi,y)}(\mathcal{X})|f_Y^\chi(y-vg) - f_Y^\chi(y)| = 0. \quad (4.20)$$

Ahora combinando (4.19) y (4.20), junto con la positividad de  $\Delta_1$  y  $K_0$ , obtenemos

$$\sup_{y \in S} |\mathbb{E}\hat{r}_4(\chi, y) - f_Y^\chi(y)| = O(1),$$

esto combinado con el Lema 4-II y con la Proposición 4-I prueba (4.17).

• Queda por comprobar que (4.18) es verdadera. Usando la compacidad de  $S$ , se

puede escribir que  $S \subset \bigcup_{k=1}^{z_n} S_k$  donde  $S_k = (t_k - l_n, t_k + l_k)$  donde  $l_n$  y  $z_n$  pueden ser elegidos tal que:

$$l_n = Cz_n^{-1} Cn^{-2\zeta}. \quad (4.21)$$

Tomando  $t_y = \arg \min_{t \in (t_1, \dots, t_{z_n})} |y - t|$ , tenemos

$$\frac{1}{\hat{r}_1(\chi)} \sup_{y \in S} |\hat{r}_4(\chi, y) - \mathbb{E}\hat{r}_4(\chi, y)| = A_1 + A_2 + A_3, \quad (4.22)$$

donde

$$\begin{aligned} A_1 &= \frac{1}{\hat{r}_1(\chi)} \sup_{y \in S} |\hat{r}_4(\chi, y) - \hat{r}_4(\chi, t_y)|, \\ A_2 &= \frac{1}{\hat{r}_1(\chi)} \sup_{y \in S} |\hat{r}_4(\chi, t_y) - \mathbb{E}\hat{r}_4(\chi, t_y)|, \\ A_3 &= \frac{1}{\hat{r}_1(\chi)} \sup_{y \in S} |\mathbb{E}\hat{r}_4(\chi, t_y) - \mathbb{E}\hat{r}_4(\chi, y)|. \end{aligned}$$

La condición de continuidad de Hölder (4.16) nos permite escribir directamente:

$$\begin{aligned} |\hat{r}_4(\chi, y) - \hat{r}_4(\chi, t_y)| &\leq \frac{1}{n} \sum_{i=1}^n \Delta_i |\Omega_1(y) - \Omega_i(t_y)| \\ &= \frac{1}{ng} \sum_{i=1}^n \Delta_i |K_0(g^{-1}(y - Y_i)) - K_0(g^{-1}(t_y - Y_i))| \\ &\leq \frac{C}{ng} \sum_{i=1}^n \Delta_i \frac{|y - t_y|}{g} \\ &\leq C\hat{r}_1(\chi)l_n g^{-2}. \end{aligned}$$

Usando (4.21), se cumple que  $A_1 \leq C/(gn^\zeta)^2$  y (4.16) implica que:

$$\lim_{n \rightarrow \infty} A_1 = 0. \quad (4.23)$$

Si siguiendo argumentos similares, podemos escribir  $A_3 \leq \frac{C}{\hat{r}_1(\chi)(gn^\zeta)^2}$ , y de acuerdo con el Lema 4-II y con la Proposición 4-I, obtenemos:

$$\lim_{n \rightarrow \infty} A_3 = 0, a.co. \quad (4.24)$$

Mirando ahora el término  $A_2$  podemos escribir para cualquier  $\epsilon > 0$ :

$$\begin{aligned}
P(\sup_{y \in S} |\hat{r}_4(\chi, t_y) - \mathbb{E}\hat{r}_4(\chi, t_y)| > \epsilon) &= P(\max_{j=1, \dots, z_n} |\hat{r}_4(\chi, t_j) - \mathbb{E}\hat{r}_4(\chi, t_j)| > \epsilon) \\
&\leq z_n \max_{j=1, \dots, z_n} P(|\hat{r}_4(\chi, t_j) - \mathbb{E}\hat{r}_4(\chi, t_j)| > \epsilon) \\
&\leq z_n \max_{j=1, \dots, z_n} P\left(\left|\frac{1}{n} \sum_{i=1}^n (U_i - \mathbb{E}U_i)\right| > \epsilon\right) \tag{4.25}
\end{aligned}$$

donde  $U_i = \Delta_i \Omega_i(t_i)$ .

Usando el Lema 1 o el Lema 2 acorde con el hecho de que  $K$  es de tipo I o II, y porque  $\Omega_i(y) \leq C/g$ , tenemos:

$$|U_i| \leq C/(g\varphi_\chi(h)). \tag{4.26}$$

Por otro lado, se tiene, después de la integración por sustitución, y usando (4.20):

$$\begin{aligned}
\mathbb{E}U_i^2 &= \mathbb{E}(\Delta_i^2 \mathbb{E}(\Omega(t_j)^2 \mid \mathcal{X} = \chi)) \\
&= \mathbb{E}\left(\Delta_i \left(\int_{\mathbb{R}} g^{-2} K_0^2(g^{-1}(t_j - u)) f_Y^\chi(u) du\right)\right) \\
&= \mathbb{E}\left(\Delta_i g^{-1} \left(\int_{\mathbb{R}} K_0^2(z) f_Y^\chi(t_j + zg) dz\right)\right) \\
&\leq C \mathbb{E}\left(\Delta_i g^{-1} \left(\int_{\mathbb{R}} K_0^2(z) dz f_Y^\chi(t_j)\right)\right).
\end{aligned}$$

Debido a que  $f_Y^\chi$  es acotada (puesto que es continua sobre un subconjunto compacto  $S$ ) y aplicando el resultado (4.11) con  $k = 2$ , obtenemos:

$$\mathbb{E}U_i^2 \leq C/(g\varphi_\chi(h)). \tag{4.27}$$

Debido a que las variables  $U_i$  son acotadas, ahora estamos en condiciones de aplicar la desigualdad de Bernstein de tipo exponencial dada por el Corolario A.2-II (ver Apéndice A). Esta desigualdad junto con (4.25), (4.26) y (4.27) dan directamente:

$$P(\sup_{y \in S} |\hat{r}_4(\chi, t_y) - \mathbb{E}\hat{r}_4(\chi, t_y)| > \epsilon) \leq z_n \exp\{-Cn\epsilon^2 g\varphi_\chi(h)\}.$$

Pero usando (4.21) obtenemos:

$$P(\sup_{y \in S} |\hat{r}_4(\chi, t_y) - \mathbb{E}\hat{r}_4(\chi, t_y)| > \epsilon) \leq Cn^2 \exp\{-Cn\epsilon^2 g\varphi_\chi(h)\}.$$

Porque  $\log n / (ng\varphi_\chi(h))$  tiende a cero, obtenemos directamente:

$$\forall \epsilon > 0, \sum_{n=1}^{\infty} P(\sup_{y \in S} |\hat{r}_4(\chi, t_y) - \mathbb{E}\hat{r}_4(\chi, t_y)| > \epsilon) < \infty. \quad (4.28)$$

El denominador de  $A_2$  es tratado directamente usando de nuevo el Lema 4-II y la Proposición 4-I . Esto es suficiente para obtener

$$\lim_{n \rightarrow \infty} A_2 = 0, a.co. \quad (4.29)$$

Finalmente, el resultado (4.18) se obtiene de (4.22), (4.23), (4.24) y (4.29). Esto culmina con la prueba del lema.  $\square$

**Teorema 4.3.** *Bajo las condiciones del Lema 6 tenemos que:*

$$\lim_{n \rightarrow \infty} \hat{\theta}(\chi) = \theta(\chi).$$

**Demostración.** La condición (3.20) asegura que la densidad condicional  $f_Y^X(\cdot)$  es continua y estrictamente creciente en  $(\theta(\chi) - \epsilon, \theta(\chi))$ . Por lo tanto, la función  $f_Y^{X^{-1}}(\cdot)$  existe y es continua. La propiedad de continuidad de  $f_Y^{X^{-1}}(\cdot)$  en el punto  $f_Y^X(\theta(\chi))$  se puede escribir como:

$$\forall \epsilon > 0, \exists \delta_1(\epsilon) > 0, \forall y \in (\theta(\chi) - \epsilon, \theta(\chi)), |f_Y^X(y) - f_Y^X(\theta(\chi))| \leq \delta_1(\epsilon) \Rightarrow |y - \theta(\chi)| \leq \epsilon,$$

debido a que  $f_Y^X(\cdot)$  es continua y estrictamente decreciente en  $(\theta(\chi), \theta(\chi) + \epsilon)$ , entonces

$$\forall \epsilon > 0, \exists \delta_2(\epsilon) > 0, \forall y \in (\theta(\chi), \theta(\chi) + \epsilon), |f_Y^X(y) - f_Y^X(\theta(\chi))| \leq \delta_2(\epsilon) \Rightarrow |y - \theta(\chi)| \leq \epsilon,$$

combinando los dos resultados podemos decir

$$\forall \epsilon > 0, \exists \delta(\epsilon) > 0, \forall y \in (\theta(\chi) - \epsilon, \theta(\chi) + \epsilon), |f_Y^X(y) - f_Y^X(\theta(\chi))| \leq \delta(\epsilon) \Rightarrow |y - \theta(\chi)| \leq \epsilon.$$

puediendo ser  $(\delta(\epsilon) = \min\{\delta_1(\epsilon), \delta_2(\epsilon)\})$ .

Por construcción  $\hat{\theta}(\chi) \in (\theta(\chi) - \epsilon, \theta(\chi) + \epsilon)$ , así que:

$$\exists \delta(\epsilon) > 0, |f_Y^X(\hat{\theta}(\chi)) - f_Y^X(\theta(\chi))| \leq \delta(\epsilon) \Rightarrow |\hat{\theta} - \theta(\chi)| \leq \epsilon.$$

Por lo tanto  $\exists \delta(\epsilon) > 0$  tal que  $P(|\hat{\theta} - \theta(\chi)| > \epsilon) \leq P(|f_Y^X(\hat{\theta}(\chi)) - f_Y^X(\theta(\chi))| > \delta(\epsilon))$

por otra parte, tenemos que

$$\begin{aligned} |f_Y^X(\theta(\chi)) - f_Y^X(\hat{\theta}(\chi))| &= |f_Y^X(\theta(\chi)) - \hat{f}_Y^X(\theta(\chi))| + |\hat{f}_Y^X(\theta(\chi)) - f_Y^X(\hat{\theta}(\chi))| \\ &= |f_Y^X(\theta(\chi)) - \hat{f}_Y^X(\theta(\chi))| + |\hat{f}_Y^X(\hat{\theta}(\chi)) - f_Y^X(\hat{\theta}(\chi))| \\ &\leq 2 \sup_{y \in (\theta(\chi) - \epsilon, \theta(\chi) + \epsilon)} |\hat{f}_Y^X(y) - f_Y^X(y)|. \end{aligned}$$

Por consiguiente

$$P(|f_Y^X(\theta(\chi)) - f_Y^X(\hat{\theta}(\chi))| > \delta(\epsilon)) \leq P\left(\sup_{y \in (\theta(\chi) - \epsilon, \theta(\chi) + \epsilon)} |\hat{f}_Y^X(y) - f_Y^X(y)| > \delta(\epsilon)\right)$$

y por el lema 6 para  $S = [\theta(\chi) - \epsilon, \theta(\chi) + \epsilon]$

$$\forall \epsilon > 0, \sum_{n \in \mathbb{N}} P(|\hat{\theta}(\chi) - \theta(\chi)| > \epsilon) < \infty.$$

□

#### 4.2.4. Estimador Cuantil condicional

Esta sección está dedicada a la generalización de los resultados dados para la estimación de cuantiles funcionales condicionales. El cuantil condicional de orden  $\alpha$ , denotada por  $t_\alpha(\chi)$ , y su estimador Kernel  $\hat{t}_\alpha(\chi)$  se definen respectivamente en (3.17) y (4.4). Como ya se ha señalado con la condición (4.12), la estimación cuantil

condicional Kernel puede ser definido como la única solución de la ecuación

$$\hat{t}_\alpha(\chi) = \hat{F}_Y^{\chi-1}(\alpha).$$

**Teorema 4.4.** *Sea  $\alpha \in (0, 1)$ . Bajo el modelo tipo-continuidad definido por (3.19) y (4.7), y si el estimador Kernel satisface (4.8) y (4.12), tenemos*

$$\lim_{n \rightarrow \infty} \hat{t}_\alpha(\chi) = t_\alpha(\chi), a.co.$$

**Demostración.** Esta prueba está estrechamente relacionada con la demostración del teorema 4.2, por tal razón será más brevemente presentada. Sea  $\epsilon > 0$  fijo. La propiedad de continuidad de  $\hat{F}_Y^{\chi-1}(\cdot)$  en el punto  $\hat{F}_Y^\chi(t_\alpha(\chi))$  se puede escribir como:

$$\exists \delta(\epsilon) > 0, \forall y, |\hat{F}_Y^\chi(y) - \hat{F}_Y^\chi(t_\alpha(\chi))| \leq \delta(\epsilon) \Rightarrow |y - t_\alpha(\chi)| \leq \epsilon.$$

Pero tomando  $y = \hat{t}_\alpha(\chi)$ , finalmente obtenemos el siguiente resultado

$$\begin{aligned} \exists \delta(\epsilon) > 0, P(|\hat{t}_\alpha(\chi) - t_\alpha(\chi)| > \epsilon) &\leq P(|\hat{F}_Y^\chi(\hat{t}_\alpha(\chi)) - \hat{F}_Y^\chi(t_\alpha(\chi))| > \delta(\epsilon)) \\ &= P(|F_Y^\chi(t_\alpha(\chi)) - \hat{F}_Y^\chi(t_\alpha(\chi))| > \delta(\epsilon)), \end{aligned}$$

el cual, combinada con la convergencia puntual casi completa del estimador *f.d.a* Kernel (ver Lema 5) da lugar a la prueba.  $\square$

### 4.3. Reimplementación y Aplicación de los Estimadores

En esta sección nos dedicaremos a la creación y aplicación de los métodos de predicción mediante los estimadores presentados anteriormente en la sección 4.1. Se construirán con los procedimientos de selección automática de los parámetros de suavizado (Ancho de banda).

Para cada uno de los métodos funcionales, se presenta el estimador Kernel. En cuanto al método de Regresión, por ser el primero, desarrollado desde el punto de vista histórico, se proponen estimadores Kernel con diferente selección automática de ancho de banda.

Teniendo en cuenta que nuestro conjunto de datos de estudio son los datos Quimiométricos, consideraremos la familia de semi-métricas basada en las derivadas como se describe en el capítulo 2, una función Kernel de las usuales (Triangular, Cuadrático) y una función Kernel integrado (Triangular, Cuadrático) que se describen en el capítulo 3.

Recordemos que nuestro problema de predicción corresponde a la situación en la que observamos  $n$  pares  $(\mathbf{x}_i, y_i)_{i=1, \dots, n}$  independientes e indenticamente distribuidos, donde  $\mathbf{x}_i = \{\chi_i(t_1), \dots, \chi_i(t_j)\}$  es la versión discretizada de la curva  $\chi_i = \{\chi_i(t), t \in \mathbb{R}\}$  valorada en los puntos  $t_1, \dots, t_j$ , mientras que los  $y_i$  son las respuestas escalares.

Además, como hemos dicho nuestra medida de cercanía entre curvas  $\chi_i, \chi_j$ , será  $d_q^{deriv}(\chi_i, \chi_j)$  la familia de semi-métricas basada en las derivadas. Así, que nuestro problema estadístico consiste en la predicción de las respuestas de las curvas.

#### 4.3.1. Estimación por Regresión

Está claro que el principal objetivo es estimar la cantidad

$$\hat{y} = \hat{r}(\mathbf{x}) = \frac{\sum_{i=1}^n y_i K(h^{-1} d_q^{deriv}(\mathbf{x}, \mathbf{x}_i))}{\sum_{i=1}^n K(h^{-1} d_q^{deriv}(\mathbf{x}, \mathbf{x}_i))}, \quad (4.30)$$

donde  $(\mathbf{x}_i, y_i)_{i=1, \dots, n}$  son los pares observados y  $\mathbf{x}$  es una curva observada en la que la regresión se estima.

En primer lugar, se logrará la predicción de la curva observada  $\mathbf{x}$  mediante la construcción de un promedio ponderado de los  $y_i$ 's para el cual los correspondientes  $\mathbf{x}_i$  son tal que la cantidad  $d_q^{deriv}(\mathbf{x}, \mathbf{x}_i)$  es más pequeña que el parámetro de suavizado  $h$ . En segundo lugar, se considerará una versión modificada en la que se reemplaza el ancho de banda  $h$  por el número  $k$  de  $\mathbf{x}_i$ 's que se toman en consideración para calcular el promedio ponderado. Para la estimación del Kernel y de los k-NN, se realizan varios procedimientos, el más básico es el caso cuando se fija el parámetro

de suavizado  $h$  o  $k$ .

- **Estimación del Estimador Kernel Funcional sin selección automática del ancho de banda**

En este caso, se debe fijar el parámetro  $h$ , el cual se da de antemano. Además, la función Kernel a considerar es la Cuadrática. La Rutina *Opnopare.kernel* calcula la estimación para cualquier curva, ver ApéndiceA.4.

- **Estimación del Estimador Kernel Funcional con selección automática del ancho de banda**

En este caso,  $h = h_{opt}$ , donde  $h_{opt}$  se obtiene de los datos por un procedimiento de validación cruzada:

$$h_{opt} = \arg \min_h CV(h)$$

donde

$$CV(h) = \sum_{i=1}^n (y_i - r_{(-i)}(\mathbf{x}_i))^2$$

con

$$r_{(-i)}(\mathbf{x}) = \frac{\sum_{j=1, j \neq i}^n y_j K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_j)/h)}{\sum_{k=1, k \neq i}^n K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_k)/h)}$$

la función Kernel  $K$  es la cuadrática. La Rutina *Opnopare.kernel.cv* calcula las estimaciones, ver ApéndiceA.4.

- **Estimación del Estimador Kernel con un número fijo de vecinos**

A quí,  $h = h_k(\mathbf{x})$  donde  $h_k(\mathbf{x})$  es un ancho de banda para el cual hay exactamente  $k$  curvas entre las  $\mathbf{x}_i$ 's talque  $d_q^{deriv}(\mathbf{x}, \mathbf{x}_i) < h_k(\mathbf{x})$ . La constante  $k$  se fija de antemano y la función Kernel es la Cuadrática. La Rutina *Opnopare.kernel.knn* calcula la estimación para cualquier curva, ver ApéndiceA.4.

- **Estimación del Estimador Kernel con elección global del número de vecinos**

En este caso,  $h = h_{k_{opt}}(\mathbf{x})$  donde  $h_{k_{opt}}(\mathbf{x})$  es el ancho de banda correspondiente al del número óptio de vecinos por una validación cruzada.

$$k_{opt} = \arg \min_k GCV(k)$$

donde

$$GCV(k) = \sum_{i=1}^n (y_i - r_{(-i)}^{kNN}(\mathbf{x}_i))^2$$

con

$$r_{(-i)}^{kNN}(\mathbf{x}) = \frac{\sum_{j=1, j \neq i}^n y_j K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_j)/h_k(\mathbf{x}))}{\sum_{k=1, k \neq i}^n K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_k)/h_k(\mathbf{x}))}.$$

La Rutina *Opnopare.kernel.knn.gcv* calcula las estimaciones, ver ApéndiceA.4.

- **Estimación del Estimador Kernel con elección local del número de vecinos**

En este caso,  $h = h_{k_{opt}}(\mathbf{x}_{i_0})$  donde  $h_{k_{opt}}(\mathbf{x}_{i_0})$  es el ancho de banda correspondiente al del número óptio de vecinos a  $\mathbf{x}_{i_0}$ , donde  $i_0 = \arg \min_{i=1, \dots, n} d_q^{deriv}(\mathbf{x}, \mathbf{x}_i)$ , obtenido por:

$$k_{opt}(\mathbf{x}_{i_0}) = \arg \min_k \left| y_{i_0} - \frac{\sum_{j=1, j \neq i_0}^n y_j K(d_q^{deriv}(\mathbf{x}_{i_0}, \mathbf{x}_j)/h_k(\mathbf{x}_{i_0}))}{\sum_{k=1, k \neq i_0}^n K(d_q^{deriv}(\mathbf{x}_{i_0}, \mathbf{x}_k)/h_k(\mathbf{x}_{i_0}))} \right|$$

La Rutina *Opnopare.kernel.knn.lcv* calcula las estimaciones, ver ApéndiceA.4.

### 4.3.2. Estimación de la Mediana Condicional

A diferencia de las técnicas de regresión anteriores, este tipo de estimador introduce un parámetro de suavizado para la respuesta además del que se necesita para las curvas. Sólo se ha desarrollado el procedimiento más sofisticado, que es el que involucra elección local automática de ancho de banda, es decir con elección local del número de vecinos.

El objetivo principal es calcular la cantidad:

$$\forall \alpha \in (0, 1), \quad y_{\alpha}^{kNN}(\mathbf{x}) = \inf_{y \in S} \{F_{k_{opt}, \kappa_{opt}}^{kNN}(\mathbf{x}, y) \geq \alpha\},$$

con

$$F_{k_{opt}, \kappa_{opt}}^{kNN}(\mathbf{x}, y) = \frac{\sum_{i \in I} K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_i)/h_k) H((y - y_i)/g_{\kappa})}{\sum_{i \in I} K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_i)/h_k)}.$$

$h_k$  es definido como anteriormente en regresión y  $g_{\kappa}$  es el ancho de banda para el cual hay exactamente  $\kappa$  entre las respuestas  $y_i$ 's tal que  $|y_i - y| < g_{\kappa}$ .

Con el fin de obtener los números óptimos de vecinos ( $k_{opt}(\mathbf{x})$  y  $\kappa_{opt}(y)$ ) aleatoriamente dividimos la muestra en dos submuestras de aprendizaje:

$$(\mathbf{x}_{i_1}, y_{i_1})_{i_1 \in I_1}, (\mathbf{x}_{i_2}, y_{i_2})_{i_2 \in I_2}, I_1 \cap I_2 = \emptyset, I_1 \cup I_2 = I$$

y definimos para cada  $\mathbf{x}$ ,  $i^* = arg \min_{i_2 \in I_2} d_q^{deriv}(\mathbf{x}, \mathbf{x}_{i_2})$ . Entonces, calculamos  $k_{opt}$  y  $\kappa_{opt}$  como sigue:

$$(k_{opt}, \kappa_{opt}) = arg \min_{(k, \kappa)} \left| y_{i^*} - \inf_{u \in S} \{F_{k, \kappa}^{kNN}(\mathbf{x}_{i^*}, u) \geq \alpha\} \right|.$$

La constante  $\alpha$  se fija de antemano y la función Kernel es la cuadrática. La Rutina *Opnopa.f.d.a.cuantil.lcv* calcula las estimaciones para  $\alpha = 0.05, 0.5, 0.95$ , ver Apéndice A.4.

### 4.3.3. Estimación de la Moda Condicional

Como en el caso anterior, el estimador necesita la selección de dos parámetros de suavizado en términos de  $k$ -vecinos más cercanos de una manera local (el número  $k$  puede variar de una unidad a otra). El procedimiento impulsado por los datos para la elección de estos parámetros también involucra dos sub-muestras de aprendizaje; una para la construcción del estimador Kernel, y otra para la selección de los parámetros de suavizado.

El objetivo principal es calcular la cantidad:

$$\theta^{kNN}(\mathbf{x}) = \arg \sup_{y \in S} f_{k_{opt}, \kappa_{opt}}^{kNN}(\mathbf{x}, y),$$

con

$$f_{k, \kappa}^{kNN}(\mathbf{x}, y) = \frac{\sum_{i \in I} K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_i)/h_k) K_0((y - y_i)/g_\kappa)}{g_\kappa \sum_{i \in I} K(d_q^{deriv}(\mathbf{x}, \mathbf{x}_i)/h_k)}.$$

$h_k$  y  $g_\kappa$  son definidos con en caso anterior. Entonces, calculamos  $k_{opt}$  y  $\kappa_{opt}$  como sigue:

$$(k_{opt}, \kappa_{opt}) = \arg \min_{(k, \kappa)} \left| y_{i^*} - \arg \sup_{u \in S} f_{k, \kappa}^{kNN}(\mathbf{x}_{i^*}, u) \right|.$$

La función Kernel es la cuadrática. La Rutina *Opnopa.f.d.moda.lcv* calcula las estimaciones, ver Apéndice [A.4](#).

#### 4.3.4. Prediciendo el contenido de grasa de las curvas de espectrometría

Como lo hemos mencionado las rutinas dadas anteriormente, se aplicaron al conjunto de datos de espectrometría, y para nuestro interés las tres rutinas usadas son basadas en la selección local del número de vecinos por ser las más sofisticadas. Para cada unidad  $i$  (entre 215 piezas de carne finamente picada), se observa una curva de espectrometría ( $\mathbf{x}_i$ ), que corresponde a la absorbancia medida a 100 canales de longitud de onda (es decir,  $\mathbf{x}_i = (\chi(\lambda_1), \dots, \chi(\lambda_{100}))$ ). Además, para cada unidad  $i$ , tenemos a la mano su contenido de grasa  $y_i$  obtenido por un proceso químico analítico. Dada una nueva curva  $\mathbf{x}$  de espectrometría, nuestra tarea principal es predecir el contenido de materia grasa  $\hat{y}$  correspondiente.

##### Medida de Rendimiento

Con el fin de destacar el desempeño de estos métodos de predicción funcional, dividimos nuestra muestra original en dos submuestras.

La primera, llamada muestra de aprendizaje, contiene las primeras 160 unidades  $((\mathbf{x}_i, y_i)_{i=1, \dots, 160})$ .

La segunda, llamada muestra de prueba, contiene las últimas 55 unidades  $((\mathbf{x}_i, y_i)_{i=161, \dots, 215})$ . La muestra de aprendizaje permite construir el estimador Kernel funcional con parámetros de suavizado óptimos; tanto los  $\mathbf{x}_i^s$  y la correspondiente  $y_i^s$  se utilizan en esta etapa. La muestra de prueba es útil para lograr predicciones y la medición de su calidad; evaluamos el estimador Kernel funcional (obtenido con la muestra de aprendizaje) en  $\mathbf{x}_{161}, \dots, \mathbf{x}_{215}$  ( $y_{161}, \dots, y_{215}$  se ignoran), el cual nos permite obtener las respuestas predichas  $\hat{y}_{161}, \dots, \hat{y}_{215}$ .

Para medir el rendimiento de cada método de predicción funcional, se considera

- la distribución de los *Errores Cuadrados* :  $se_i = (y_i - \hat{y}_i)^2, i = 162, \dots, 215$ .
- el *Error Cuadrado Medio Empírico* :  $MSE = \frac{1}{55} \sum_{i=161}^{215} se_i$ .
- ANOVA (one way) y prueba de Kruskal-Wallis, para las respuestas verdaderas y las predichas por los tres métodos de predicción.

Finalmente, se ejecutan las tres rutinas

*Opnopare.kernel.knn.lcv*, *Opnopa.f.d.a.cuantil.lcv*, *Opnopa.f.d.moda.lcv* en el conjunto de datos de espectrometría, correspondientes a los tres métodos de predicción: la esperanza condicional (Es decir, regresión), la mediana condicional y la moda condicional.

```
file("http://www.lsp.ups-tlse.fr/staph/npfda/npfda-spectrometric.dat", open="r")
DATA<-read.table("http://www.math.univ-toulouse.fr/staph/npfda/npfda-
spectrometric.dat", header=F)
CURVASPEC <- DATA[,1:100]           # muestra de curvas
Respuestas <-DATA[,101]            # muestras de respuestas
Aprendizaje <- 1:160
Prueba <- 161:215
CURVASPEC1 <- CURVASPEC[Aprendizaje,] # Muestra de curvas de aprendizaje
CURVASPEC2 <- CURVASPEC[Prueba,]     # Muestra de curvas de prueba
```

```

Resp1 <- Respuestas[Aprendizaje]      # Muestras de respuestas de aprendizaje
Resp2 <- Respuestas[Prueba]          # Muestra de respuestas de prueba

```

```

prediccion.regresion=0pnopare.kernel.knn.lcv(Resp1,CURVASPEC1,CURVASPEC2)

```

```

prediccion.mediana=0pnopa.f.d.a.cuartil.lcv(Resp1,CURVASPEC1,CURVASPEC2)

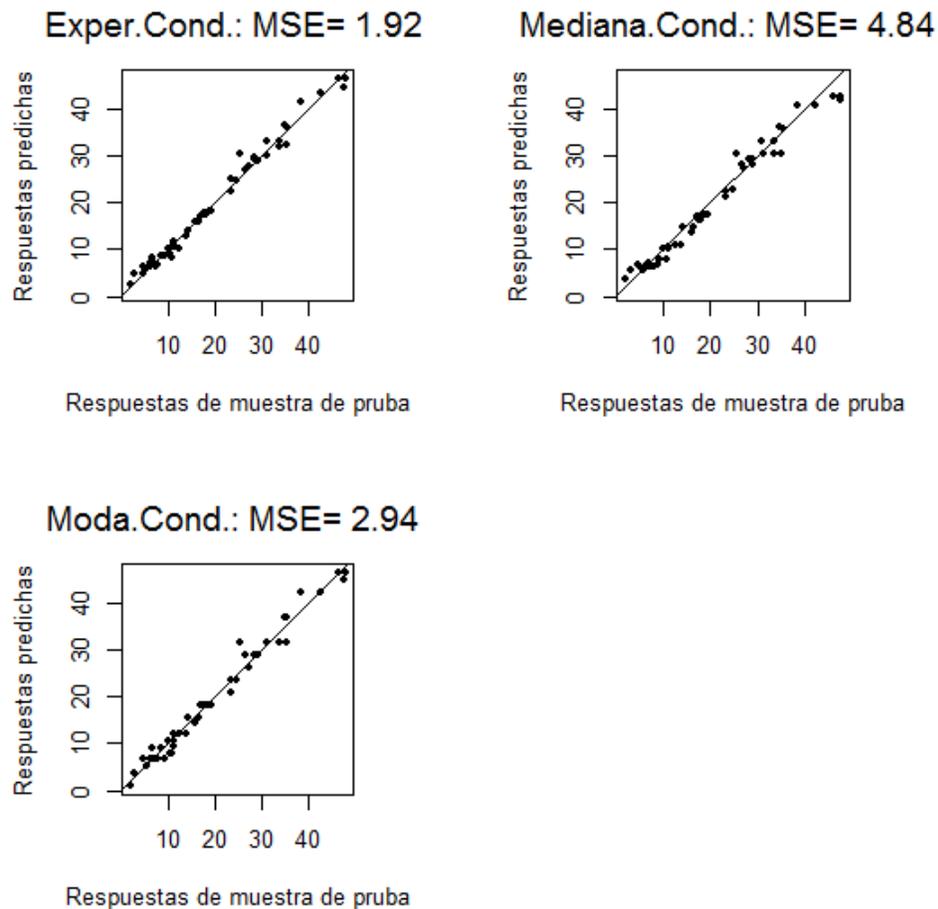
```

```

prediccion.moda=0pnopa.f.d.moda.lcv(Resp1,CURVASPEC1,CURVASPEC2)

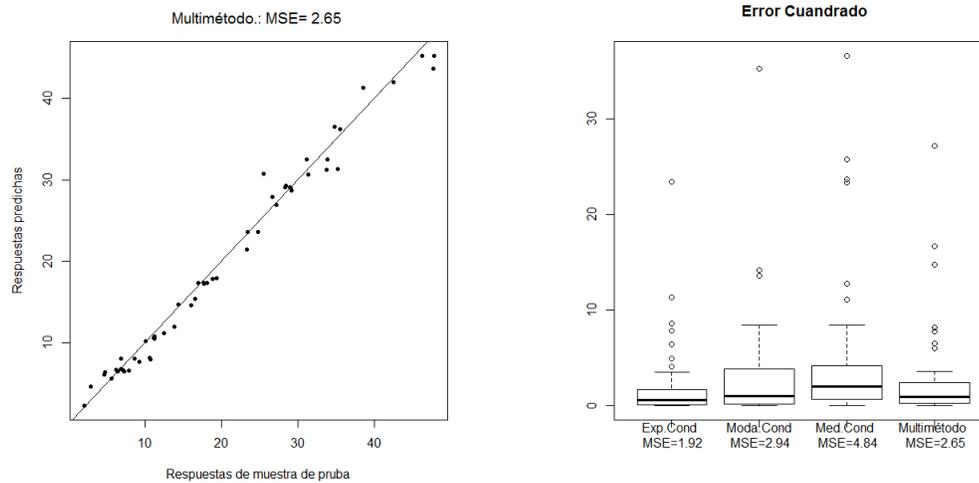
```

Los resultados se resumen en la Figura 4-2, como podemos ver los tres métodos dan buenas predicciones.



**Figura 4-2:** Rendimiento de los tres métodos de predicción

Debido a que existen algunas diferencias de un método a otro, una forma de mejorar o no los resultados es producir predicciones promediando los obtenidos con cada método, la metodología correspondiente se denomina multimétodos. Como se muestra en la figura 4-3, el resultado es muy interesante.



**Figura 4-3:** Comparando los tres métodos de predicción y el método multimétodo

No es que haya un aumento significativo en términos del error cuadrado medio pero si en cuanto a la dispersión del error cuadrado, pero aún así sigue siendo una alternativa como método adicional.

## Capítulo 5

# CONCLUSIONES Y TRABAJOS FUTUROS

### 5.1. Conclusiones

De los tres métodos de predicción desarrollados anteriormente podemos deducir varios aspectos interesantes:

- Son de fácil implementación y dieron buenos resultados de predicción en el experimento realizado.
- Se obtubieron mejores resultados considerando la derivada de segundo orden para la familia de semi-métrica basada en las derivadas.
- Hubo menos dispersión del error cuadrado para el método de predicción basado en la esperanza condicional.
- Haciendo prueba de ANOVA y prueba no paramétrica de Kruskal-Willes no hubo diferencias significativas entre los tres métodos de predicción.
- Se pueden lograr mejoras sustanciales en términos de errores de predicción mediante el uso de las tres técnicas de predicción funcional (por ejemplo, a través de la media de los tres predictores).

- La convergencia de los estimadores fueron demostradas en términos de convergencia casi completa.

## 5.2. Trabajos futuros

- Investigar el problema de predicción pero considerando que tanto la variable predictora como la variable de respuesta sean funcionales.
- Desarrollar una forma de optimizar la familia de semi-métricas a ser considerada como medida de cercanía entre curvas.

# APÉNDICES

# Apéndice A

## ALGUNOS RESULTADOS PROBABILÍSTICOS E IMPLEMENTACIÓN

### A.1. Definiciones y Teoremas Clásicos

**Definición A.1.** Sea  $(\Omega, \mathcal{A})$  un espacio medible y sea  $\{X_n\}_{n \in \mathbb{N}}$  una sucesión de variables aleatorias sobre  $(\Omega, \mathcal{A})$

$$\begin{aligned}
 a. \quad & \left\{ \begin{array}{l} \sup_n X_n : \Omega \rightarrow \mathbb{R} \\ (\sup_n X_n)(s) = \sup_n X_n(s) \end{array} \right. \\
 b. \quad & \left\{ \begin{array}{l} \inf_n X_n : \Omega \rightarrow \mathbb{R} \\ (\inf_n X_n)(s) = \inf_n X_n(s) \end{array} \right.
 \end{aligned}$$

**Definición A.2.** Sea  $\{A_n\}_{n \in \mathbb{N}}$  una sucesión de eventos en  $\mathcal{A}$

$$\limsup_n A_n = \bigcap_{m=1} \left( \bigcup_{n=m} A_n \right)$$

$$\liminf_n A_n = \bigcup_{m=1} \left( \bigcap_{n=m} A_n \right)$$

**Teorema A.1. Lema de Borel-Cantelli's.**

Sea  $(\Omega, \mathcal{A}, P)$  un espacio de probabilidad y sea  $\{A_n\}_{n \in \mathbb{N}}$  una sucesión de eventos en  $\mathcal{A}$  tal que  $\sum_n P(A_n) < \infty$  entonces  $P(\limsup_n A_n) = 0$

**Teorema A.2.** Sea  $X, Y, Z$  variables aleatorias reales,  $a, b \in \mathbb{R}$ , y  $g : \mathbb{R} \rightarrow \mathbb{R}$ .

Asumiendo que todas las siguiente esperanzas existen, tenemos que:

1.  $\mathbb{E}[a | Y] = a$
2.  $\mathbb{E}[aX + bZ | Y] = a\mathbb{E}[X | Y] + b\mathbb{E}[Z | Y]$
3.  $\mathbb{E}[X | Y] \geq 0$  si  $X \geq 0$ .
4.  $\mathbb{E}[X | Y] = \mathbb{E}[X]$  si  $X$  y  $Y$  son independientes
5.  $\mathbb{E}[\mathbb{E}[X | Y]] = \mathbb{E}[X]$
6.  $\mathbb{E}[Xg(Y) | Y] = g(Y)\mathbb{E}[X | Y]$ . En particular  $\mathbb{E}[g(Y) | Y] = g(Y)$
7.  $\mathbb{E}[X | Y, g(Y)] = \mathbb{E}[X | Y]$
8.  $\mathbb{E}[\mathbb{E}[X | Y, Z] | Y] = \mathbb{E}[X | Y]$

## A.2. Convergencia casi Completa

**Definición A.3.** Decimos que una sucesión  $(X_n)_{n \in \mathbb{N}}$  de variables aleatorias converge casi completamente a alguna v.a.r  $X$  si y solo si

$$\forall \epsilon > 0, \quad \sum_{n \in \mathbb{N}} P(|X_n - X| > \epsilon) < \infty$$

y la convergencia casi completa de  $(X_n)_{n \in \mathbb{N}}$  a  $X$  es denotada por  $\lim_{n \rightarrow \infty} X_n = X$ , a.co

Esta noción, puede ser llamada algunas veces como convergencia completa. La primera parte de la Proposición 1 a continuación estudiará la relación entre la convergencia casi completa y converge en probabilidad. Este último modo de convergencia se hace referencia a partir de ahora como convergencia  $p$  y es definido por la siguiente propiedad

$$\lim_{n \rightarrow \infty} X_n = X, p \iff \forall \epsilon > 0, \quad \lim_{n \rightarrow \infty} P(|X_n - X| > \epsilon) = 0.$$

La segunda parte de la proposición 1 hace lo mismo pero con la convergencia casi completa. Este último modo de convergencia es definido por la siguiente propiedad

$$\lim_{n \rightarrow \infty} X_n = X, a.s \iff P(\lim_{n \rightarrow \infty} X_n = X) = 1,$$

y será referido a partir de ahora como convergencia *a.s.*

**Teorema A.3.**  $\lim_{n \rightarrow \infty} X_n = X, a.s$  si y solo si  $P(\bigcap_{m=1}^{\infty} \{|X_n - X| < \epsilon\}) = 1$  para todo  $\epsilon > 0$

**Proposición 1.** Si  $\lim_{n \rightarrow \infty} X_n = X, a.co$ , entonces tenemos:

- I.  $\lim_{n \rightarrow \infty} X_n = X, p$
- II.  $\lim_{n \rightarrow \infty} X_n = X, a.s.$

**Demostración:** Sin pérdida de generalidad mostraremos el resultado para  $X = 0$ .

- I. Para todo  $\epsilon > 0$ , tenemos  $\sum_{n \in \mathbb{N}} P(|X_n| > \epsilon) < \infty$ , entonces por el lema de Borel-Cantelli's, tenemos que

$$P(\limsup_n \{|X_n| > \epsilon\}) = 0$$

y por consiguiente como tenemos que

$$\limsup_n P(|X_n| > \epsilon) \leq P(\limsup_n \{|X_n| > \epsilon\}) = 0$$

entonces

$$\limsup_n P(|X_n| > \epsilon) = 0$$

y por tanto

$$\lim_n P(|X_n| > \epsilon) = 0$$

- II. Para todo  $\epsilon > 0$ , tenemos que  $\sum_{n \in \mathbb{N}} P(|X_n| > \epsilon) < \infty$ . Por el lema de Borel-Cantelli's, se cumple que

$$P(\limsup_n \{|X_n| > \epsilon\}) = 0$$

$$P(\limsup_n \{|X_n| \leq \epsilon\}) = 1$$

esto equivale a decir que

$$P\left(\bigcap_{m=1}^{\infty} B_m\right) = 1$$

donde

$$B_m = \bigcup_{n=m}^{\infty} \{|X_n| \leq \epsilon\}$$

por lo tanto

$$P(\forall \epsilon, \exists n, \forall m > n, |X_n| \leq \epsilon) = 1.$$

□

En la literatura clásica, la razón de convergencia casi segura a cero(0) para una sucesión de v.a.r es definida por la condición

$$\begin{aligned} X_n - X = O_{a.s}(u_n) &\iff P(X_n - X = O(u_n)) = 1, \\ &\iff P(\exists C < \infty, \exists n, \forall m > n, |X_n - X| \leq C u_m) = 1. \end{aligned}$$

Que la razón de convergencia en probabilidad es definida por

$$X_n - X = O_p(u_n) \iff \lim_{m \rightarrow \infty} \limsup_n P(|X_n - X| < m u_n) = 1.$$

**Definición A.4.** Decimos que la razón de convergencia casi completa es de orden  $u_n$  si y solo si

$$\exists \epsilon_0 > 0 \quad \sum_{n \in \mathbb{N}} P(|X_n - X| > \epsilon_0 u_n) < \infty$$

y lo escribimos  $X_n - X = O_{a.co}(u_n)$

**Proposición 2.** Supongamos que  $X_n - X = O_{a.co}(u_n)$ , tenemos:

1.  $X_n - X = O_p(u_n)$
2.  $X_n - X = O_{a.s}(u_n)$

**Proposición 3.** Supongamos que  $\lim_{n \rightarrow \infty} u_n = 0$ ,  $\lim_{n \rightarrow \infty} X_n = l_x, a.co$  y  $\lim_{n \rightarrow \infty} Y_n = l_y, a.co$  donde  $l_x$  y  $l_y$  son dos números reales determinísticos.

I. tenemos:

- a.  $\lim_{n \rightarrow \infty} (X_n + Y_n) = l_x + l_y, a.co$
- b.  $\lim_{n \rightarrow \infty} X_n Y_n = l_x l_y, a.co$
- c.  $\lim_{n \rightarrow \infty} \frac{1}{Y_n} = \frac{1}{l_y}, a.co$ , siempre que  $l_y \neq 0$ .

II. si  $X_n - l_x = O_{a.co}(u_n)$  y  $Y_n - l_y = O_{a.co}(u_n)$  tenemos:

- a.  $(X_n + Y_n) - (l_x + l_y) = O_{a.co}(u_n), a.co$
- b.  $X_n Y_n - l_x l_y = O_{a.co}(u_n), a.co$
- c.  $\frac{1}{Y_n} - \frac{1}{l_y} = O_{a.co}(u_n), a.co$ , siempre que  $l_y \neq 0$ .

**Demostración:** Sea  $\epsilon > 0$

I. a. por hipótesis tenemos que

$$\sum_{n \in \mathbb{N}} P(|X_n - l_x| > \epsilon/2) < \infty \quad \text{y} \quad \sum_{n \in \mathbb{N}} P(|Y_n - l_y| > \epsilon/2) < \infty$$

por otra parte

$$\{|X_n + Y_n - (l_x + l_y)| > \epsilon\} \subset \{|X_n - l_x| > \epsilon/2\} \cup \{|Y_n - l_y| > \epsilon/2\}$$

por consiguiente

$$P(|X_n + Y_n - (l_x + l_y)| > \epsilon) \leq P(|X_n - l_x| > \epsilon/2) + P(|Y_n - l_y| > \epsilon/2)$$

y por tanto

$$\sum_{n \in \mathbb{N}} P(|X_n + Y_n - (l_x + l_y)| > \epsilon) \leq \sum_{n \in \mathbb{N}} P(|X_n - l_x| > \epsilon/2) + \sum_{n \in \mathbb{N}} P(|Y_n - l_y| > \epsilon/2) < \infty.$$

II. a. en esta último desigualdad tomemos  $\epsilon = \epsilon_0 u_n$ , entonces

$$\sum_{n \in \mathbb{N}} P(|X_n + Y_n - (l_x + l_y)| > \epsilon_0 u_n) < \infty.$$

ya que  $X_n - l_x = O_{a.co}(u_n)$  y  $Y_n - l_y = O_{a.co}(u_n)$

I. b. Sin pérdida de generalidad, consideremos solo el caso cuando  $l_x = 0$ .

Sea  $\epsilon > 0$  entonces

$$\begin{aligned}
 P(|X_n Y_n| > \epsilon) &= P(|X_n(Y_n - l_y) + X_n l_y| > \epsilon) \\
 &\leq P(|X_n||Y_n - l_y| > \epsilon/2) + P(|X_n l_y| > \epsilon/2) \\
 &\leq P(|X_n| > \sqrt{\epsilon/2}) + P(|Y_n - l_y| > \sqrt{\epsilon/2}) + P(|X_n l_y| > \epsilon/2) \\
 &\leq P(|X_n| > \epsilon/2) + P(|Y_n - l_y| > \epsilon/2) + P(|X_n l_y| > \epsilon/2)
 \end{aligned}$$

por lo tanto

$$\sum_{n \in \mathcal{N}} P(|X_n Y_n| > \epsilon) \leq \sum_{n \in \mathcal{N}} P(|X_n| > \epsilon/2) + \sum_{n \in \mathcal{N}} P(|Y_n - l_y| > \epsilon/2) + \sum_{n \in \mathcal{N}} P(|X_n l_y| > \epsilon/2) < \infty$$

II. b. de esta última ecuación tomemos  $\epsilon = \epsilon_0 u_n$  entonces

$$\begin{aligned}
 \sum_{n \in \mathcal{N}} P(|X_n Y_n| > \epsilon = \epsilon_0 u_n) &\leq \sum_{n \in \mathcal{N}} P(|X_n| > \epsilon = \epsilon_0 u_n/2) + \sum_{n \in \mathcal{N}} P(|Y_n - l_y| > \epsilon = \epsilon_0 u_n/2) \\
 &+ \sum_{n \in \mathcal{N}} P(|X_n l_y| > \epsilon = \epsilon_0 u_n/2) < \infty.
 \end{aligned}$$

Haciendo uso de la razón de convergencia  $X_n - l_x = O_{a.co}(u_n)$  y  $Y_n - l_y = O_{a.co}(u_n)$ .

I. c. La convergencia casi completa de  $Y_n$  a  $l_y$  implica que existe  $\delta > 0$  tal que

$$\sum_{n \in \mathcal{N}} P(|Y_n| \leq \delta) < \infty$$

entonces

$$\begin{aligned}
P\left(\left|\frac{1}{Y_n} - \frac{1}{l_y}\right| > \epsilon\right) &= P(|Y_n - l_y| > \epsilon|Y_n l_y|) \\
&\leq P(|Y_n - l_y| > \epsilon|Y_n l_y| \wedge |Y_n| > \delta) \\
&+ P(|Y_n - l_y| > \epsilon|Y_n l_y| \wedge |Y_n| \leq \delta) \\
&\leq P(|Y_n - l_y| > \epsilon|Y_n l_y| \wedge |Y_n| > \delta) \\
&+ P(|Y_n| < \frac{|l_y|}{\epsilon(|l_y| - 1)} \wedge |Y_n| \leq \delta) \\
&= P(|Y_n - l_y| > \epsilon|Y_n l_y| \wedge |Y_n| > \delta) \\
&+ P(|Y_n| \leq \delta) \\
&\leq P(|Y_n - l_y| > \epsilon\delta|l_y|) + P(|Y_n| \leq \delta)
\end{aligned}$$

y por tanto

$$\begin{aligned}
\sum_{n \in \mathcal{N}} P\left(\left|\frac{1}{Y_n} - \frac{1}{l_y}\right| > \epsilon\right) &\leq \sum_{n \in \mathcal{N}} P(|Y_n - l_y| > \epsilon\delta|l_y|) + \sum_{n \in \mathcal{N}} P(|Y_n| \leq \delta) \\
&< \infty
\end{aligned}$$

II. c. esta prueba sale directamente de la última desigualdad

□

**Proposición 4.** *Supongamos que  $\lim_{n \rightarrow \infty} u_n = 0$ ,  $\lim_{n \rightarrow \infty} Y_n = l_y$ , a.co donde  $l_y$  es un número real determinístico, muestre que:*

- I.  $X_n Y_n = O_{a.co}(u_n)$
- II.  $\frac{X_n}{Y_n} = O_{a.co}(u_n)$  siempre que  $l_y \neq 0$

**Demostración:** .

I. Como  $Y_n$  converge casi completamente a  $l_y$  implica que existe  $\delta > 0$  tal que

$$\sum_{n \in \mathcal{N}} P(|Y_n| > \delta) < \infty$$

entonces

$$\begin{aligned}
P(|X_n Y_n| > \epsilon u_n) &= P(|X_n Y_n| > \epsilon u_n \wedge |Y_n| \leq \delta) + P(|X_n Y_n| > \epsilon u_n \wedge |Y_n| > \delta) \\
&\leq P\left(|X_n| > \frac{\epsilon u_n}{|Y_n|} \wedge \frac{1}{|Y_n|} \geq \frac{1}{\delta}\right) + P(|X_n Y_n| > \epsilon u_n \delta) \\
&\leq P(|X_n| > \epsilon u_n \delta^{-1}) + P(|X_n Y_n| > \epsilon u_n \delta)
\end{aligned}$$

así que

$$\begin{aligned}
\sum_{n \in \mathcal{N}} P(|X_n Y_n| > \epsilon u_n) &\leq \sum_{n \in \mathcal{N}} P(|X_n| > \epsilon u_n \delta^{-1}) + \sum_{n \in \mathcal{N}} P(|X_n Y_n| > \epsilon u_n \delta) \\
&< \infty.
\end{aligned}$$

II. Por la parte (I.c) de la proposición 3  $\lim_{n \rightarrow \infty} \frac{1}{Y_n} = \frac{1}{l_y}, a.co$  y por la parte I de esta proposición  $\frac{X_n}{Y_n} = O_{a.co}(u_n)$

□

### A.3. Desigualdad exponencial para *v.a.r* independientes.

En toda esta sección  $Z_1, Z_2, \dots, Z_n$  serán r.r.v independiente con media cero(0). Como podemos ver a lo largo de este documento, la afirmación de propiedades de convergencia casi completa necesitan hallar cotas superiores para algunas probabilidades que envuelven suma de variables aleatorias tal como

$$P\left(\left|\sum_{i=1}^n Z_i\right| > \epsilon\right),$$

donde eventualmente el número real  $\epsilon$  decrece con  $n$ . En este contexto existe una potente herramienta probabilística, generalmente llamada desigualdad exponencial. Nuestro enfoque aquí es la llamada desigualdad de Bernstein's. Esta elección fue hecha porque la forma de la desigualdad de Bernstein's es más fácil para los desarrollos teóricos en un estadístico funcional que han sido establecidos a lo largo de este documento.

**Proposición 5.** *Supongamos que*

$$\forall m \geq 2, |\mathbb{E}Z_{i=1}^m| \leq (m!/2)(a_i)^2 b^{m-2}$$

*y sea*  $(A_n)^2 = (a_1)^2 + \dots + (a_n)^2$ , *entonces*

$$\forall \epsilon \geq 0, \quad P\left(\left|\sum_{i=1}^n Z_i\right| > \epsilon A_n\right) \leq 2 \exp\left\{-\frac{\epsilon^2}{2\left(1 + \frac{\epsilon b}{A_n}\right)}\right\}$$

**Corolario A.1.** .

I. *Si*  $\forall m \geq 2, \exists C_m > 0, \mathbb{E}|Z_1^m| \leq C_m a^{2(m-1)}$ , *entonces*

$$\forall \epsilon \geq 0, \quad P\left(\left|\sum_{i=1}^n Z_i\right| > \epsilon n\right) \leq 2 \exp\left\{-\frac{\epsilon^2 n}{2a^2(1 + \epsilon)}\right\}$$

II. *Supongamos que las variables dependen de*  $n$  *(esto es, supongamos que*  $Z_i = Z_{i,n}$ *).*

*Si*  $\forall m \geq 2, \exists C_m > 0, \mathbb{E}|Z_1^m| \leq C_m a_n^{2(m-1)}$  *y si*  $u_n = n^{-1} a_n^2 \log n$  *verifica que*

$\lim_{n \rightarrow \infty} u_n = 0$ , *entonces*

$$\frac{1}{n} \sum_{i=1}^n Z_i = O_{a.co}(\sqrt{u_n}).$$

**Corolario A.2.** .

I. *Si*  $\exists M < \infty, |Z_1| \leq M$ , *y denotando*  $\sigma^2 = \mathbb{E}Z_1^2$ , *tenemos*

$$\forall \epsilon \geq 0, \quad P\left(\left|\sum_{i=1}^n Z_i\right| > \epsilon n\right) \leq 2 \exp\left\{-\frac{\epsilon^2 n}{2a^2\left(1 + \epsilon \frac{M}{\sigma^2}\right)}\right\}$$

II. *Supongamos que las variables dependen de*  $n$  *(esto es, supongamos que*  $Z_i = Z_{i,n}$ *) y*

*son tal que*  $\exists M = M_n < \infty, |Z_1| \leq M$  *y definiendo*  $\sigma^2 = \mathbb{E}Z_1^2$ . *si*  $u_n = n^{-1} \sigma^2 \log n$

*verifica que*  $\lim_{n \rightarrow \infty} u_n = 0$ , *y si*  $M/\sigma_n^2 < C < \infty$ , *entonces tenemos*

$$\frac{1}{n} \sum_{i=1}^n Z_i = O_{a.co}(\sqrt{u_n}).$$

#### A.4. Algunas Rutinas

##### Operador Kernel sin selección automática de ancho de banda.

```

Opnopare.kernel <- function(Respuesta, CURVAS, PRED, ANCHO.DE.BANDA,
  ..., Tipo.de.kernel = "quadratic",semi.metrica = "semimetric.deriv")
{
#####
# Ejecuta la prediccion funcional(regresion) de una respuesta escalar
# dada una muestra de curvas a traves del estimador kernel funcional.
# Se considera un ancho de banda global sin seleccion automatica.
# "Respuesta" vector que contiene las observaciones de la respuesta escalar
# "CURVAS" matriz que contiene el conjunto de datos de las curvas
# (fila por fila) utilizado para la etapa de estimacion
# "PRED" matriz que contiene las nuevas curvas usadas para la prediccion
# "ANCHO.DE.BANDA" el valor del ancho de banda
# "..." argumentos necesarios para la llamada de la funcion de
# calculo de la semi-metrica entre las curvas
# "Tipo.de.kernel" la funcion kernel usada para el calculo del
# estimador kernel; se puede elegir "triangle" o "quadratic (por defecto)
# "semi.metrica" la funcion de calculo de la semi-metrica; se puede elegir
# "deriv" (default)
# Arroja una lista que contiene:
# "valor.estimado" vector que contiene las respuestas estimadas para
# cada curva de "CURVAS"
# "valor.predicho" si PRED es diferente de CURVAS, este vector
# contiene las respuestas predichas para cada curva de PRED
# "band" valor del ancho de banda concurrente
# "Mse" error cuadrado medio entre valores estimados y valores observados

```

```
#####

  Respuesta <- as.vector(Respuesta)
  if(is.vector(PRED)) PRED <- as.matrix(t(PRED))
  testparadim <- sum(dim(CURVAS)==dim(PRED))==2
  sdocondato <- T
  if(testparadim) sdocondato <- sum(CURVAS==PRED)!=prod(dim(CURVAS))
  sm <- get(semi.metrica)
  SEMIMETRIC1 <- sm(CURVAS, CURVAS, ...)
  kernel <- get(Tipo.de.kernel)
  KERNEL1 <- kernel(SEMIMETRIC1/ANCHO.DE.BANDA)
  KERNEL1[KERNEL1 < 0] <- 0
  KERNEL1[KERNEL1 > 1] <- 0
  diag(KERNEL1) <- 0
  RESPKERNEL1 <- KERNEL1 * Respuesta
  Denom1 <- apply(KERNEL1, 2, sum)
  if(sum(Denom1 == 0) > 0)
  stop(paste("probar un ancho de banda mayor que", ANCHO.DE.BANDA))
  Respuesta.estimada <- apply(RESPKERNEL1, 2, sum)/Denom1
  Mse.estimado <- sum((Respuesta.estimada - Respuesta)^2)/length(Respuesta)
  if(sdocondato) {
  SEMIMETRIC2 <- sm(CURVAS, PRED, ...)
  KERNEL2 <- kernel(SEMIMETRIC2/ANCHO.DE.BANDA)
  KERNEL2[KERNEL2 < 0] <- 0
  KERNEL2[KERNEL2 > 1] <- 0
  Denom2 <- apply(KERNEL2, 2, sum)
  if(sum(Denom2 == 0) > 0)
  stop(paste("probar un ancho de banda mayor que ", ANCHO.DE.BANDA))
  }
```

```

RESPKERNEL2 <- KERNEL2 * Respuesta
Respuesta.predicha <- apply(RESPKERNEL2, 2, sum)/Denom2
  Mse.predicho<- sum((Respuesta.predicha - Resp2)^2)/length(Resp2)
return(list(Valor.Estimado = Respuesta.estimada,
Valor.Predicho = Respuesta.predicha, band = ANCHO.DE.BANDA,
Mse = Mse.estimado, Mse1=Mse.predicho))
}else {
return(list(Valor.Estimado =Respuesta.estimada , band =
ANCHO.DE.BANDA, Mse = Mse.estimado))
}
}
}

```

#### Operador Kernel con selección automática de ancho de banda.

```

Opnopare.kernel.cv <-function(Respuesta,CURVAS,PRED, ...,Tipo.de.kernel
="quadratic",semi.metrica="semimetric.deriv",h.rango = NULL)
{
#####
# Realiza la prediccion funcional (regresion) de respuesta escalar
# de una muestra de curvas mediante el estimador kernel funcional.
# Un ancho de banda global se selecciona automaticamente
# con un procedimiento de validacion cruzada.
# "Respuesta" vector que contiene las respuestas escalares
# "CURVAS" matriz que contiene el conjunto de datos curvas (fila por fila)
# usados para la etapa de estimacion
# "PRED" matriz que contiene nuevas curvas almacenadas fila por fila
# usadas para el calculo de la prediccion
# "..." argumentos necesarios para la llamada de la funcion de
# calculo de la semi-metrica entre las curvas

```

```

# "Tipo.de.kernel" la funcion kernel usada para el calculo del
# estimador kernel; se puede elegir "triangle" o "quadratic (por defecto)
# "semi.metrica" calcula la distancia entre curvas
# "h.rango" un vector de longitud 2 que da el rango para el ancho de banda.
# Por defecto, el procedimiento define una secuencia de candidatos
# de anchos de banda de acuerdo con los valores de las curvas de matriz
# Arroja una lista que contiene:
# "valores.estimados" vector que contiene las respuestas estimadas
# para cada curva de "CURVAS"
# "valores.predichos" si PRED es diferente de CURVAS, este vector
# contiene las respuestas predichas para cada curva de PRED
# "hopt" valor optimo de ancho de banda
# "hseq" la secuencia utilizada de posibles anchos de banda
# "Mse" error cuadrado medio entre valores estimados y valores observados
#####
    Respuesta <- as.vector(Respuesta)
if(is.vector(PRED)) PRED <- as.matrix(t(PRED))
testparadim <- sum(dim(CURVAS)==dim(PRED))==2
sdocondato <- T
if(testparadim) sdocondato <- sum(CURVAS==PRED)!=prod(dim(CURVAS))
sm <- get(semi.metrica)
SEMIMETRIC1 <- sm(CURVAS, CURVAS, ...)
Semimetric1 <- SEMIMETRIC1[row(SEMIMETRIC1) > col(SEMIMETRIC1)]
if(is.null(h.rango)) {
h.seq <- quantile(Semimetric1, seq(0.05, 0.5, length = 20))
}else {
h.seq <- seq(h.rango[1], h.rango[2], length = 20)

```

```

}

kernel <- get(Tipo.de.kernel)

contar1 <- 0

contar2 <- 0

Mse <- 0

h.seq.correcto <- 0

h.seq.longitud <- length(h.seq)
  h <- h.seq[1]

while(count1 < h.seq.longitud) {
  contar1 <- contar1 + 1

  h <- 1.1 * h

  KERNEL1 <- kernel(SEMIMETRIC1/h)

  diag(KERNEL1) <- 0

  KERNEL1[KERNEL1 < 0] <- 0
  KERNEL1[KERNEL1 > 1] <- 0

  Denom1 <- apply(KERNEL1, 2, sum)
  Logic <- (Denom1 == 0)

  while(sum(Logic) >= 1) {
    h <- 1.1 * h

    KERNEL1 <- kernel(SEMIMETRIC1/h)

    diag(KERNEL1) <- 0

    KERNEL1[KERNEL1 < 0] <- 0
    KERNEL1[KERNEL1 > 1] <- 0

    Denom1 <- apply(KERNEL1, 2, sum)
    Logic <- (Denom1 == 0)
  }

  RESPKERNEL1 <- KERNEL1 * Respuesta

```

```

Respuesta.estimada <- apply(RESPKERNEL1, 2, sum)/Denom1
contar2 <- contar2 + 1
h.seq.correcto[contar2] <- h
Mse[contar2] <- sum((Respuesta.estimada - Respuesta)^2)
}

index.opt <- order(Mse)[1]
h.opt <- h.seq.correcto[index.opt]
KERNEL1 <- kernel(SEMIMETRIC1/h.opt)
KERNEL1[KERNEL1 < 0] <- 0
KERNEL1[KERNEL1 > 1] <- 0
diag(KERNEL1) <- 0
Denom1 <- apply(KERNEL1, 2, sum)
RESPKERNEL1 <- KERNEL1 * Respuesta
Respuesta.estimada <- apply(RESPKERNEL1, 2, sum)/Denom1
Mse.estimado <- sum((Respuesta.estimada - Respuesta)^2)/length(Respuesta)
if(sdocondato) {
  SEMIMETRIC2 <- sm(CURVAS, PRED, ...)
  KERNEL2 <- kernel(SEMIMETRIC2/h.opt)
  KERNEL2[KERNEL2 < 0] <- 0
  KERNEL2[KERNEL2 > 1] <- 0
  Denom2 <- apply(KERNEL2, 2, sum)
  Logic <- (Denom2 == 0)
  while(sum(Logic) >= 1) {
    h.opt <- 1.1 * h.opt
    KERNEL2 <- kernel(SEMIMETRIC2/h.opt)
    diag(KERNEL2) <- 0
    KERNEL2[KERNEL2 < 0] <- 0
  }
}

```

```

KERNEL2[KERNEL2 > 1] <- 0
Denom2 <- apply(KERNEL2, 2, sum)
Logic <- (Denom2 == 0)
}

RESPKERNEL2 <- KERNEL2 * Respuesta
Respuesta.predicha <- apply(RESPKERNEL2, 2, sum)/Denom2
Mse.predicho<- sum((Respuesta.predicha - Resp2)^2)/length(Resp2)
return(list(Valor.Estimado = Respuesta.estimada,
Valor.Predicho = Respuesta.predicha, hopt = h.opt,
hseq = h.seq.correcto, Mse = Mse.estimado,Mse1=Mse.predicho))
}else {
return(list(Valor.Estimado =Respuesta.estimada, hopt = h.opt,
hseq = h.seq.correcto, Mse = Mse.estimado))
}
}
}

```

### Operador Kernel con un número fijo de vecinos

```

Opnopare.kernel.knn <-function(Respuesta,CURVAS,PRED,vecinos,...,Tipo.de.kernel
="quadratic",semi.metrica="semimetric.deriv")
{
#####
# Realiza la prediccion funcional (regresion) de una respuesta escalar
# de una muestra de curvas mediante el estimador kernel funcional.
# Un ancho de banda correspondiente al numero de vecinos, que tiene que ser dado.
# "Respuesta" vector que contiene las observaciones de las respuestas escalares
# "CURVAS" matriz que contiene el conjunto de datos curvas (fila por fila)
# usados para la etapa de estimacion
# "PRED" matriz que contiene nuevas curvas almacenadas fila por fila

```

```

# usadas para el calculo de la prediccion
# "..." argumentos necesarios para la llamada de la funcion de
# calculo de la semi-metrica entre las curvas
# "Tipo.de.kernel" la funcion kernel usada para el calculo del
# estimador kernel; se puede elegir "triangle" o "quadratic (por defecto)
# "semi.metrica" calcula la distancia entre curvas;se utiliza "deriv"
# "vecinos" Numero de vecinos fijo para calcular el estimador kernel funcional
# Arroja una lista que contiene:
# "valores.estimados" vector que contiene las respuestas estimadas
# para cada curva de "CURVAS"
# "valores.predichos" si PRED es diferente de CURVAS, este vector
# contiene las respuestas predichas para cada curva de PRED
# "kNN" valor del concurrente argumento "vecinos".
# "Mse" error cuadrado medio entre valores estimados y valores observados
#####
    Respuesta <- as.vector(Respuesta)
if(is.vector(PRED)) PRED <- as.matrix(t(PRED))
testparadim <- sum(dim(CURVAS)==dim(PRED))==2
sdocondato <- T
if(testparadim) sdocondato <- sum(CURVAS==PRED)!=prod(dim(CURVAS))
sm <- get(semi.metrica)
SEMIMETRIC1 <- sm(CURVAS, CURVAS, ...)
kernel <- get(Tipo.de.kernel)
p1 <- ncol(SEMIMETRIC1)
n1 <- nrow(SEMIMETRIC1)
if(vecinos >= n1)
stop(paste("probar un numero menor de vecinos que", vecinos))

```

```

vecinos.knn1 <- 0
for(j in 1:p1) {
  Sem <- SEMIMETRIC1[, j]
  knn.para.banda <- Sem[order(Sem)[vecinos:(vecinos + 1)]]
  vecinos.knn1[j] <- 0.5 * sum(knn.para.banda)
}

KERNEL1 <- kernel(t(t(SEMIMETRIC1)/vecinos.knn1))
KERNEL1[KERNEL1 < 0] <- 0
KERNEL1[KERNEL1 > 1] <- 0
  diag(KERNEL1) <- 0
RESPKERNEL1 <- KERNEL1 * Respuesta
Denom1 <- apply(KERNEL1, 2, sum)
Respuesta.estimada <- apply(RESPKERNEL1, 2, sum)/Denom1
Mse.estimado <- sum((Respuesta.estimada - Respuesta)^2)/length(Respuesta)
if(sdocondato) {
  SEMIMETRIC2 <- sm(CURVAS, PRED, ...)
  p2 <- ncol(SEMIMETRIC2)
  vecinos.knn2 <- 0
  for(j in 1:p2) {
    Sem <- SEMIMETRIC2[, j]
    knn.para.banda <- Sem[order(Sem)[vecinos:(vecinos + 1)]]
    vecinos.knn2[j] <- 0.5 * sum(knn.para.banda)
  }

  KERNEL2 <- kernel(t(t(SEMIMETRIC2)/vecinos.knn2))
  KERNEL2[KERNEL2 < 0] <- 0
  KERNEL2[KERNEL2 > 1] <- 0
  Denom2 <- apply(KERNEL2, 2, sum)

```

```

RESPKERNEL2 <- KERNEL2 * Respuesta
Respuesta.predicha <- apply(RESPKERNEL2, 2, sum)/Denom2
  Mse.predicho<- sum((Respuesta.predicha - Resp2)^2)/length(Resp2)
return(list(Valor.Estimado = Respuesta.estimada,
Valor.Predicho = Respuesta.predicha, kNN=vecinos, Mse = Mse.estimado,
Mse1=Mse.predicho))
}
else {
return(list(Valor.Estimado = Respuesta.estimada, knn =
vecinos, Mse = Mse.estimado))
}
}

```

#### Operador Kernel con elección global del número de vecinos.

```

Opnopare.kernel.knn.gcv <-function(Respuesta,CURVAS,PRED,...,Tipo.de.kernel
= "quadratic",semi.metrica="semimetric.deriv")
{
#####
# Realiza la prediccion funcional (regresion) de una respuesta escalar
# de una muestra de curvas por medio del estimador kernel funcional.
# Un ancho de banda global (i.e. un numero de vecinos) es
# seleccionado por un procedimiento de validacion-cruzada.
# "Respuesta" vector que contiene las observaciones de la respuesta escalar
# "CURVAS" matriz que contiene el conjunto de datos de las curvas
# (fila por fila) utilizado para la etapa de estimacion
# "PRED" matriz que contiene las nuevas curvas usadas para la prediccion
# "..." argumentos necesarios para la llamada de la funcion de calculo de
# la semi-metrica entre las curvas

```

```

# "Tipo.de.kernel" la funcion kernel usada para el calculo del
# estimador kernel; se puede elegir "triangle" o "quadratic (por defecto)
# "semi.metrica" la funcion que calcula la distancia entre curvas
# Arroja una lista que contiene:
# "valor.estimado" vector que contiene las respuestas estimadas para cada
# curva de "CURVAS"
# "valor.predicho" si PRED es diferente de CURVAS, este vector
# contiene las respuestas predichas para cada curva de PRED
# "Anchos.de.banda" vector que contiene los anchos de banda Para cada curva
# de la matriz "CURVAS"
# "kvecino.opt" numero optimo de vecinos
# "Mse" error cuadrado medio entre valores estimados y valores observados
#####
  Respuesta <- as.vector(Respuesta)
  if(is.vector(PRED)) PRED <- as.matrix(t(PRED))
  testparadim <- sum(dim(CURVAS)==dim(PRED))==2
  sdocondato <- T
  if(testparadim) sdocondato <- sum(CURVAS==PRED)!=prod(dim(CURVAS))
  sm <- get(semi.metrica)
SEMIMETRIC1 <- sm(CURVAS, CURVAS, ...)
kernel <- get(Tipo.de.kernel)
n1 <- ncol(SEMIMETRIC1)
pasos <- ceiling(n1/100)
if(pasos == 0)
pasos <- 1
Kvecino <- seq(from = 10, to = n1 %/% 2, by = pasos)
kmax <- max(Kvecino)

```

```

# el vector kvecino contiene la sucesion de los k-vecinos
# cercanos usados para calcular el ancho de banda optimo
Respuesta.estimada <- 0
Bandwidth.opt <- 0
HAT.RESP <- matrix(0, nrow = n1, ncol = length(Kvecino))
BANDWIDTH <- matrix(0, nrow = n1, ncol = kmax)
for(i in 1:n1) {
  Norm.diff <- SEMIMETRIC1[, i]
  # "norm.order" da una sucesion k_1, k_2,... tal que
  # dq(X_{k_1},X_i) < dq(X_{k_2},X_i) < ...
  Norm.order <- order(Norm.diff)
  # "zz" contiene dq(X_{k_2},X_i), dq(X_{k_3},X_i),...,
  # dq(X_{j_{kmax+2}},X_i)
  zz <- sort(Norm.diff)[2:(kmax + 2)]
  # BANDWIDTH[i, l-1] contiene (dq(X_{j_l},X_i) +
  # dq(X_{j_l},X_i))/2 for l=2,...,kmax+2
  BANDWIDTH[i, ] <- 0.5 * (zz[-1] + zz[ - (kmax + 1)])
  z <- zz[ - (kmax + 1)]
  ZMAT <- matrix(rep(z, kmax), nrow = kmax, byrow = T)
  UMAT <- ZMAT/BANDWIDTH[i, ]
  KMAT <- kernel(UMAT)
  KMAT[col(KMAT) > row(KMAT)] <- 0
  Ind.curvas <- Norm.order[2:(kmax + 1)]
  Ind.resp <- Respuesta[Ind.curvas]
  YMAT <- matrix(rep(Ind.resp, kmax), nrow = kmax, byrow = T)
  HAT.RESP[i, ] <- apply(YMAT[Kvecino, ] * KMAT[Kvecino, ],
  1, sum)/apply(KMAT[Kvecino, ], 1, sum)
}

```

```

}
CRITERIUM <- (HAT.RESP - Respuesta)^2
Criterium <- apply(CRITERIUM, 2, sum)
index.opt <- order(Criterium)[1]
Respuesta.estimada <- HAT.RESP[, index.opt]
kvecino.opt <- Kvecino[index.opt]
Bandwidth.opt <- BANDWIDTH[, kvecino.opt]
Mse.estimado <- sum((Respuesta.estimada - Respuesta)^2)/n1
if(sdocondato) {
SEMIMETRIC2 <- sm(CURVAS, PRED, ...)
Bandwidth2 <- 0
n2 <- ncol(SEMIMETRIC2)
for(k in 1:n2) {
  Sm2 <- SEMIMETRIC2[, k]
  Bandwidth2[k] <- sum(sort(Sm2)[kvecino.opt:(kvecino.opt+1)])*0.5
}
KERNEL <- kernel(t(t(SEMIMETRIC2)/Bandwidth2))
KERNEL[KERNEL < 0] <- 0
KERNEL[KERNEL > 1] <- 0
Denom <- apply(KERNEL, 2, sum)
RESPKERNEL <- KERNEL * Respuesta
Respuesta.predicha <- apply(RESPKERNEL, 2, sum)/Denom
Mse.predicho<- sum((Respuesta.predicha - Resp2)^2)/length(Resp2)
return(list(Valor.Estimado = Respuesta.estimada,
Valor.Predicho = Respuesta.predicha, Anchos.de.banda =
Bandwidth.opt, kvecino.opt = kvecino.opt, Mse =
Mse.estimado, Mse1=Mse.predicho))

```

```

}else {
return(list(Valor.Estimado = Respuesta.estimada, Anchos.de.banda
= Bandwidth.opt, kvecino.opt = kvecino.opt, Mse =
Mse.estimado))
}
}

```

### Operador Kernel con elección local del número de vecinos.

```

Opnopare.kernel.knn.lcv <-function(Respuesta, CURVAS, PRED, ...,
Tipo.de.kernel="quadratic",semi.metrica ="semimetric.deriv")
{
#####
# Realiza la prediccion funcional (regresion) de respuesta escalar
# de una muestra de curvas mediante el estimador kernel funcional.
# Un ancho de banda local (i.e. un numero local de vecinos) es seleccionado
# por un procedimiento de validacion cruzada.
# "Respuesta" vector que contiene las observaciones de la respuesta escalar
# "CURVAS" matriz que contiene el conjunto de datos de las curvas
# (fila por fila) utilizado para la etapa de estimacion
# "PRED" matriz que contiene las nuevas curvas usadas para la prediccion
# "..." argumentos necesarios para la llamada de la funcion de calculo de
# la semi-metrica entre las curvas
# "Tipo.de.kernel" la funcion kernel usada para el calculo del estimador
# kernel; se puede elegir "triangle" o "quadratic (por defecto)
# "semi.metrica" la funcion de calcula la distancia entre curvas
# Arroja una lista que contiene:
# "valor.estimado" vector que contiene las respuestas estimadas para cada
# curva de "CURVAS"

```

```

# "valor.predicho" si PRED es diferente de CURVAS, este vector
# contiene las respuestas predichas para cada curva de PRED
# "Anchos.de.banda" vector que contiene los datos globales impulsados por
# anchos de banda Para cada curva de la matriz "CURVAS"
# "Mse" error cuadrado medio entre valores estimados y valores observados
#####
    Respuesta <- as.vector(Respuesta)
if(is.vector(PRED)) PRED <- as.matrix(t(PRED))
testparadim <- sum(dim(CURVAS)==dim(PRED))==2
    sdocondato <- T
if(testparadim) sdocondato <- sum(CURVAS==PRED)!=prod(dim(CURVAS))
sm <- get(semi.mÃ©trica)
SEMIMETRIC1 <- sm(CURVAS, CURVAS, ...)
kernel <- get(Tipo.de.kernel)
n1 <- ncol(SEMIMETRIC1)
pasos <- ceiling(n1/100)
if(pasos == 0)
pasos <- 1
Kvecino <- seq(from = 10, to = n1 %/% 2, by = pasos)
kmax <- max(Kvecino)
    # el vector kvecino contiene la sucesion de los k-vecinos
    # cercanos usados para calcular el ancho de banda optimo
Respuesta.estimada <- 0
    Bandwidth.opt <- 0
    Knn1 <- 0
for(i in 1:n1) {
Norm.diff <- SEMIMETRIC1[, i]

```

```

# "norm.order" da una sucesion k_1, k_2,... tal que
# dq(X_{k_1},X_i) < dq(X_{k_2},X_i) < ...
Norm.order <- order(Norm.diff)
# "zz" contiene dq(X_{k_2},X_i), dq(X_{k_3},X_i),...,
# dq(X_{j_{kmax+2}},X_i)
zz <- sort(Norm.diff)[2:(kmax + 2)]
# Bandwidth[l-1] contiene (dq(X_{j_l},X_i) +
# dq(X_{j_l},X_i))/2 for l=2,...,kmax+2
Bandwidth <- 0.5 * (zz[-1] + zz[ - (kmax + 1)])
z <- zz[ - (kmax + 1)]
ZMAT <- matrix(rep(z, kmax), nrow = kmax, byrow = T)
UMAT <- ZMAT/Bandwidth
KMAT <- kernel(UMAT)
KMAT[col(KMAT) > row(KMAT)] <- 0
Ind.curvas <- Norm.order[2:(kmax + 1)]
Ind.resp <- Respuesta[Ind.curvas]
YMAT <- matrix(rep(Ind.resp, kmax), nrow = kmax, byrow = T)
Hat.resp <- apply(YMAT[Kvecino, ] * KMAT[Kvecino, ], 1, sum
)/apply(KMAT[Kvecino, ], 1, sum)
Criterium <- abs(Hat.resp - Respuesta[i])
index <- order(Criterium)[1]
Knn1[i] <- Kvecino[index]
Respuesta.estimada[i] <- Hat.resp[index]
Bandwidth.opt[i] <- Bandwidth[index]
}
Mse.estimado <- sum((Respuesta.estimada - Respuesta)^2)/n1
if(sdocondato) {

```

```

SEMIMETRIC2 <- sm(CURVAS, PRED, ...)
Bandwidth2 <- 0
n2 <- ncol(SEMIMETRIC2)
  knn2 <- 0
for(k in 1:n2) {
  Sm2 <- SEMIMETRIC2[, k]
  Sm2.ord <- order(SEMIMETRIC2[, k])
  knn2[k]<-Knn1[Sm2.ord[1]]
  knn <- Knn1[Sm2.ord[1]]
  Bandwidth2[k] <- sum(sort(Sm2)[knn:(knn+1)])*0.5
}
KERNEL <- kernel(t(t(SEMIMETRIC2)/Bandwidth2))
KERNEL[KERNEL < 0] <- 0
KERNEL[KERNEL > 1] <- 0
Denom <- apply(as.matrix(KERNEL), 2, sum)
RESPKERNEL <- KERNEL * Respuesta
Respuesta.predicha <- apply(as.matrix(RESPKERNEL), 2, sum)/Denom
  Mse.predicho<- sum((Respuesta.predicha - Resp2)^2)/length(Resp2)
return(list(Valor.Estimado = Respuesta.estimada,
Valor.Predicho = Respuesta.predicha, Anchos.de.banda =
Bandwidth.opt, kvecinos.opt = knn2, Mse =
Mse.estimado, Mse1=Mse.predicho))
}
else {
return(list(Valor.Estimado = Respuesta.estimada, Anchos.de.banda
= Bandwidth.opt, Mse=Mse.estimado,kvecinos.opt=knn1))
}

```

}

### Operador Mediana condicional

```

Opnopa.f.d.a.cuantil.lcv <- function(Respuesta,CURVAS,PRED,...,alpha=
c(0.05, 0.5, 0.95),Tipo.de.kernel="quadratic",semi.metrica="semimetric.deriv")
{
#####
# realiza la prediccion funcional de una respuesta escalar de una
# de una muestra de curvas por el calculo de la mediana condicional funcional.
# Un ancho de banda local (i.e. un numero local de vecinos) es seleccionado
# por un "trivial" procedimiento de validacion cruzada.
# "Respuesta" vector que contiene las observaciones de la respuesta escalar
# "CURVAS" matriz que contiene el conjunto de datos de las curvas
# (fila por fila) utilizado para la etapa de estimacion
# "PRED" matriz que contiene las nuevas curvas usadas para la prediccion
# "..." argumentos necesarios para la llamada de la funcion de calculo
# de la semi-metrica entre las curvas
# "alpha" vector que da los cuantiles a ser calculados.
# "Tipo.de.kernel" la funcion kernel usada para el calculo del estimador
# kernel; se puede elegir "triangle" o "quadratic" (por defecto)
# "semi.metrica" la funcion de calculo la distancia entre curvas
# Arroja una lista que contiene:
# "valor.estimado" es una matriz cuyas columnas da los correspondientes
# cuantiles condicionales estimados para todas las curvas de aprendizaje
# "valor.predicho" si PRED es diferente de CURVAS, esta matriz
# contiene los cuantiles condicionales predichos para cada curva de PRED
# "Respuestas.originales" vector que contiene las respuestas observadas.
# "Mse" error cuadrado medio entre valores estimados y valores observados

```

```
#####
  Respuesta <- as.vector(Respuesta)
lalpha <- length(alpha)
  logicalalpha <- alpha==0.5
testalpha <- lalpha > 1
if(testalpha) posmediana <- (1:lalpha)[logicalalpha]
if(is.vector(PRED)) PRED <- as.matrix(t(PRED))
unafila <- nrow(PRED) == 1
testparadim <- sum(dim(CURVAS)==dim(PRED))==2
sdocondato <- T
if(testparadim) sdocondato <- sum(CURVAS==PRED)!=prod(dim(CURVAS))
sm <- get(semi.metrica)
kernel <- get(Tipo.de.kernel)
kernel.int <- get("integrado.quadratic")
laprendizaje <- nrow(CURVAS)
Aprendizaje1 <- seq(2, laprendizaje, by=2)
lAprendizaje1 <- length(Aprendizaje1)
APRENDIZAJE1 <- CURVAS[Aprendizaje1, ]
APRENDIZAJE2 <- CURVAS[ - Aprendizaje1, ]
SMAPRENDIZAJE1 <- sm(APRENDIZAJE1, APRENDIZAJE1, ...)
SMAPRENDIZAJE12 <- sm(APRENDIZAJE1, APRENDIZAJE2, ...)
SML12.ORDEN <- apply(SMAPRENDIZAJE12, 2, sort)
Respu1 <- Respuesta[Aprendizaje1]
Respu2 <- Respuesta[ - Aprendizaje1]
Resp.rango <- range(Respuesta)
Red.Respuestas <- seq(from = Resp.rango[1] * 0.9, to = Resp.rango[2] *
1.1, length = 100)
```

```

# DIFRESPUESTA[i,j]=yi-yj con yi en Red.Respuestas y yj en Respu1
DIFRESPUESTA <- outer(Red.Respuestas, Respu1, "-")
DIFRESPU.ORDEN <- t(apply(abs(DIFRESPUESTA), 1, sort))
laprendizaje2 <- nrow(APRENDIZAJE2)
lred <- length(Red.Respuestas)
Kvecino.min <- max(ceiling(lAprendizaje1 * 0.05), 10)
Kvecino.max <- ceiling(lAprendizaje1 * 0.25)
  if(Kvecino.max <= Kvecino.min){
    Kvecino.min <- ceiling(lAprendizaje1 * 0.05)
  }
pasos <- ceiling((Kvecino.max - Kvecino.min)/10)
Kvecino <- seq(Kvecino.min, Kvecino.max, by = pasos)
lkvecino <- length(Kvecino)
BANDL12.CUR <-0.5*(SML12.ORDEN[Kvecino,]+SML12.ORDEN[Kvecino + 1,])
BAND.RESP <-0.5*(DIFRESPU.ORDEN[,Kvecino]+DIFRESPU.ORDEN[,Kvecino+1])
CV <- matrix(0, nrow = lkvecino^2, ncol = laprendizaje2)
if(testalpha){
  CUANT <- array(0, dim = c(lkvecino^2, laprendizaje2, lalpha))
} else {
  CUANT <- matrix(0, nrow = lkvecino^2, ncol = laprendizaje2)
}
contar <- 0
for(kk in 1:lkvecino) {
  ARG <- t(t(SMAPRENDIZAJE12)/BANDL12.CUR[kk,])
  KERNEL.CURVAS <- kernel(ARG)
  KERNEL.CURVAS[KERNEL.CURVAS < 0] <- 0
  KERNEL.CURVAS[KERNEL.CURVAS > 1] <- 0
}

```

```

Denom <- apply(KERNEL.CURVAS, 2, sum)
for(hh in 1:lvecino) {
  contar <- contar + 1
  KERNEL.I.RESP <- apply(DIFRESPUESTA/BAND.RESP[, hh], 1, kernel.int)
  FDA.EST <- (t(KERNEL.CURVAS)/Denom) %*% KERNEL.I.RESP
  if(testalpha){
    for(ii in 1:lalpha){
      Ind.cuant <- apply(FDA.EST<alpha[ii], 1, sum)
      Ind.cuant[Ind.cuant==0] <- 1
      CUANT[contar,,ii] <- Red.Respuestas[Ind.cuant]
    }
    CV[contar,] <- (Respu2 - CUANT[contar,,posmediana])^2
  } else {
    Ind.cuant <- apply(FDA.EST < alpha, 1, sum)
    Ind.cuant[Ind.cuant==0] <- 1
    CUANT[contar, ] <- Red.Respuestas[Ind.cuant]
    CV[contar, ] <- (Respu2 - CUANT[contar, ])^2
  }}
  Ind.kvecino.opt <- apply(CV, 2, order)[1, ]
  IND.OPT <- cbind(Ind.kvecino.opt, 1:laprendizaje2)
  if(testalpha){
    Indkopt <- rep(Ind.kvecino.opt, lalpha)
    Units <- rep(1:laprendizaje2, lalpha)
    Typeofest <- sort(rep(1:lalpha,laprendizaje2))
    RESPUESTA.ESTIMADA <- matrix(CUANT[cbind(Indkopt,Units,Typeofest)],
  nrow=laprendizaje2, byrow=F)
    dimnames(RESPUESTA.ESTIMADA) <- list(NULL, as.character(alpha))
  }
}

```

```

} else { RESPUESTA.ESTIMADA<- CUANT[IND.OPT]
}

Mse.estimado <- sum(CV[IND.OPT])/laprendizaje2
if(sdocondato) {
SMAPRENDIZAJE2N <- sm(APRENDIZAJE2, PRED, ...)
Nuevo.Orden <- apply(SMAPRENDIZAJE2N, 2, order)[1, ]
if(testalpha){
if(unafila){
RESPUESTA.PREDICHA <- as.matrix(t(RESPUESTA.ESTIMADA[Nuevo.Orden,]))
} else {
RESPUESTA.PREDICHA <- as.matrix(RESPUESTA.ESTIMADA[Nuevo.Orden,])
dimnames(RESPUESTA.PREDICHA) <- list(NULL, as.character(alpha))}
} else {
RESPUESTA.PREDICHA <- RESPUESTA.ESTIMADA[Nuevo.Orden]
}
return(list(valor.estimado = RESPUESTA.ESTIMADA,
valor.predicho = RESPUESTA.PREDICHA, Respuestas.originales
= Resp2, Mse = Mse.estimado))
}
else {
return(list(valor.estimado = RESPUESTA.ESTIMADA,
Respuestas.originales = Resp2, Mse = Mse.estimado))
}
}
}

```

### Operador Moda condicional

```

Opnopa.f.d.moda.lcv <-function(Respuesta,CURVAS,PRED,...,Tipo.de.kernel=
"quadratic", semi.metrica = "semimetric.deriv")
{
#####
# realiza la prediccion funcional de una respuesta escalar de una
# de una muestra de curvas por el calculo de la moda condicional funcional.
# Un ancho de banda local (i.e. un numero local de vecinos) es seleccionado
# por un "trivial" procedimiento de validacion cruzada.
# "Respuesta" vector que contiene las observaciones de la respuesta escalar
# "CURVAS" matriz que contiene el conjunto de datos de las curvas
# (fila por fila) utilizado para la etapa de estimacion
# "PRED" matriz que contiene las nuevas curvas usadas para la prediccion
# "..." argumentos necesarios para la llamada de la funcion de calculo
# de la semi-metrica entre las curvas
# "Tipo.de.kernel" la funcion kernel usada para el calculo del estimador
# kernel; se puede elegir "triangle" o "quadratic (por defecto)
# "semi.metrica" la funcion de calcula la distancia entre curvas
# Arroja una lista que contiene:
# "valor.estimado" vector que contiene las respuestas estimadas
# para cada curva de "CURVAS"
# "valor.predicho" si PRED es diferente de CURVAS, este vector
# contiene las respuestas predichas para cada curva de PRED
# "Respuestas.originales" vector que contiene las respuestas observadas.
# "Mse" error cuadrado medio entre valores estimados y valores observados
#####
  Respuesta <- as.vector(Respuesta)

```

```

if(is.vector(PRED)) PRED <- as.matrix(t(PRED))
testparadim <- sum(dim(CURVAS)==dim(PRED))==2
sdocondato <- T
if(testparadim) sdocondato <- sum(CURVAS==PRED)!=prod(dim(CURVAS))
sm <- get(semi.metrica)
kernel <- get(Tipo.de.kernel)
laprendizaje <- nrow(CURVAS)
Aprendizaje1 <- seq(2, laprendizaje, by=2)
lAprendizaje1 <- length(Aprendizaje1)
APRENDIZAJE1 <- CURVAS[Aprendizaje1, ]
APRENDIZAJE2 <- CURVAS[ -Aprendizaje1, ]
SMAPRENDIZAJE1 <- sm(APRENDIZAJE1, APRENDIZAJE1, ...)
SMAPRENDIZAJE12 <- sm(APRENDIZAJE1, APRENDIZAJE2, ...)
SML12.ORDEN <- apply(SMAPRENDIZAJE12, 2, sort)
Respu1 <- Respuesta[Aprendizaje1]
Respu2 <- Respuesta[ -Aprendizaje1]
Rango.Resp <- range(Respuesta)
Red.Respuestas <- seq(from = Rango.Resp[1] * 0.9, to = Rango.Resp[2] *
1.1, length = 100)
# DIFRESPUESTA[i,j]=yi-yj con i en Red.Respuestas and j en Respu1
DIFRESPUESTA <- outer(Red.Respuestas, Respu1, "-")
DIFRES.ORDEN <- t(apply(abs(DIFRESPUESTA), 1, sort))
laprendizaje2 <- nrow(APRENDIZAJE2)
lRed <- length(Red.Respuestas)
Kvecino.min <- max(ceiling(lAprendizaje1 * 0.05), 10)
Kvecino.max <- ceiling(lAprendizaje1 * 0.25)
  if(Kvecino.max <= Kvecino.min){

```

```

    Kvecino.min <- ceiling(lAprendizaje1 * 0.05)
  }
pasos <- ceiling((Kvecino.max - Kvecino.min)/10)
Kvecino <- seq(Kvecino.min, Kvecino.max, by = pasos)
lKvecino <- length(Kvecino)
BANDL12.CUR <- 0.5 * (SML12.ORDEN[Kvecino,]+SML12.ORDEN[Kvecino+1,])
BAND.RESP <- 0.5*(DIFRES.ORDEN[,Kvecino]+DIFRES.ORDEN[,Kvecino+1])
CV <- matrix(0, nrow = lKvecino^2, ncol = laprendizaje2)
MODA <- matrix(0, nrow = lKvecino^2, ncol =laprendizaje2)
contar1 <- 0
contar2 <- 0
for(kk in Kvecino) {
  contar2 <- contar2 + 1
  ARG <- t(t(SMAPRENDIZAJE12)/BANDL12.CUR[contar2, ])
  KERNEL.CURVAS <- kernel(ARG)
  KERNEL.CURVAS[KERNEL.CURVAS < 0] <- 0
  KERNEL.CURVAS[KERNEL.CURVAS > 1] <- 0
  Denom <- apply(KERNEL.CURVAS, 2, sum)
  contar3 <- 0
  for(hh in Kvecino) {
    contar1 <- contar1 + 1
    contar3 <- contar3 + 1
    KERNEL.RESP <- apply(abs(DIFRESPUESTA)/BAND.RESP[, contar3], 1, kernel)
    KERNEL.RESP[KERNEL.RESP < 0] <- 0
    KERNEL.RESP[KERNEL.RESP > 1] <- 0
    DENS.ESTIMADA<-(t(KERNEL.CURVAS)/Denom)%*(KERNEL.RESP/BAND.RESP[,contar3])
    Ind.moda <-apply(DENSIDAD.ESTIMADA,1,order)[lRed,]
  }
}

```

```

MODA[contar1, ] <- Red.Respuestas[Ind.moda]
CV[contar1, ] <- (Respu2 - MODA[contar1, ])^2
}}

Ind.Kvecino.opt <- apply(CV, 2, order)[1, ]
IND.OPT <- cbind(Ind.Kvecino.opt, 1:laprendizaje2)
Respuesta.estimada <- MODA[IND.OPT]
Mse.estimado <- sum(CV[IND.OPT])/laprendizaje2
if(sdocondato) {
  SMAPRENDIZAJE2N <- sm(APRENDIZAJE2, PRED, ...)
  Nuevo.Orden <- apply(SMAPRENDIZAJE2N, 2, order)[1, ]
  Respuesta.predicha <- Respuesta.estimada[Nuevo.Orden]
  return(list(valor.estimado = Respuesta.estimada,
  valor.predicho = Respuesta.predicha, Respuesta.originales
  = Respu2, Mse = Mse.estimado))
}
else {
  return(list(valor.estimado = Respuesta.estimada,
  Respuesta.originales = Respu2, Mse = Mse.estimado))
}
}

```

### Calculando los errores cuadrados

```

Se.reg <- (Resp2 - prediccion.regresion$Valor.Predicho)^2
Se.moda <- (Resp2 - prediccion.moda$valor.predicho)^2
Se.mediana <- (Resp2 - prediccion.mediana$valor.predicho[,2])^2
mse.reg <- round(sum(Se.reg)/length(Resp2),2)
mse.moda <- round(sum(Se.moda)/length(Resp2),2)
mse.mediana <- round(sum(Se.mediana)/length(Resp2),2)

```

### Rutina para generar la figura4-2

```
#####
# PLOTEANDO RESPUESTAS PREDICHAS #####
#####
par(mfrow=c(2,2), pty="s")
limiteX=range(Resp2)
limiteY=range(c(prediccion.regresion$Valor.Predicho,
prediccion.moda$valor.predicho,prediccion.mediana$valor.predicho[,2]))
plot(Resp2,prediccion.regresion$Valor.Predicho,xlim=limiteX,ylim=limiteY,
xlab="Respuestas de muestra de prueba",ylab="Respuestas predichas",pch=20)
mtext(paste("Exper.Cond.: MSE=",mse.reg,""), cex=1.15, line=1)
abline(0,1)
plot(Resp2,prediccion.mediana$valor.predicho[,2],xlim=limiteX,ylim=limiteY,
xlab="Respuestas de muestra de prueba",ylab="Respuestas predichas",pch=20)
mtext(paste("Mediana.Cond.: MSE=",mse.mediana,""), cex=1.15, line=1)
abline(0,1)
plot(Resp2,prediccion.moda$valor.predicho,xlim=limiteX,ylim=limiteY,
xlab="Respuestas de muestra de prueba",ylab="Respuestas predichas",pch=20)
mtext(paste("Moda.Cond.: MSE=",mse.moda,""), cex=1.15, line=1)
abline(0,1)
```

### Rutina para generar la figura4-3

```
#####
# La mejora de las predicciones mediante el promedio de los tres#
# valores predichos (tres metodos) #
#####
```

```

result.multimetodo <- (prediccion.regresion$Valor.Predicho+
prediccion.moda$valor.predicho + prediccion.mediana$valor.predicho[,2])/3
Se.multimetodo <-(Resp2 - result.multimetodo)^2
mse.multimetodo <- round(sum(Se.multimetodo)/length(Resp2),2)
nombre.error <- c(paste("Moda.Cond\n MSE=",mse.moda,sep=""),
paste("Exp.Cond \n MSE=",mse.reg,sep=""),
paste("Med.Cond \n MSE=",mse.mediana,sep=""),
paste("Multimetodo \n MSE=",mse.multimetodo,sep=""))
par(mfrow=c(1,2))
plot(Resp2,result.multimetodo,xlim=limiteX,xlab="Respuestas
de muestra de prueba",ylab="Respuestas predichas",pch=20)
mtext(paste("Multimetodo.: MSE=",mse.multimetodo,""), cex=1.15, line=1)
abline(0,1)
boxplot(Se.reg, Se.moda, Se.mediana, Se.multimetodo, names = nombre.error)
title(" Error Cuadrado")

```

### Rutina para generar las funciones Kernel

```

#####
# KERNEL CUADRATICO #
#####
quadratic <- function(u)
{
  3/4*(1 - (u)^2)
}
#####
### KERNEL CUADRATICO ASIMETRICO ###
#####

```

```

quadratic2 <- function(u)
{
u[u<0] <- 1
u[u>1] <- 1
return(1 - (u)^2)
}

#####
# Kernel triangular A #
#####
triangle <- function(u)
{
u[u<0] <- 1
u[u>1] <- 1
return( 1 - abs(u))
}

#####
# KERNEL CUADRATICO INTEGRADO#
#####
integrado.quadratic <- function(u)
{
resultado <- u
Logic0 <- (u <= -1)
Logic1 <- (u >= 1)
Logic2 <- (u > -1 & u < 1)
resultado[Logic0] <- 0
resultado[Logic1] <- 1
Dval <- resultado[Logic2]
}

```

```

resultado[Logic2] <- 0.75 * Dval * (1 - (Dval^2)/3) + 0.5
return(resultado)
}

#####
# KERNEL TRIANGULAR INTEGRADO #
#####
integrado.triangle <- function(u)
{
  resultado <- u
  Logic0 <- (u <= -1 | u >= 1)
  Logic1 <- (u > -1 & u < 0)
  Logic2 <- (u >= 0 & u < 1)
  Dneg <- u[Logic1]
  Dpos <- u[Logic2]
  resultado[Logic1] <- Dneg + (Dneg^2)/2 + 0.5
  resultado[Logic2] <- Dpos - (Dpos^2)/2 + 0.5
  resultado[Logic0] <- 0
  return(resultado)
}

```

### Rutina para generar la semi-métrica

```

symsolve <- function(Asym, Bmat)
{
  #Interpretado por J. O. Ramsay y disponible en su sitio web
  #http://ego.psych.mcgill.ca/misc/fda que contiene una gran cantidad
  #de funciones para el analisis de datos funcionales de una manera
  #diferente: Resuelve el sistema ASYM X = BMAT

```

```

#para X donde ASYM es simetrica.regresa X
n <- ncol(Asym)
if(max(abs(Asym - t(Asym)))/max(abs(Asym)) > 1e-10)
stop("Argument not symmetric.")
Lmat <- chol(Asym, T)
if(attr(Lmat, "rank") < n)
stop("Argument singular.")
Lmatinv <- solve(Lmat[, order(attr(Lmat, "pivot"))])
Xmat <- Lmatinv %*% t(Lmatinv) %*% Bmat
return(Xmat)
}

#####
# SEMI-METRICA DERIVADA #
#####

semimetric.deriv <- function(DATA1, DATA2, q=2, nknot=20,
range.grid=c(0,1))
{
#####
#Calcula la distancia entre curvas basada en las derivadas.
#"DATA1" matriz que contiene el primer conjunto de curvas
#"DATA2" matriz que contiene el segundo conjunto de curvas
#"q" orden de derivacion
#"nknot" numero de nodos interiores (necesario para la
#definicion de la base B-spline)
#"range.grid" vector de longitud 2 que contiene el rango de
#la red para el cual las curvas son evaluadas
#Retorna

```

```

#"semimetric" una matriz que contiene la medida entre las curvas.
#####
library(splines)
if(is.vector(DATA1)) DATA1 <- as.matrix(t(DATA1))
if(is.vector(DATA2)) DATA2 <- as.matrix(t(DATA2))
testfordim <- sum(dim(DATA1)==dim(DATA2))==2
twodatasets <- T
if(testfordim) twodatasets <- sum(DATA1==DATA2)!=prod(dim(DATA1))
#####
# aproximacion B-splines de las curvas contenidas en el conjunto da datos:
#-----
# "knot" y "x" permiten definir la base B-splines
# "coef.mat1[, i]" corresponde a la expansion B-splines de las
# curvas discretizadas contenidas en DATASET[i, ].
# la expansion B-splines de las curvas contenidas en "DATA1[i, ]"
# es dado por "Bspline %*% coef.mat1[, i]"
#####
p <- ncol(DATA1)
a <- range.grid[1]
b <- range.grid[2]
x <- seq(a, b, length = p)
order.Bspline <- q + 3
nknotmax <- (p - order.Bspline - 1)%/%2
if(nknot > nknotmax){
stop(paste("dar un nunero de nodos mas pequeno que
",nknotmax, " para evitar una matriz mal condicionada"))
}

```

```

Knot <- seq(a, b, length = nknot + 2)[- c(1, nknot + 2)]
delta <- sort(c(rep(c(a, b), order.Bspline), Knot))
Bspline <- splineDesign(delta, x, order.Bspline)
Cmat <- crossprod(Bspline)
Dmat1 <- crossprod(Bspline, t(DATA1))
coef.mat1 <- symsolve(Cmat, Dmat1)
#####
# Integracion numerica para el method Gauss:
# -----
# Los objetos que terminan por "gauss" nos permiten calcular numericamente
# integrales por medio del "metodo de Gauss" (lx.gauss=6==>el calculo
# de la integral es exacta para polinomio de grado menor o igual a 11).
#####
point.gauss <- c(-0.9324695142, -0.6612093865, -0.2386191861,
0.2386191861, 0.6612093865, 0.9324695142)
weight.gauss <- c(0.1713244924, 0.360761573, 0.4679139346,
0.4679139346, 0.360761573, 0.1713244924)
x.gauss <- 0.5 * ((b + a) + (b - a) * point.gauss)
lx.gauss <- length(x.gauss)
Bspline.deriv <- splineDesign(delta, x.gauss, order.Bspline,
rep(q, lx.gauss))
H<-t(Bspline.deriv)%*%(Bspline.deriv*(weight.gauss*0.5*(b-a)))
eigH <- eigen(H, sym = T)
eigH$values[eigH$values < 0] <- 0
Hhalf <- t(eigH$vectors %*%(t(eigH$vectors)*sqrt(eigH$values)))
COEF1 <- t(Hhalf %*% coef.mat1)
if(twodatasets){

```

```
Dmat2 <- crossprod(Bspline, t(DATA2))
coef.mat2 <- symsolve(Cmat, Dmat2)
COEF2 <- t(Hhalf %*% coef.mat2)
} else {
COEF2 <- COEF1
}
SEMIMETRIC <- 0
nbasis <- nrow(H)
for(f in 1:nbasis)
SEMIMETRIC <- SEMIMETRIC +outer(COEF1[, f],COEF2[,f],"-")^2
return(sqrt(SEMIMETRIC))
}
```

## Bibliografía

- [1] Borggaard C. and Thodberg H. *H. Optimal Minimal Neural Interpretation of Spectra*. Analytical Chemistry, 1992.
- [2] de Boor C. *A practical guide to splines*. Springer, New York., 1978.
- [3] Preda C. and Saporta G. *Clusterwise PLS regression on a stochastic process*. Comput.Statist. Data Anal., 2005.
- [4] Preda C. and Saporta G. *PLS regression on a stochastic process*. Comput.Statist. Data Anal., 2005.
- [5] Bosq D. and Lecoutre J. *Théorie de l'estimation fonctionnelle (in french)*. Economica, 1987.
- [6] Nadaraya E. *On estimating regression*. Theory Proba. Appl., 1964.
- [7] Ferraty F. and Vieu P. *Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics)*. Springer-Verlag New York, Inc., 2006.
- [8] Collomb G. *Estimation nonparamétrique de la régression (in french)*. PhD Université Paul Sabatier, Toulouse. 1976.
- [9] Roussas G. *Nonparametric estimation of the transition distribution function of a Markov process*. Annals of Mathematical Statistics, 1969.
- [10] Watson G. *Smooth regression analysis*. Sankhya Ser.A, 1964.
- [11] Martens H. and Naes T. *Multivariate calibration*. 1989.
- [12] Wold H. *Soft modelling by latent variables; the non linear iterative partialleast squares approach*. in: J. gani (ed.). *Perspectives in Probability and Statistics, Papers in Honour of M.S. Bartlett.*, Academic Press, London(1975).
- [13] Frank I.E. and Friedman J.H. *A statistical view of some chemometrics regression tools (with discussions).*, volume (2). 1993.

- [14] Dauxois J., Pousse A., and Romain Y. *Asymptotic theory for the principal component analysis of a random vector function: some application to statistical inference.* J. Multivariate Anal., 1982.
- [15] Samanta M. *Nonparametric estimation of conditional quantiles.* Statist. Proba. Letters,, 1989.
- [16] Schumaker M. *Spline functions: basic theory.* Wiley., 1981.
- [17] Dabo-Niang S. and Rhomari N. *Estimation non paramétrique de la régression avec variable explicative dans un espace métrique (in french).* Compte Rendus Acad. Sci. Paris,, 2003.
- [18] Leurgans S.E. and Moyeed R.A. *Silverman, B.W. Canonical correlation analysis when the data are curves.* J. R. Statist. Soc. B, 1993.
- [19] Hardle W. *Applied nonparametric regression.* Cambridge Univ. Press, UK., 1990.
- [20] Massy W.F. *Principal Components Regression in Exploratory Statistical Research.* 1965.
- [21] Paola Bermolen y Federico Larroca. *Cálculo de la banda de ancho óptima para el estimador de Nadaraya-Watson.* Monografía presentada para la aprobación del curso "Automatic Learning Machines" dictado por el Prof. Gonzalo Perera en el IMERL., 2005.

# ANÁLISIS NO PARAMÉTRICO PARA LA PREDICCIÓN DE DATOS FUNCIONALES

Daiver De Jesús Vélez Ramos

[daiver.velez@upr.edu](mailto:daiver.velez@upr.edu)

Departamento de Ciencias Matemáticas

Consejero: Edgar Acuña Fernández Ph.D

Grado: Maestría en Ciencias

Fecha de Graduación: 2013