

FACULTAD DE CIENCIAS
TRABAJO FIN DE GRADO EN MATEMÁTICAS

Algunos métodos estadísticos aplicados a la astrofísica

Borja Pérez Díaz

2018-2019



Universidad de Oviedo

Universidá d'Uviéu

University of Oviedo

Dirigido por: María Teresa López García

Índice general

Motivación	3
1. Introducción	4
1.1. Conceptos preliminares	4
1.1.1. Función de distribución y de densidad	4
1.1.2. Muestras aleatorias simples	4
1.1.3. Función de distribución empírica	5
1.1.4. Propiedades de interés de los estimadores	5
1.2. Ejemplo de cuásares SDSS	6
2. Estimación de la función de densidad	9
2.1. Introducción	9
2.2. Histogramas	9
2.3. Estimador naïve	13
2.4. Estimadores kernel	17
2.4.1. Introducción y propiedades generales	17
2.4.2. Aproximación teórica a la ventana óptima	19
2.4.3. Aproximación teórica a la función kernel óptima	21
2.4.4. Funciones kernel más usuales	22
2.4.5. Generalización a funciones multidimensionales	26
3. Selección de ventana óptima	27
3.1. Introducción	27
3.2. Aproximación práctica a la ventana óptima	27
3.2.1. Elección subjetiva	27
3.2.2. Elección en el caso normal	28
3.2.3. Elección por métodos <i>plug-in</i>	31
3.2.4. Elección por sobre suavización	38
3.3. Elección por validación cruzada	41
3.3.1. Validación cruzada por máxima verosimilitud	42
3.3.2. Validación cruzada por mínimos cuadrados	43
3.3.3. Validación cruzada por sesgo	45
3.3.4. Validación cruzada suave	48

3.3.5. Validación cruzada indirecta	52
3.3.6. Evaluación de los métodos de validación cruzada	56
3.4. Procedimientos bootstrap	58
3.4.1. Introducción y procedimiento básico	58
3.4.2. Bootstrap suave sin anchura pivote	59
3.4.3. Bootstrap suave con anchura pivote	60
3.4.4. Bootstrap no-suave	60
3.4.5. Evaluación de los procedimientos bootstrap	61
4. Otras mejoras en la estimación	68
4.1. Suavidad adaptada	68
4.1.1. Método del k-vecino más cercano	68
4.1.2. Método kernel adaptado	74
4.2. Evaluación de los métodos de suavidad adaptada	78
4.2.1. Aplicación del método del k-vecino más cercano	78
4.2.2. Aplicación del método kernel adaptado	80
Anexo de programas en R	83
A.1. Código del Ejemplo 3.6	83
A.2. Código del estudio de la Subsección 3.3.6	84
A.3. Código del estudio de la Subsección 3.4.4	89
A.4. Código del estudio de la Subsección 4.2.1	93
A.5. Código del estudio de la Subsección 4.2.2	96
Bibliografía	98

Motivación

El objetivo del presente trabajo fin de grado es el conocimiento, análisis y aplicación de la estimación de la función de densidad: técnica estadística destinada a resumir la información de un conjunto de datos mediante una función suave que es la función de densidad de una distribución continua. Esta técnica es de amplio uso en varios campos científicos, entre ellos, la astrofísica.

En estadística, la función de densidad juega un papel esencial para caracterizar una variable continua, además de ser la base del desarrollo teórico de muchos procedimientos inferenciales como los basados en la verosimilitud, de ahí la importancia de su estimación. En el contexto astrofísico, juega un papel importante en muchos aspectos: analizar las distribuciones de materia en las galaxias para determinar la distribución de materia oscura; estimar las curvas de luminosidad de diferentes objetos para determinar si tiene un comportamiento periódico o episódico; o, simplemente, analizar la posible relación entre diferentes magnitudes físicas.

En este trabajo, no sólo nos centraremos en el análisis teórico de diferentes técnicas empleadas en la estimación de la función de densidad, sino que también aplicaremos estos métodos para un caso particular: la distribución de cuásares a través del redshift. Esta conjunción de teoría y práctica queda plasmada en la elaboración de los programas en código R, para la ejecución de los ejemplos, recogidos en el Anexo A.

El conjunto de librerías R, de amplio uso en el campo de la estadística y la probabilidad, cuenta ya con una gran cantidad de funciones implementadas que se pueden usar en la estimación de la función de densidad. Por ejemplo, la función *hist* permite la visualización directa de histogramas mientras que la función *density* permite la construcción del estimador kernel (que ocupará la mayor parte del presente trabajo) sin más que introducir la muestra aleatoria y el tipo de función kernel con el que se quiere trabajar, o su generalización multidimensional dada por la función *kde*. Incluso métodos menos generalizados, como los métodos bootstrap, se pueden realizar con la función *kernelboot*. Sin embargo, puesto que analizaremos a nivel teórico las propiedades y los fundamentos de los diferentes métodos de estimación, construiremos nuestras propias funciones en R, y con las que aplicaremos dichos métodos. Por este motivo, aunque los códigos se presentan como un anexo, son una parte tan importante del trabajo como las demostraciones y desarrollos teóricos que contiene.

1. Introducción

1.1. Conceptos preliminares

1.1.1. Función de distribución y de densidad

En el presente trabajo, denotaremos como *función de distribución* de una variable aleatoria X a la probabilidad de que dicha variable aleatoria tome un valor igual o inferior a x , es decir, $P(X \leq x)$. Dicha función quedará denotada por $F(x)$, y por definición de probabilidad será una función monótona creciente que toma valores en el intervalo $[0, 1]$. Si bien las variables aleatorias pueden presentar una función de distribución escalonada (en el que caso de que la variable aleatoria tome valores discretos), en el presente trabajo nos centraremos en las variables aleatorias que presentan una función de distribución absolutamente continua. Para estas variables aleatorias la función de distribución $F(x)$ se puede caracterizar a partir de la denominada *función de densidad de probabilidad*, que se corresponde con la derivada de la función de distribución y que denotaremos como $f(x)$. Esto conduce a la siguiente definición:

Definición 1.1. La función de distribución $F(x)$ para una variable aleatoria continua X viene dada por:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(y) dy \quad (1.1)$$

donde la función $f(x)$ es la función de densidad de dicha variable aleatoria que cumple:

$$a) \ f(x) \geq 0 \ \forall x \in \mathbb{R} \quad ; \quad b) \ \int_{-\infty}^{\infty} f(y) dy = 1 \quad (1.2)$$

1.1.2. Muestras aleatorias simples

Consideremos que tenemos un conjunto de valores (X_1, X_2, \dots, X_m) obtenidos a partir de realizar mediciones en m individuos elegidos al azar de la población de la variable aleatoria X . Dicho conjunto de mediciones se denomina *muestra aleatoria* de la variable X . En el caso de que las mediciones se hayan realizado de forma independiente, entonces diremos que la muestra aleatoria es *simple*, es decir:

Definición 1.2. Una *muestra aleatoria simple* de una variable aleatoria X de tamaño m es un vector m -dimensional (X_1, X_2, \dots, X_m) donde cada componente tiene el mismo comportamiento (el asociado a la variable aleatoria X) y que han sido elegidas de forma independiente unas de otras, es decir, las variables X_i son independientes e idénticamente distribuidas (de aquí en adelante i.i.d.) $\forall i \in \{1, 2, \dots, m\}$.

1.1.3. Función de distribución empírica

Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) de una variable aleatoria X , podemos aproximar la función de distribución asociada $F(x)$ por su estimador $\hat{F}(x)$, la denominada *función de distribución empírica*.

Definición 1.3. La función de distribución empírica es un funcional aleatorio $\hat{F}(x)$, definido en el conjunto de muestras como:

$$\begin{aligned} \hat{F} : \quad \mathcal{X}^m &\longrightarrow \mathcal{F} \\ (X_1, X_2, \dots, X_m) &\longmapsto \hat{F} : \mathbb{R} \rightarrow [0, 1] \\ x &\mapsto \hat{F}(x) \equiv \hat{F}(x) = \frac{\#(X_i \leq x)}{m} \end{aligned} \quad (1.3)$$

donde $\#$ denota el cardinal.

Así expresada, observamos que la función $\hat{F}(x)$ toma valores en $[0, 1]$ y que es una función no decreciente y escalonada, que estima la función de distribución en el punto x por la proporción de valores menores o iguales que x en la muestra aleatoria simple. Esta aproximación es muy importante ya que el siguiente teorema garantiza un buen comportamiento.

Teorema 1.4 (Teorema de Glivenko-Cantelli, 1933). Sean X una variable aleatoria con función de distribución $F(x)$, (X_1, X_2, \dots, X_m) una muestra aleatoria simple de dicha variable y $\hat{F}(x)$ la función de distribución empírica asociada en el punto x . Definimos el siguiente estadístico:

$$\begin{aligned} D_m : \quad \mathbb{R}^m &\longrightarrow \mathbb{R} \\ (X_1, X_2, \dots, X_m) &\longmapsto D_m \equiv \sup_{x \in \mathbb{R}} \{|F(x) - \hat{F}(x)|\} \end{aligned}$$

Se verifica entonces que:

$$D_m \xrightarrow[c.s.]{} 0$$

El teorema de Glivenko-Cantelli¹ nos garantiza que, controlando el tamaño muestral m , podemos asegurar que la función de distribución $F(x)$ se encontrará en una banda de confianza alrededor de la función de distribución empírica $\hat{F}(x)$, cuya anchura dependerá de dicho tamaño. Es decir, podemos garantizar que $F(x) \in [\hat{F}(x) - \delta(m), \hat{F}(x) + \delta(m)]$ con probabilidad 1 si el tamaño de muestra m es lo suficientemente grande.

1.1.4. Propiedades de interés de los estimadores

En (1.3) hemos definido un estimador para la función de distribución. De forma general, para evaluar la calidad de un estimador se analizan una serie de propiedades del mismo, entre las que destacamos el sesgo, el error cuadrático medio (de aquí en adelante MSE) para estimadores puntuales y el error cuadrático medio integral (de aquí en adelante MISE²) para estimadores funcionales.

¹Hemos omitido la demostración de dicho teorema en este trabajo ya que se corresponde con un contenido propio del grado de Matemáticas.

²Las notaciones introducidas de MSE y MISE responden a las siglas empleadas con mayor frecuencia en la literatura anglosajona.

Definición 1.5. El sesgo de un estimador, en este caso de $\hat{F}(x)$ estimador puntual de $F(x)$, se define como la diferencia entre el valor esperado del estimador y el verdadero valor del mismo, es decir:

$$\text{sesgo}(\hat{F}(x)) = E(\hat{F}(x)) - F(x) \quad (1.4)$$

El sesgo se puede interpretar entonces como una medida de la desviación del estimador³. Los estimadores que cumplen que su sesgo es nulo se denominan *estimadores insesgados*, mientras que si el sesgo se hace nulo cuando el tamaño de muestra tiende al infinito ($m \rightarrow \infty$) diremos que el estimador es *asintóticamente insesgado*.

Proposición 1.6. La función de distribución empírica $\hat{F}(x)$ es un estimador insesgado⁴ de $F(x)$.

Como el sesgo en sí mismo no es un error de la estimación, sino que nos indica cuánto se aleja el valor esperado de nuestro procedimiento de lo que deseamos estimar, es importante introducir alguna magnitud que permiten evaluar la precisión del estimador.

Definición 1.7. El error cuadrático medio (MSE) de un estimador, en este caso de $\hat{F}(x)$ estimador puntual de $F(x)$, se define como:

$$\text{MSE}(\hat{F}(x)) = E\left(\left(\hat{F}(x) - F(x)\right)^2\right) = \text{sesgo}^2(\hat{F}(x)) + \text{Var}(\hat{F}(x)) \quad (1.5)$$

Esta propiedad nos permite evaluar la precisión del estimador ya que además del sesgo tiene en cuenta la varianza del estimador: pequeños valores del MSE nos indican que el estimador no se alejará notablemente de los valores reales de $F(x)$.

En los estudios de grado hemos empleado tanto el sesgo como el MSE en el contexto de la estimación paramétrica, pero también se pueden definir para la estimación puntual de funciones. Sin embargo, en este último caso el MSE pasa a ser un valor dependiente de x , por lo que es interesante tener una medida global del MSE en todos los posibles valores del espacio muestral.

Definición 1.8. El error cuadrático medio integral (MISE) de un estimador funcional, en este caso de \hat{F} estimador de F , se define como:

$$\text{MISE}(\hat{F}) = \int_{-\infty}^{\infty} \text{MSE}(\hat{F}(x)) dx = E\left(\int_{-\infty}^{\infty} \left(\hat{F}(x) - F(x)\right)^2 dx\right) = \int_{-\infty}^{\infty} E\left(\left(\hat{F}(x) - F(x)\right)^2\right) dx \quad (1.6)$$

Esta propiedad tiene la misma interpretación del MSE, pero ahora consideramos la calidad de la estimación global, y no puntual.

1.2. Ejemplo de cuásares SDSS

Dado que nuestro objetivo es analizar, con rigor matemático, algunos de los procedimientos estadísticos que se aplican con frecuencia en la astrofísica, nos centraremos en el análisis de la distribución de cuásares

³En base a esta definición, es frecuente encontrar en la literatura física la expresión *off-set*, por su analogía con medidas experimentales.

⁴Omitimos la demostración de este resultado ya que se trata de un contenido propio del grado de Matemáticas.

en el Universo. Los cuásares se corresponden con objetos extragalácticos considerablemente brillantes y que están formados por galaxias con agujeros negros supermasivos en su interior. A lo largo del presente trabajo, aplicaremos los procedimientos analizados con el ejemplo de los datos recogidos en la campaña de observación *Sloan Digital Sky Survey* (de aquí en adelante SDSS) la cual se encuentra disponible en <http://astrostatistics.psu.edu/MSMA/datasets/SDSS/QSO.dat>. En la [Figura 1.1](#) se muestra un ejemplo de la muestra.

	SDSS	z	u_mag	sig_u_mag	g_mag	sig_g_mag	r_mag	sig_r_mag	i_mag	sig_i_mag	z_mag	sig_z_mag
1	000006.53+003055.2	1.8227	20.389	0.066	20.468	0.034	20.332	0.037	20.099	0.041	20.053	0.121
2	000008.13+001634.6	1.8365	20.233	0.054	20.200	0.024	19.945	0.032	19.491	0.032	19.191	0.068
3	000009.26+151754.5	1.1986	19.921	0.042	19.811	0.036	19.386	0.017	19.165	0.023	19.323	0.069
4	000009.38+135618.4	2.2400	19.218	0.026	18.893	0.022	18.445	0.018	18.331	0.024	18.110	0.033
5	000009.42+102751.9	1.8442	19.249	0.036	19.029	0.027	18.980	0.021	18.791	0.018	18.751	0.047
6	000011.41+145545.6	0.4596	19.637	0.030	19.466	0.024	19.362	0.022	19.193	0.025	19.005	0.047
7	000011.96+000225.3	0.4790	18.237	0.028	17.971	0.020	18.025	0.019	17.956	0.014	17.911	0.029
8	000012.25+003220.5	1.4365	20.424	0.068	20.502	0.033	20.253	0.028	20.148	0.037	20.381	0.160
9	000013.14+141034.6	0.9491	19.519	0.034	19.281	0.028	19.115	0.016	19.155	0.024	19.071	0.053
10	000013.80+005446.8	1.8361	20.541	0.063	20.212	0.034	20.062	0.027	19.602	0.030	19.477	0.075
11	000014.82+011030.6	1.8902	19.484	0.038	19.366	0.031	19.322	0.022	19.157	0.025	19.200	0.062
12	000015.47+005246.8	1.8476	19.777	0.046	19.615	0.023	19.551	0.022	19.217	0.036	19.138	0.062
13	000016.43+001833.4	0.7026	18.782	0.032	18.129	0.024	17.907	0.014	17.788	0.019	17.698	0.025
14	000017.38+085123.7	1.2499	19.151	0.034	18.722	0.020	18.263	0.021	18.276	0.036	18.260	0.037
15	000017.88+082612.6	0.5459	20.183	0.059	19.911	0.028	19.858	0.031	19.641	0.030	19.609	0.083
16	000024.02+152005.4	0.9885	19.413	0.044	19.183	0.035	18.988	0.015	19.079	0.023	19.133	0.060
17	000026.29+134604.6	0.7678	19.345	0.030	18.998	0.023	18.922	0.023	19.010	0.022	18.838	0.042
18	000028.82+102755.7	1.1377	20.461	0.086	19.697	0.030	19.176	0.016	19.143	0.023	19.105	0.061
19	000030.37+002732.4	1.8030	20.415	0.070	20.251	0.030	20.146	0.026	19.839	0.030	19.721	0.090
20	000031.86+010305.2	1.0925	20.566	0.070	20.100	0.023	19.649	0.020	19.596	0.030	19.568	0.083

Figura 1.1: Ejemplo de los 20 primeros datos que se presentan en el archivo.

La muestra la componen un total de 77429 cuásares y, pese a la cantidad de datos que se muestran para cada uno de ellos, nos centraremos en el dato representado por z , que se corresponde con el redshift. El redshift es una forma indirecta de medir la distancia de los objetos en un Universo en expansión, y que además se encuentra relacionado intrínsecamente con el tiempo transcurrido desde el Big Bang. De esta manera, analizar la distribución de cuásares en el Universo en base al redshift es una forma informal indirecta de analizar en qué época del Universo tuvieron mayor presencia estos objetos.

Definición 1.9. Sea una fuente que realiza una emisión en reposo con longitud de onda λ_{lab} y que es detectada con una longitud de onda λ_{obs} , el redshift z se define como el cociente:

$$z = \frac{\lambda_{obs} - \lambda_{lab}}{\lambda_{lab}} \quad (1.7)$$

La definición que presentamos del redshift se corresponde con la forma empírica de determinarlo. El cambio en la longitud se debe al fenómeno físico del Efecto Doppler, el cual establece que las fuentes emisoras en movimiento son detectadas con una frecuencia (alternativamente longitud de onda) diferente a cuando se encuentran en reposo. De esta manera, se determina a qué velocidad se mueve el objeto emisor. El modelo actual del Universo en expansión establece además que la velocidad de movimiento v de las fuentes cosmológicas se relaciona con su distancia d como $v = H_0 d$, de acuerdo a la conocida Ley de Hubble. De esta manera, se relacionan el redshift y la distancia.

Proposición 1.10. Para fuentes emisoras con velocidades no relativistas $v \ll c$, se verifica la siguiente

relación⁵ entre su distancia d y el redshift z :

$$z \sim \frac{H_0}{c}d \quad (1.8)$$

siendo c la velocidad de la luz y H_0 la constante de Hubble.

La distribución de los cuásares en función del redshift es una cuestión bastante analizada desde mediados de los años 70, apenas una década después de la primera detección de un cuásar en 1963 por parte del astrónomo Maarten Schmidt. Ejemplos de trabajos en este campo se pueden observar en las publicaciones de Karlsson[10], Zhou y su equipo de trabajo[21] o Bell y McDiarmid[1], entre otros muchos. Como no existe un modelo teórico puro que explique la distribución de cuásares en función de z , a lo largo del presente trabajo asumiremos como la distribución verdadera aquella que nos proporciona la muestra completa de cuásares, y para evaluar las técnicas aplicadas tomaremos muestras aleatorias simples a partir de la muestra completa.

⁵Omitimos la demostración de este resultado ya que se trata de un contenido propio del grado de Física.

2. Estimación de la función de densidad

2.1. Introducción

Con *estimación de densidad* entendemos, usualmente, el conjunto de procedimientos estadísticos empleados para estimar la función de densidad $f(x)$ de una variable aleatoria X , a partir de un conjunto de observaciones de la misma. Dicha estimación la denotaremos por $\hat{f}(x)$ en este trabajo.

En el presente capítulo nos centraremos en estimaciones no paramétricas de la densidad, con el objetivo de buscar una forma suave de caracterizar dicha distribución (curva o superficie, dependiendo del número de variables aleatorias). Así, habiendo determinado la función de densidad, no sólo podremos estimar la función de distribución por (1.1), sino que además podemos determinar otros parámetros de la distribución, como la esperanza o la varianza de una forma más sencilla.

2.2. Histogramas

Consideremos en primer lugar el método más sencillo de estimación de una función de densidad: el histograma. Sea X una variable aleatoria real y sea (X_1, X_2, \dots, X_m) una muestra aleatoria simple de la misma. El estimador histograma considera una descomposición de la recta real en clases C_j y para cada clase se estima la probabilidad asociada $p_j = P(X \in C_j)$ en base a la muestra por:

$$\hat{p}_j = \frac{\sum_{i=1}^m I_{C_j}(X_i)}{m}$$

Por lo tanto, dado que la función de densidad tiene que cumplir la restricción (1.3) y suponiendo que toma valores constantes en las clases C_j , tenemos que:

$$p_j = \int_{C_j} f(y) dy = f_j \ell_j$$

donde ℓ_j es la amplitud de la clase C_j y f_j es el valor que toma la función de densidad en dicha clase, que estamos suponiendo es constante, tenemos entonces que:

$$\hat{f}_j = \frac{\hat{p}_j}{\ell_j} = \frac{\sum_{i=1}^m I_{C_j}(X_i)}{m \cdot \ell_j}$$

Definición 2.1. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) , se define el estimador histograma

para la función de densidad como la función:

$$\hat{f}_H(x) = \sum_j \hat{f}_j I_{C_j}(x) \quad (2.1)$$

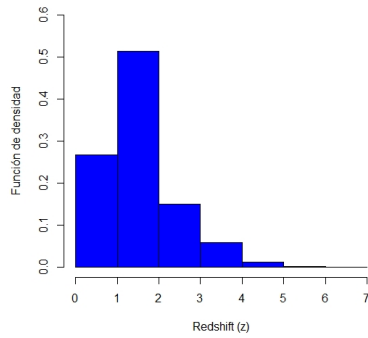
Tenemos, en consecuencia, una estimación de la función de densidad \hat{f} , dada por una función constante a trozos $\hat{f}(x) = \hat{f}_j$ si $x \in C_j$.

Proposición 2.2. *El estimador histograma es una función de densidad.*

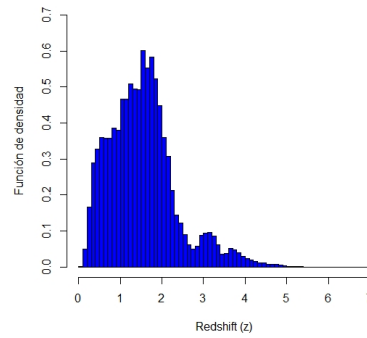
Demostración. Para que el estimador histograma sea una función de densidad debe cumplir las condiciones a) y b) introducidas en (1.2). Como la función indicador I sólo toma valores 0 o 1, deducimos que $\hat{f}_H(x) \geq 0 \forall x \in \mathbb{R}$, cumpliéndose a). Por otro lado, para demostrar que se cumple b) realizamos el cálculo:

$$\int_{-\infty}^{\infty} \hat{f}_H(y) dy = \sum_j \int_{C_j} \hat{f}_j dy = \sum_j \hat{f}_j \ell_j = \sum_j \frac{\sum_{i=1}^m I_{C_j}(X_i)}{m} = \frac{\sum_{i=1}^m \sum_j I_{C_j}(X_i)}{m} = 1$$

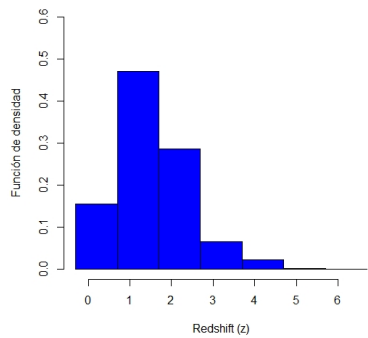
□



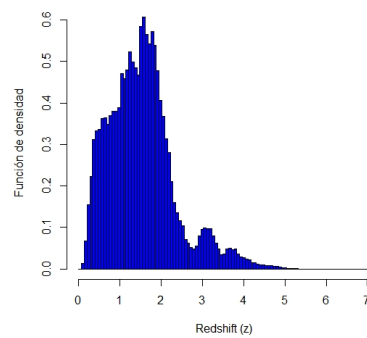
(a) Histograma con $x_0 = 0$ y $h = 1$.



(b) Histograma con $x_0 = 0$ y $h = 0.1$.



(c) Histograma con $x_0 = -0.3$ y $h = 1$.



(d) Histograma con $x_0 = 0$ y $h = 0.0673$.

Figura 2.1: Histogramas para la muestra de cuásares de SDSS con variación en la definición de las clases.

Llegados a este punto, antes de analizar las propiedades de este estimador podemos observar que existe una clara ambigüedad en su definición: es necesario establecer la amplitud ℓ_j y el punto de comienzo de la primera clase. Por lo general, la construcción de las clases se suele realizar de la siguiente manera: fijamos un punto inicial x_0 y una amplitud ℓ_j , la cual denominaremos *amplitud de ventana*, que usualmente tomaremos constante para todas las clases $\ell_j = h \forall j$; la clase C_k se corresponde con el intervalo (*bin*) dado por $C_k \equiv [x_0 + kh, x_0 + (k + 1)h)$. Para visualizar el problema anteriormente comentado, consideremos el ejemplo propuesto en la [Subsección 1.1.2](#). En la [Figura 2.1](#) se presentan varios histogramas para la misma muestra. Claramente los histogramas (a) y (c), que dividen la recta real en clases de la misma longitud, son diferentes como consecuencia de la definición del punto inicial. Por su parte, los histogramas (a) y (b) muestran que la determinación del tamaño de las clases también tiene una gran influencia: si tomamos intervalos muy grandes, perdemos gran cantidad de información respecto a la distribución de la variable; si por otro lado se toman intervalos muy pequeños, el histograma puede ajustarse demasiado a la muestra aleatoria, lo que conduciría a mucha varianza en los histogramas para distintas muestras de la variable.

Para analizar las propiedades este estimador, comenzaremos estudiando el sesgo y la varianza, para así poder determinar el MSE.

Proposición 2.3. *Sea X una variable aleatoria con función de densidad $f(x)$ tal que es derivable y de derivada acotada $|f'(x)| \leq L$. Se verifica entonces que el sesgo y la varianza del estimador histograma en un punto $x \in C_j$ vienen dados por:*

$$\text{sesgo}(\hat{f}_H(x)) \leq L\ell_j \quad (2.2)$$

$$\text{Var}(\hat{f}_H(x)) = \frac{1}{m \cdot \ell_j} f(x^*) (1 - \ell_j f(x^*)) \quad (2.3)$$

con $x^* \in C_j$, ℓ_j la amplitud de la clase y m el tamaño de la muestra.

Demostración. Por construcción, las clases C_j constituyen una partición de \mathbb{R} , por lo que existe un j tal que $x \in C_j$. En dicho punto, el estimador histograma toma el valor \hat{f}_j , con lo que la esperanza vendrá dada por:

$$E(\hat{f}(x)) = E(\hat{f}_j) = \frac{1}{m \cdot \ell_j} \sum_{i=1}^m E(I_{C_j}(X_i)) \stackrel{i.i.d.}{=} \frac{1}{\ell_j} E(I_{C_j}(X)) = \frac{p_j}{\ell_j}$$

Por otro lado, la esperanza de que la variable X se encuentre en la clase C_j es precisamente la probabilidad real de dicha clase, p_j . Puesto que la probabilidad p_j es, por definición, $\int_{C_j} f(y) dy$, por el Teorema del Valor Medio de la Integración, podemos deducir que, si f es continua en C_j :

$$E(\hat{f}_j) = f(x^*) \quad \text{para algún } x^* \in C_j$$

Aplicando ahora el Teorema del Valor Medio de la Derivada, y como por hipótesis la función densidad es derivable y con derivada acotada $|f'(x)| \leq L$, podemos expresar el sesgo como:

$$\text{sesgo}(\hat{f}_j) = f(x^*) - f(x) = f'(x^{**}) [x^* - x] \leq |f'(x^{**})| |x^* - x| \leq L\ell_j$$

Por otra parte, un estudio de la varianza del estimador nos permite obtener que:

$$\begin{aligned}\text{Var}(\hat{f}_j) &= \frac{1}{m^2 \ell_j^2} \text{Var} \left(\sum_{i=1}^m I_{C_j}(X_i) \right) \stackrel{i.i.d.}{=} \frac{1}{m \cdot \ell_j^2} \text{Var}(I_{C_j}(X)) = \\ &= \frac{1}{m \cdot \ell_j^2} p_j (1 - p_j) = \frac{1}{m \cdot \ell_j} f(x^*) (1 - \ell_j f(x^*))\end{aligned}$$

□

El sesgo y la varianza muestran claramente los problemas que habíamos comentado con anterioridad respecto a la elección del tamaño de los intervalos. Si elegimos intervalos muy pequeños, entonces el sesgo se hará cada vez más pequeño, pudiendo aproximar mejor en promedio la función de densidad en el punto x . No obstante, hay que tener en cuenta que la varianza tiene una dependencia inversa respecto a dicho parámetro: reducir el tamaño implica también un aumento de la varianza en la estimación.

Para evitar un problema de elección entre un sesgo razonable o una varianza razonable, se busca la amplitud de clase ℓ_k que optimiza el MSE (1.5)[3].

Corolario 2.4. *La amplitud de clase $\tilde{\ell}_j$ que minimiza el MSE del estimador histograma viene dada por:*

$$\tilde{\ell}_j = \left(\frac{f(x^*)}{2mL^2} \right)^{\frac{1}{3}} \quad (2.4)$$

Demostración. Para calcular el MSE nos basamos en los resultados obtenidos de la Proposición 2.3, con lo que se obtiene:

$$\text{MSE}(\hat{f}(x)) \leq L^2 \ell_j^2 + \frac{1}{m \cdot \ell_j} f(x^*) (1 - \ell_j f(x^*))$$

Para buscar el valor que minimice la cota del MSE, derivamos dicha cota e igualamos a 0:

$$\begin{aligned}\frac{d\text{MSE}}{d\ell_j} &= 2L^2 \ell_j - \frac{f(x^*)}{m \cdot \ell_j^2} \\ \left| \frac{d\text{MSE}}{d\ell_j} \right|_{\tilde{\ell}_j} &= 0 \Rightarrow \tilde{\ell}_j = \left(\frac{f(x^*)}{2mL^2} \right)^{\frac{1}{3}}\end{aligned}$$

Calculamos la segunda derivada para comprobar que el extremo relativo obtenido es un mínimo:

$$\begin{aligned}\frac{d}{dh} \left[2L^2 \ell_j - \frac{f(x^*)}{m \cdot \ell_j^2} \right] &= 2L^2 + 2 \frac{f(x^*)}{m \cdot \ell_j^3} \Rightarrow \\ 2L^2 + 2 \frac{f(x^*)}{m \cdot \tilde{\ell}_j^3} &= 4L^2 > 0\end{aligned}$$

□

Este resultado evidencia dos hechos muy importantes:

1. Es imposible conocer dada una muestra aleatoria simple el tamaño óptimo de la ventana empleado en el histograma, ya que de los parámetros implicados sólo conocemos el tamaño m .

2. El intervalo óptimo depende del valor de la densidad en el punto $x^* \in C_j$, asentando así la base de trabajo para el estudio de estimadores con anchuras de intervalo variables.

En el ejemplo (d) de la [Figura 2.1](#) se muestra una solución a este problema propuesta por Scott en 1979. Dicha alternativa se basa en considerar una aproximación de normalidad en la variable aleatoria y, basándose en la estimación de la desviación estándar $\hat{\sigma}$, definir una anchura de intervalo $h = 3.49\hat{\sigma}m^{-1/3}$ [16]. En la [Subsección 3.2.2](#) justificaremos esta elección.

2.3. Estimador naïve

El estimador naïve, también denominado *histograma móvil*, responde a la interpretación de la función de densidad como derivada de la función de distribución, $f(x) = \lim_{h \rightarrow 0} \frac{F(x+h) - F(x-h)}{2h}$. Por lo tanto, fijando un valor de h pequeño, e introduciendo la estimación de la función de distribución (1.3), se tiene:

$$\hat{f}(x) = \frac{\hat{F}(x+h) - \hat{F}(x-h)}{2h} = \frac{1}{2mh} \sum_{i=1}^m I_{[x-h, x+h]}(X_i)$$

Definición 2.5. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) , se define el estimador naïve de la función de densidad como:

$$\hat{f}_{naive}(x) = \frac{1}{2mh} \sum_{i=1}^m I_{[x-h, x+h]}(X_i) = \frac{1}{2mh} \sum_{i=1}^m I_{[-1, 1]} \left(\frac{x - X_i}{h} \right) \quad (2.5)$$

Proposición 2.6. El estimador naïve es una función de densidad.

Demostración. Para que el estimador naïve sea una función de densidad debe cumplir las condiciones a) y b) introducidas en (1.2). Como la función indicador I sólo toma valores 0 o 1, deducimos que $\hat{f}_{naive}(x) \geq 0 \forall x \in \mathbb{R}$, cumpliéndose a). Por otro lado, podemos observar que:

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{f}_{naive}(x) dx &= \frac{1}{2mh} \int_{-\infty}^{\infty} \sum_{i=1}^m I_{[-1, 1]} \left(\frac{x - X_i}{h} \right) dx = \frac{1}{2mh} \sum_{i=1}^m \left[\int_{-\infty}^{\infty} I_{[-1, 1]} \left(\frac{x - X_i}{h} \right) dx \right] = \\ &= \frac{1}{2mh} \sum_{i=1}^m \left[\int_{-h+X_i}^{h+X_i} dx \right] = \frac{1}{2mh} \sum_{i=1}^m 2h = 1 \end{aligned}$$

con lo que demostramos que se cumple b) también. □

Si bien este estimador guarda un gran parecido con el estimador histograma definido en (2.1), presenta una diferencia esencial: las clases se centran en cada punto x y tienen una amplitud fija h .

Definición 2.7. El valor h empleado en la construcción de las clases del estimador naïve recibe el nombre de *amplitud de ventana* o *anchura de ventana*.

De esta manera, solucionamos el problema de la arbitrariedad en la elección de las clases, pero se mantiene el problema de la elección la anchura de la ventana. La [Figura 2.2](#) muestra las estimaciones de la densidad para diferentes valores de h .

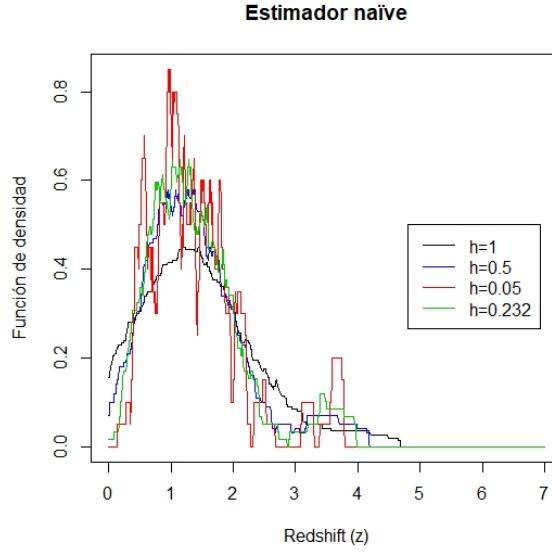


Figura 2.2: Ejemplos del estimador naïve del redshift de una muestra aleatoria de 1000 cuásares en la campaña SDSS.

Por construcción, el estimador naïve es también una función escalonada (véase la [Figura 2.2](#)), como el histograma, pero la amplitud de ventana h nos permite suavizar dicha función. A nivel descriptivo, podemos observar que elegir un valor de h pequeño conduce a un menor sesgo del estimador pero aumenta su varianza, resultado ya comentado con el estimador histograma. Para justificar estos resultados, analizaremos el sesgo y la varianza del estimador.

Proposición 2.8. Sean X una variable aleatoria continua con función de densidad $f(x) \in \mathcal{C}^2$, una muestra aleatoria simple de dicha variable (X_1, X_2, \dots, X_m) y \hat{f}_{naive} el estimador naïve asociado. El sesgo y la varianza del estimador en el punto x verifican:

$$\text{sesgo} \left(\hat{f}_{naive}(x) \right) = \frac{h^2}{6} f''(x) + o(h^2) \quad (2.6)$$

$$\text{Var} \left(\hat{f}_{naive}(x) \right) = \frac{1}{2hm} f(x) + o\left(\frac{1}{mh}\right) \quad (2.7)$$

Demostración. Para el cálculo del sesgo y la varianza, podemos observar que en ambos casos llegamos a las siguientes expresiones:

$$E \left(\hat{f}_{naive}(x) \right) = \frac{1}{2mh} \sum_{i=1}^m E \left(I_{[-1,1]} \left(\frac{x - X_i}{h} \right) \right) \stackrel{i.i.d.}{=} \frac{1}{2h} E \left(I_{[x-h, x+h]}(X) \right)$$

$$\text{Var} \left(\hat{f}_{naive}(x) \right) \stackrel{X_i \text{ i.i.d.}}{=} \frac{1}{4h^2m} \text{Var} \left(I_{[x-h, x+h]}(X) \right)$$

es decir, pasamos a trabajar con la variable aleatoria $Y = I_{[x-h, x+h]}(X)$, que sigue una distribución de Bernoulli cuyo parámetro característico será la probabilidad de la clase $[x-h, x+h]$, y que denotaremos como $p(x \pm h)$. La esperanza y la varianza serán $p(x \pm h)$ y $p(x \pm h)[1 - p(x \pm h)]$ respectivamente. Por

definición de función de densidad, podemos calcular $p(x \pm h)$ como:

$$p(x \pm h) = F(x + h) - F(x - h) = \int_{x-h}^{x+h} f(y) dy$$

y aplicando la hipótesis de que $f \in \mathcal{C}^2$, obtenemos entonces:

$$\begin{aligned} p(x \pm h) &= \int_{x-h}^{x+h} \left[f(x) + (y-x)f'(x) + \frac{1}{2}(y-x)^2 f''(x) + o(|y-x|^3) \right] dy = \\ &= 2hf(x) + \frac{h^3}{3} f''(x) + o(|h^3|) \end{aligned}$$

De esta manera, por (1.4) el sesgo será:

$$\text{sesgo}(\hat{f}_{naive}(x)) = \frac{h^2}{6} f''(x) + o(h^2) \sim \frac{h^2}{6} f''(x)$$

y la varianza:

$$\begin{aligned} \text{Var}(\hat{f}_{naive}(x))_{X_i \text{ i.i.d.}} &= \frac{1}{4h^2 m} \text{Var}(I_{[x-h, x+h]}(X)) = \frac{p(x \pm h)(1 - p(x \pm h))}{4h^2 m} = \\ &= \frac{1}{2hm} f(x) + \frac{h}{12m} f''(x) + \frac{1}{m} f(x)^2 + o(h) = \frac{1}{2hm} f(x) + o\left(\frac{1}{mh}\right) \sim \frac{1}{2hm} f(x) \end{aligned}$$

□

Para afrontar de nuevo el problema de optimizar la anchura de la ventana, suponiendo que h es el mismo para todas las clases, podemos realizar el tratamiento con el MISE en lugar de emplear el MSE, como se procedió en el caso particular de los histogramas.

Corolario 2.9. *Bajo las hipótesis de la Proposición 2.8, el MISE para el estimador naïve viene dado por:*

$$\text{MISE}(\hat{f}_{naive}) = \frac{1}{2mh} + \frac{h^4}{36} R(f) + o\left(\frac{1}{mh} + h^4\right) \quad (2.8)$$

donde se define $R(f) := \int_{-\infty}^{\infty} (f''(x))^2 dx$.

Demostración. Empleando la Proposición 2.8 y de acuerdo a la definición de MISE (1.6), tenemos entonces que:

$$\begin{aligned} \text{MISE}(\hat{f}_{naive}(x)) &= \int_{-\infty}^{\infty} \left[\frac{1}{2mh} f(x) + \frac{h^4}{36} (f''(x))^2 + o\left(\frac{1}{mh} + h^4\right) \right] dx = \\ &= \frac{1}{2mh} \int_{-\infty}^{\infty} f(x) dx + \frac{h^4}{36} \int_{-\infty}^{\infty} (f''(x))^2 dx + o\left(\frac{1}{mh} + h^4\right) = \\ &= \frac{1}{2mh} + \frac{h^4}{36} R(f) + o\left(\frac{1}{mh} + h^4\right) \end{aligned}$$

□

Como el MISE se obtiene de realizar desarrollos en serie, es bastante útil trabajar con la *parte asintótica*

del mismo, es decir, la parte dominante del MISE cuando $h \rightarrow 0$ y $m \rightarrow \infty$, que en el caso particular del Corolario 2.9, toma la expresión:

$$\text{MISE}(\hat{f}_{naive}) \sim \frac{1}{2mh} + \frac{h^4}{36}R(f) \quad (2.9)$$

asumiendo que $R(f)$ toma un valor finito⁶.

Corolario 2.10. *La amplitud de ventana que minimiza la parte asintótica del MISE del estimador naïve viene dada por:*

$$\tilde{h} = \left[\frac{9}{2R(f)} \right]^{\frac{1}{5}} m^{-1/5} \quad (2.10)$$

y recibe el nombre de anchura de ventana óptima o anchura óptima.

Demostración. Considerando la parte asintótica dada por (2.9), derivando respecto de h e igualando a 0 para encontrar los extremos relativos obtenemos:

$$\begin{aligned} \frac{d}{dh} \left[\frac{1}{2mh} + \frac{h^4}{36}R(f) \right] &= -\frac{1}{2mh^2} + \frac{h^3}{9}R(f) \\ -\frac{1}{2m\tilde{h}^2} + \frac{\tilde{h}^3}{9}R(f) &= 0 \Rightarrow \tilde{h} = \left[\frac{9}{2R(f)} \right]^{\frac{1}{5}} m^{-1/5} \end{aligned}$$

Si calculamos la segunda derivada y evaluamos, comprobamos que efectivamente es un mínimo:

$$\begin{aligned} \frac{d}{dh} \left[-\frac{1}{2mh^2} + \frac{h^3}{9}R(f) \right] &= \frac{1}{mh^3} + \frac{h^2}{3}R(f) \Rightarrow \\ \frac{1}{m\tilde{h}^3} + \frac{\tilde{h}^2}{3}R(f) &= \left[\left(\frac{2^3}{3^6} \right)^{\frac{1}{5}} + \left(\frac{1}{6} \right)^{\frac{2}{5}} \right] R(f)^{\frac{3}{5}} m^{-\frac{2}{5}} > 0 \end{aligned}$$

□

La anchura óptima calculada en (2.10) presenta resultados análogos a los discutidos en la anchura óptima del histograma (2.4). Ambas muestran que para grandes tamaños de muestra, las anchuras se deben escoger más pequeñas, pero también que es necesario tener un conocimiento de la función de densidad real (algo que no es posible en la estimación no paramétrica). Sin embargo, el estimador naïve requiere de una mayor regularidad de la función de densidad ($f \in \mathcal{C}^{(2)}$) que el histograma ($f \in \mathcal{C}^{(1)}$).

Es importante resaltar que tanto el estimador $\hat{f}(x)$ como el estimador $\hat{f}_{naive}(x)$ proporcionan funciones escalonadas, si bien el salto podemos suavizarlo en base a las amplitudes de las clases ℓ_k o de h . Este escalonamiento es consecuencia del hecho de que el estimador otorga el mismo peso, en términos de la función densidad, a todos los puntos en el intervalo $[x-h, x+h]$. Para solucionar este problema, se introducen los denominados *estimadores kernel*, que estudiaremos a continuación. Sin embargo, hay que tener presente la función de densidad que se desea estimar: no tiene sentido plantearse la suavización de la

⁶Esta acotación tiene una interpretación en términos de suavidad (o rigidez) de la curva $f(x)$. De hecho, este término se suele utilizar como un funcional de penalización para estimar funciones de densidad por el método de máxima verosimilitud[17].

estimación en los extremos cuando la función de densidad es uniforme, ya que no cumple las condiciones de regularidad, y es el motivo por el cual algunos estimadores kernel no funcionan bien.

2.4. Estimadores kernel

2.4.1. Introducción y propiedades generales

Inicialmente propuestos por Rosenblatt en 1956[14] y posteriormente analizados por Parzen en 1962[13], los *estimadores kernel* surgen como una generalización del estimador naïve (2.5), que podemos expresar de forma alternativa:

$$\hat{f}_{naive}(x) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - X_i}{h}\right) \quad \text{con} \quad K(x) = \begin{cases} \frac{1}{2} & \text{si } |x| < 1 \\ 0 & \text{resto} \end{cases}$$

de donde observamos que la función K en este caso se corresponde con la función de densidad de una distribución uniforme $U_{[-1,1]}$. Rosenblatt planteó en 1956 una generalización de las funciones K de manera que se pudieran utilizar para suavizar la propia función de densidad.

Definición 2.11. Se denomina función kernel (función núcleo) a toda función K de variable real con valores reales que verifica las propiedades:

$$\begin{aligned} a) \quad & K(x) \geq 0 \quad \forall x \in \mathbb{R} \quad ; \quad b) \quad \int_{-\infty}^{\infty} K(x) dx = 1 \\ c) \quad & \exists \kappa > 0 \text{ t.q. } \int_{|x| < \kappa} K(x) dx \sim 1 \quad ; \quad d) \quad \int_{-\infty}^{\infty} xK(x) dx = 0 \\ e) \quad & \int_{-\infty}^{\infty} x^2 K(x) dx = V_K < \infty \quad ; \quad f) \quad K(-x) = K(x) \quad \forall x \in \mathbb{R} \end{aligned}$$

Como estudiaremos a continuación, las propiedades a) y b) son necesarias para que los estimadores construidos a partir de ellas sean funciones de densidad (Proposición 2.13). Por otro lado, parece intuitivo que la información sobre el punto x en el que deseamos estimar la función de densidad de mayor peso a los valores muestrales que se encuentren próximos a él. Esta es la justificación de la propiedad c). Por su parte, el resto de propiedades nos permitirán obtener mayores órdenes de convergencia del sesgo (Proposición 2.14).

Definición 2.12. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) y una función kernel K , se define el estimador kernel como la función:

$$\hat{f}_K(x) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - X_i}{h}\right) \quad (2.11)$$

Proposición 2.13. El estimador kernel \hat{f}_K es una función de densidad.

Demostración. Veamos que el estimador kernel cumple las propiedades a) y b) (1.2) de función de densidad. Por la propiedad a) de función kernel, $K(x) \geq 0 \quad \forall x \in \mathbb{R}$, con lo que $\hat{f}_K(x) \geq 0 \quad \forall x \in \mathbb{R}$. Para ver

que se cumple la segunda condición, realizamos el siguiente cálculo:

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{f}_K(x) dx &= \int_{-\infty}^{\infty} \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - X_i}{h}\right) dx = \frac{1}{mh} \sum_{i=1}^m \int_{-\infty}^{\infty} K\left(\frac{x - X_i}{h}\right) dx = \\ &\stackrel{hy=x-X_i}{=} \frac{1}{mh} \sum_{i=1}^m \int_{-\infty}^{\infty} hK(y) dy = \frac{1}{mh} \sum_{i=1}^m h = 1 \end{aligned}$$

□

Proposición 2.14. Sean X una variable aleatoria continua con función de densidad $f(x) \in \mathcal{C}^2$, una muestra aleatoria simple de dicha variable (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a la función K . El sesgo y la varianza del estimador en el punto x verifican:

$$\text{sesgo}(\hat{f}_K(x)) = \frac{h^2}{2} f''(x) V_K + o(h^2) \quad (2.12)$$

$$\text{Var}(\hat{f}_K(x)) = \frac{1}{mh} f(x) \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} + o\left(\frac{1}{mh}\right) \quad (2.13)$$

siendo $V_K := \int_{-\infty}^{\infty} y^2 K(y) dy$, denominada varianza de la función kernel.

Demostración. Para calcular el sesgo comenzamos calculando la esperanza del estimador:

$$E(\hat{f}_K(x)) = \frac{1}{mh} \sum_{i=1}^m E\left(K\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x - y}{h}\right) f(y) dy \stackrel{x=y+\tilde{y}h}{=} \int_{-\infty}^{\infty} K(\tilde{y}) f(x - h\tilde{y}) d\tilde{y}$$

Como por hipótesis la función de densidad real $f(x) \in \mathcal{C}^2$, realizamos el desarrollo:

$$E(\hat{f}_K(x)) = \int_{-\infty}^{\infty} K(\tilde{y}) \left[f(x) - h\tilde{y}f'(x) + \frac{h^2\tilde{y}^2}{2} f''(x) + o(h^2) \right] d\tilde{y}$$

y haciendo uso de las condiciones a), d) y e) de función kernel, obtenemos finalmente:

$$E(\hat{f}_K(x)) = f(x) + \frac{h^2}{2} f''(x) V_K + o(h^2)$$

Este resultado nos permite deducir que el sesgo tomará la forma:

$$\text{sesgo}(\hat{f}_K(x)) = \frac{h^2}{2} f''(x) V_K + o(h^2) \sim \frac{h^2}{2} f''(x) V_K$$

Ahora estudiemos la varianza del estimador:

$$\begin{aligned} \text{Var}(\hat{f}_K(x)) &= \frac{1}{m^2 h^2} \text{Var}\left(\sum_{i=1}^m K\left(\frac{x - X_i}{h}\right)\right) \stackrel{X_i \text{ i.i.d.}}{=} \frac{1}{mh^2} \text{Var}\left(K\left(\frac{x - X_i}{h}\right)\right) = \\ &= \frac{1}{mh^2} \left[\int_{-\infty}^{\infty} K^2\left(\frac{x - y}{h}\right) f(y) dy - \left[hf(x) + \frac{h^3}{2} f''(x) V_K + o(h^3) \right]^2 \right] \stackrel{x=y+h\tilde{y}}{=} \end{aligned}$$

$$= \frac{1}{mh} f(x) \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} + o\left(\frac{1}{mh}\right) \sim \frac{1}{mh} f(x) \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y}$$

□

Las propiedades d) y e) nos permiten deducir los resultados (2.12) y (2.13), que además evidencian nuevamente el problema que presentaba el estimador naïve: es asintóticamente insesgado cuando $h \rightarrow 0$, pero su varianza puede crecer enormemente. Una posible forma evitar esta dicotomía es seleccionar tamaños de muestra grandes de manera que $\frac{1}{mh} \rightarrow 0$, con lo que se conseguiría que el estimador fuese asintóticamente insesgado y con una varianza cercana a cero.

2.4.2. Aproximación teórica a la ventana óptima

El cálculo del intervalo óptimo para los estimadores kernel generaliza el obtenido para el estimador naïve (2.10). Para demostrarlo, nos basamos en el siguiente resultado:

Corolario 2.15. *Bajo las condiciones de la Proposición 2.14, el MSE y el MISE del estimador kernel vienen dados por:*

$$\text{MSE}(\hat{f}_K(x)) = \frac{h^4}{4} (f''(x))^2 V_K^2 + \frac{1}{mh} f(x) \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} + o\left(\frac{1}{mh} + h^4\right) \quad (2.14)$$

$$\text{MISE}(\hat{f}_K) = \frac{h^4 V_K^2}{4} R(f) + \frac{1}{mh} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} + o\left(\frac{1}{mh} + h^4\right) \quad (2.15)$$

Demostración. Introduciendo los resultados de la Proposición 2.14 en (1.5) obtenemos:

$$\text{MSE}(\hat{f}_K(x)) = \frac{h^4}{4} (f''(x))^2 V_K^2 + \frac{1}{mh} f(x) \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} + o\left(\frac{1}{mh} + h^4\right)$$

y por definición de MISE, integrando la expresión anterior obtenemos:

$$\begin{aligned} \text{MISE}(\hat{f}_K) &= \int_{-\infty}^{\infty} \text{MSE}(\hat{f}_K(x)) dx = \frac{h^4 V_K^2}{4} \int_{-\infty}^{\infty} (f''(x))^2 dx + \frac{1}{mh} \int_{-\infty}^{\infty} f(x) dx \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} + \\ &+ o\left(\frac{1}{mh} + h^4\right) = \frac{h^4 V_K^2}{4} R(f) + \frac{1}{mh} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} + o\left(\frac{1}{mh} + h^4\right) \sim \\ &\sim \frac{h^4 V_K^2}{4} R(f) + \frac{1}{mh} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} \end{aligned}$$

□

Repitiendo el procedimiento seguido en la Sección 2.3 para estimar una amplitud de ventana óptima para el estimador naïve, buscamos la ventana que minimice la parte asintótica del MISE.

Corolario 2.16. *La amplitud de ventana óptima \tilde{h} que minimiza la parte asintótica del MISE para el estimador kernel viene dada por:*

$$\tilde{h} = \left[\frac{\int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 R(f)} \right]^{\frac{1}{5}} m^{-\frac{1}{5}} \quad (2.16)$$

Demostración. De acuerdo a los resultados del Corolario 2.15, la parte asintótica del MISE viene dada por:

$$\text{MISE}(\hat{f}_K) \sim \frac{h^4 V_K^2}{4} R(f) + \frac{1}{mh} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y}$$

Para minimizar dicha expresión, calculamos la derivada y la igualamos a 0:

$$\frac{d}{dh} \text{MISE}(\hat{f}_K) = h^3 V_K^2 R(f) - \frac{1}{mh^2} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} \Rightarrow$$

$$\tilde{h}^3 V_K^2 R(f) - \frac{1}{m\tilde{h}^2} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} = 0 \Rightarrow \tilde{h} = \left[\frac{\int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 R(f)} \right]^{\frac{1}{5}} m^{-\frac{1}{5}}$$

Ahora comprobamos que el extremo relativo calculado se trata de un mínimo:

$$\frac{d}{dh} \left[h^3 V_K^2 R(f) - \frac{1}{mh^2} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} \right] = 3h^2 V_K^2 R(f) + \frac{2}{mh^3} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} \Rightarrow$$

$$3\tilde{h}^2 V_K^2 R(f) + \frac{2}{m\tilde{h}^3} \int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} = 5 \left(\int_{-\infty}^{\infty} K^2(\tilde{y}) d\tilde{y} \right)^{\frac{2}{5}} V_K^{\frac{6}{5}} R(f)^{\frac{3}{5}} m^{-\frac{2}{5}} > 0$$

□

El resultado (2.16) es claramente la generalización de la amplitud óptima calculada para el estimador naïve (2.10) y por tanto, presenta el mismo problema: es necesario tener un conocimiento de la función de densidad real subyacente a la muestra.

Corolario 2.17. *El estimador kernel, construido con la amplitud de ventana óptima \tilde{h} , verifica que el MISE se comporta asintóticamente como un infinitésimo de orden $m^{-\frac{4}{5}}$, es decir, $\text{MISE} \sim m^{-\frac{4}{5}}$.*

Demostración. Por el Corolario 2.15 sabemos que el MISE tiene un orden de convergencia $o\left(\frac{1}{mh} + h^4\right)$ y por el Corolario 2.16 sabemos que la amplitud óptima viene dada por $\tilde{h} = Am^{-\frac{1}{5}}$ con A una constante positiva dependiente de la función kernel y de la función de densidad real subyacente de la variable aleatoria. Por lo tanto, introduciendo la amplitud de ventana óptima \tilde{h} :

$$\frac{1}{m\tilde{h}} + \tilde{h}^4 = \frac{1}{A} m^{-\frac{4}{5}} + A^4 m^{-\frac{4}{5}} = Bm^{-\frac{4}{5}}$$

□

Observamos que la amplitud decrece con el tamaño de la muestra m , pero lo hace de una forma tan lenta que el término $\frac{1}{hm} \propto m^{-\frac{4}{5}}$, por lo que sería teóricamente posible encontrar estimadores asintóticamente insesgados con varianza cercana a 0.

2.4.3. Aproximación teórica a la función kernel óptima

En la anterior subsección hemos obtenido una expresión explícita de la ventana óptima que minimiza el MISE para una función kernel genérica. Considerando la aproximación teórica del tamaño de ventana óptimo (2.16) e introduciendo dicho valor en el desarrollo asintótico del MISE, se obtiene la expresión:

$$\text{MISE}(\hat{f}_K(x)) \sim \frac{5}{4} \left(\int_{-\infty}^{\infty} K^2(y) dy \right)^{\frac{4}{5}} \left(\int_{-\infty}^{\infty} y^2 K(y) dy \right)^{\frac{2}{5}} \left(\int_{-\infty}^{\infty} f''(y) dy \right)^{\frac{1}{5}}$$

Nuestro objetivo será buscar la función kernel K que optimice la anterior expresión del MISE, en la que hay dos factores que dependen de esa función kernel. Observamos que uno de ellos se puede mantener fijo sin mayor problema, ya que el término asociado con la varianza V_K siempre se puede establecer como 1 realizando el siguiente cambio:

$$V_K = \int_{-\infty}^{\infty} y^2 K(y) dy \Rightarrow 1 = \int_{-\infty}^{\infty} \frac{y^2}{V_K} K(y) dy = \int_{\tilde{y}=yV_K^{-\frac{1}{2}}}^{\infty} \tilde{y}^2 V_K^{\frac{1}{2}} K\left(V_K^{\frac{1}{2}} \tilde{y}\right) d\tilde{y}$$

Proposición 2.18. *La función que optimiza el problema:*

$$\min_K \left(\int_{-\infty}^{\infty} K^2(y) dy \right) \quad \text{sujeeto a} \quad \begin{cases} K(y) \geq 0 \quad \forall y \in \mathbb{R} \\ \int_{-\infty}^{\infty} K(y) dy = 1 \\ \int_{-\infty}^{\infty} y K(y) dy = 0 \\ \int_{-\infty}^{\infty} y^2 K(y) dy < \infty \end{cases}$$

viene dada por:

$$K_E(x) := \frac{3}{4} (1 - x^2) I_{[-1,1]}(x) \quad (2.17)$$

y recibe el nombre de función de Epanechnikov en honor del matemático ruso que la propuso en 1969[4].

Demostración. Dado el problema de optimización del enunciado, podemos aplicar la teoría de los multiplicadores de Lagrange y la derivada de Gateaux al problema. De esta manera, bastará con encontrar la función $K(y)$ que permita minimizar:

$$\int_{-\infty}^{\infty} [K^2(y) + (-a + cy^2) K(y)] dy$$

que viene dada por $K(y) = \frac{a}{2} (1 - \frac{c}{a} y^2)$. Ahora bien, para determinar las constantes debemos introducir nuevamente las condiciones del problema. Tomando las constantes $a, c \geq 0$, entonces de la condición de no negatividad se obtiene que necesariamente $y^2 \leq \frac{a}{c}$, o equivalentemente $y \in [-\sqrt{\frac{a}{c}}, \sqrt{\frac{a}{c}}]$. Por otro lado,

aplicando las restricciones integrales del problema se obtiene:

$$\left. \begin{aligned} \int_{-\infty}^{\infty} K(y) dy = 1 &\Rightarrow \frac{2a}{3} \sqrt{\frac{a}{c}} = 1 \\ \int_{-\infty}^{\infty} y^2 K(y) dy = 1 &\Rightarrow \frac{2a}{15} \sqrt{\left(\frac{a}{c}\right)^3} = 1 \end{aligned} \right\} \Rightarrow \begin{cases} a = \frac{3}{2\sqrt{5}} \\ c = \frac{3}{10\sqrt{5}} \end{cases}$$

Es decir se obtiene la función:

$$K(y) = \frac{3}{4\sqrt{5}} \left(1 - \frac{1}{5}y^2\right) I_{[-5,5]}(y) \quad (2.18)$$

Realizando el cambio de variable $y = 5x$ y teniendo en cuenta la transformación $V_K^{\frac{1}{2}} K \left(V_K^{\frac{1}{2}} \tilde{y} \right)$, la expresión (2.18) adopta la forma:

$$K_E(x) = \frac{3}{4} (1 - x^2) I_{[-1,1]}(x)$$

□

Proposición 2.19. *La función de Epanechnikov (2.17) es una función kernel.*

Demostración. Debemos comprobar que se cumplen las 6 propiedades de la Definición 2.11. Podemos observar que las condiciones a), b), d) y e) se cumplen como consecuencia de las restricciones del problema de optimización. Por lo tanto, sólo tenemos que comprobar las condiciones c) y f).

Como la función de Epanechnikov se encuentra limitada al intervalo $[-1, 1]$, la condición c) también se cumple. Por último, podemos observar que:

$$K_E(-x) = \frac{3}{4} (1 - x^2) I_{[-1,1]}(-x) = \frac{3}{4} (1 - x^2) I_{[-1,1]}(x) = K_E(x)$$

por lo que se cumple también la condición f). □

2.4.4. Funciones kernel más usuales

Hasta ahora hemos realizado un estudio general sobre las funciones kernel, pero es conveniente conocer qué modelos concretos se suelen utilizar en los paquetes estadísticos y cuáles son las respectivas ventajas de cada uno.

Además de la función uniforme, que constituye la base del estimador naïve, también podemos utilizar funciones kernel como la normal, la triangular, Epanechnikov, biweight y triweight, entre otras. En la [Tabla 2.1](#) mostramos un resumen de las funciones anteriormente mencionadas mientras que en la [Figura 2.3](#) mostramos una representación gráfica de las mismas. Por su parte, la [Tabla 2.2](#) muestra las principales propiedades de los kernel listados así como de los estimadores asociados.

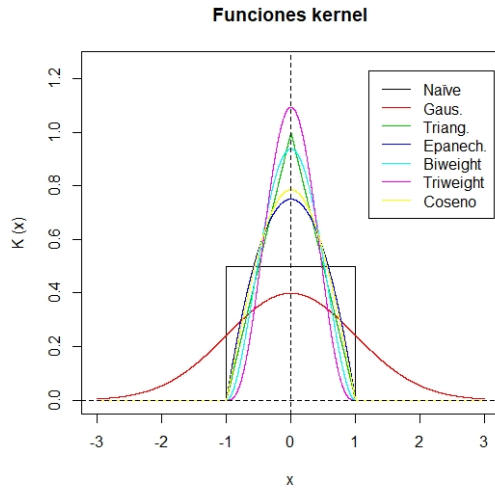


Figura 2.3: Gráfica de las funciones kernel presentadas en la [Tabla 2.2](#).

Tabla 2.1: Funciones kernel más usuales.

Nombre	Forma $K(x)$
Naïve	$\frac{1}{2} I_{[-1,1]}(x)$
Normal	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$
Triangular	$(1 - x) I_{[-1,1]}(x)$
Epanechnikov	$\frac{3}{4} (1 - x^2) I_{[-1,1]}(x)$
Biweight	$\frac{15}{16} (1 - x^2)^2 I_{[-1,1]}(x)$
Triweight	$\frac{35}{32} (1 - x^2)^3 I_{[-1,1]}(x)$
Coseno	$\frac{\pi}{4} \cos\left(\frac{\pi}{2}x\right) I_{[-1,1]}(x)$

Tabla 2.2: Principales propiedades de las funciones kernel más usuales, V_K y $\int_{-\infty}^{\infty} K^2(x) dx$ y propiedades de los estimadores asociados.

Nombre	V_K	$\int_{-\infty}^{\infty} K^2(x) dx$	Sesgo	Varianza	MISE
Naïve	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{h^2}{6} f''(x)$	$\frac{f(x)}{2mh}$	$\frac{1}{2mh} + \frac{h^4}{36} R(f)$
Normal	1	$\frac{1}{2\sqrt{\pi}}$	$\frac{h^2}{2} f''(x)$	$\frac{1}{2mh\sqrt{\pi}} f(x)$	$\frac{1}{2mh\sqrt{\pi}} + \frac{h^4}{4} R(f)$
Triangular	$\frac{1}{6}$	$\frac{2}{3}$	$\frac{h^2}{12} f''(x)$	$\frac{2}{3mh} f(x)$	$\frac{2}{3mh} + \frac{h^4}{144} R(f)$
Epanechnikov	$\frac{1}{5}$	$\frac{3}{5}$	$\frac{h^2}{10} f''(x)$	$\frac{3}{5mh} f(x)$	$\frac{3}{5mh} + \frac{h^4}{100} R(f)$
Biweight	$\frac{1}{7}$	$\frac{5}{7}$	$\frac{h^2}{14} f''(x)$	$\frac{5}{7mh} f(x)$	$\frac{5}{7mh} + \frac{h^4}{196} R(f)$
Triweight	$\frac{1}{9}$	$\frac{350}{429}$	$\frac{h^2}{18} f''(x)$	$\frac{350}{429mh} f(x)$	$\frac{350}{429mh} + \frac{h^4}{324} R(f)$
Coseno	$1 - \frac{8}{\pi^2}$	$\frac{\pi^2}{16}$	$\frac{h^2}{2} \left(1 - \frac{8}{\pi^2}\right) f''(x)$	$\frac{\pi^2}{16mh} f(x)$	$\frac{\pi^2}{16mh} + \frac{h^4}{4} \left(1 - \frac{8}{\pi^2}\right)^2 R(f)$

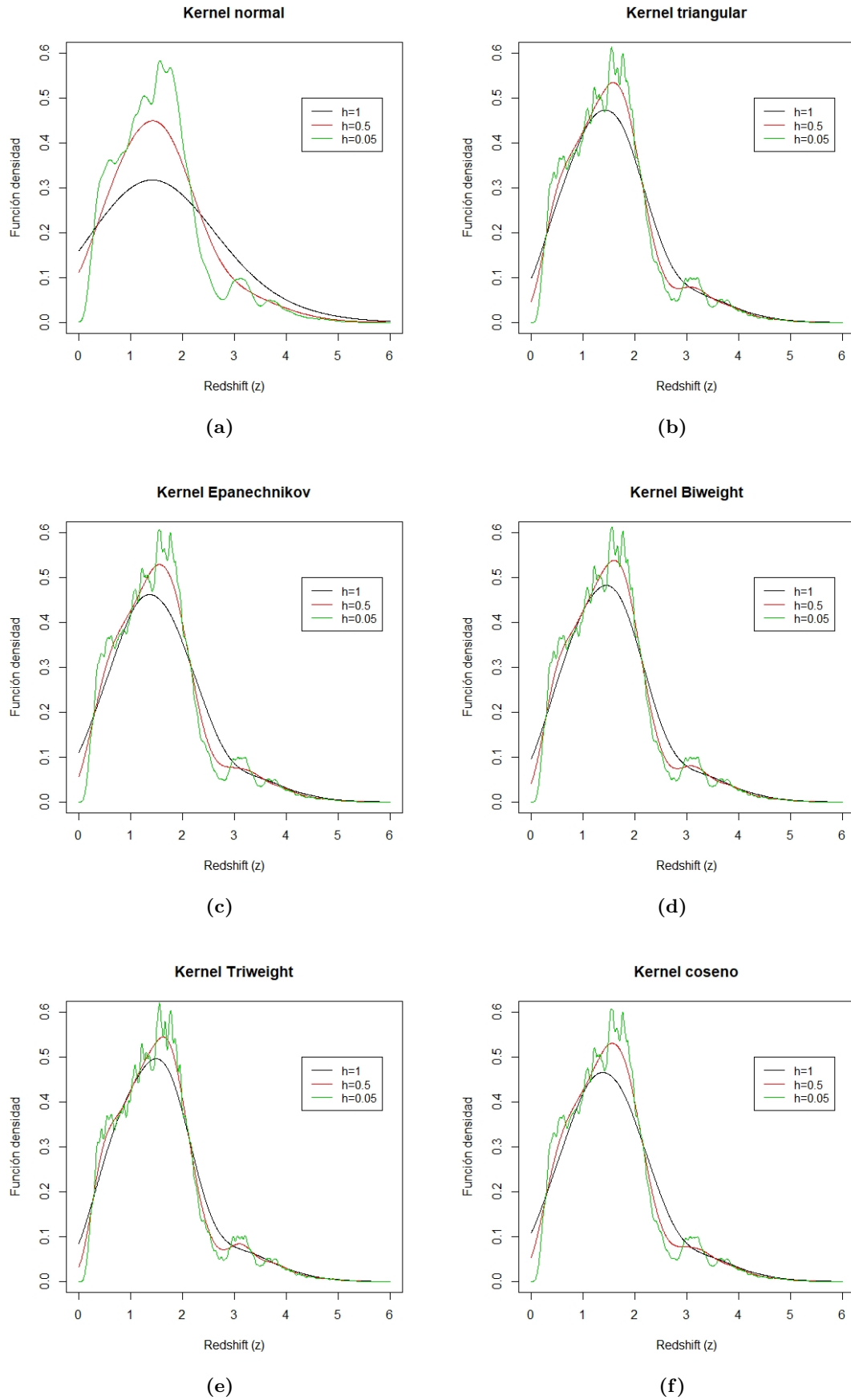


Figura 2.4: Estimaciones realizadas con las funciones kernel de la [Tabla 2.2](#).

La deducción de la función Epanechnikov que presentamos en la [Subsección 2.4.3](#) muestra que la optimización del MISE depende exclusivamente del término $\int_{-\infty}^{\infty} K^2(y) dy$. Por lo tanto, es intuitivo introducir el concepto de la *eficiencia* de la función kernel $\text{eff}(K)$, de la misma manera que hace Silverman[17], es decir, comparando el MISE asintótico de cada kernel con el presentado por la función Epanechnikov:

Definición 2.20. Dada una función kernel K , se define la eficiencia de la función como el cociente:

$$\text{eff}(K) = \frac{\left[\int_{-\infty}^{\infty} K_E^2(y) dy \right] \left[\int_{-\infty}^{\infty} y^2 K_E(y) dy \right]^{\frac{1}{2}}}{\left[\int_{-\infty}^{\infty} K^2(y) dy \right] \left[\int_{-\infty}^{\infty} y^2 K(y) dy \right]^{\frac{1}{2}}} \quad (2.19)$$

Corolario 2.21. La eficiencia para cualquier función kernel K se puede expresar como:

$$\text{eff}(K) = \frac{3}{5\sqrt{5}} \left[\int_{-\infty}^{\infty} K^2(y) dy \right]^{-1} \left[\int_{-\infty}^{\infty} y^2 K(y) dy \right]^{-\frac{1}{2}} \quad (2.20)$$

Demostración. De acuerdo a los resultados de la [Tabla 2.2](#), para la función Epanechnikov tenemos:

$$\int_{-\infty}^{\infty} K_E^2(y) dy = \frac{3}{5}$$

$$\int_{-\infty}^{\infty} y^2 K_E(y) dy = V_K = \frac{1}{5}$$

Introduciendo ambos resultados en (2.19) se obtiene la expresión (2.20). □

En (2.19) introducimos la potencia $\frac{5}{4}$ [17] ya que nos permite trabajar de una forma más sencilla con la expresión. Además, de acuerdo al Corolario 2.17, al introducir dicha potencia en el MISE tendremos un comportamiento asintótico dado por $\tilde{B}m^{-1}$, lo que implica que el MISE obtenido para una función kernel K con un tamaño de muestra m , es el mismo que el obtenido al hacer uso de la función de Epanechnikov pero con un tamaño de muestra menor, $m \cdot \text{eff}(K)$.

Tabla 2.3: Eficiencias de las diferentes funciones kernel presentadas en orden decreciente.

Función	Eficiencia exacta	Valor aproximado
Epanechnikov	1	1.0000
Coseno	$\frac{48\sqrt{\frac{\pi^2}{\pi^2-8}}}{5\pi^2\sqrt{5}}$	0.9995
Biweight	$\frac{21\sqrt{7}}{25\sqrt{5}}$	0.9939
Triweight	$\frac{3861}{1750\sqrt{5}}$	0.9867
Triangular	$\frac{9\sqrt{6}}{10\sqrt{5}}$	0.9859
Normal	$\frac{6\sqrt{\pi}}{15\sqrt{5}}$	0.9512
Naïve	$\frac{6\sqrt{3}}{5\sqrt{5}}$	0.9295

Las eficiencias que se presentan en la [Tabla 2.3](#) son bastante grandes, incluso para la estimación naïve que se correspondería con la aproximación menos suave.

Visualicemos, mediante el ejemplo de la distribución de cuásares medidos en la SDSS, que resultados ofrecen cada uno de ellos. En la [Figura 2.4](#) se muestran las estimaciones realizadas. Al haber prefijado las mismas ventanas para todos ellos, aparentemente el kernel normal (a) nos ofrece una mayor suavidad y un buen ajuste a la función de densidad. Sin embargo, la aproximación teórica del intervalo óptimo [\(2.16\)](#) muestra que el intervalo óptimo cambia para cada tipo de función kernel, motivo por el cual la anchura de ventana que funciona bien con el kernel normal (a) no ofrece tan buenos resultados con otros kernel como el triweight (e) o el coseno (f).

2.4.5. Generalización a funciones multidimensionales

En todo lo discutido hasta ahora nos hemos centrado en el desarrollo de la estimación de densidad de una variable aleatoria X unidimensional. Sin embargo, también es importante analizar la situación cuando la función de densidad pasa a depender de una variable aleatoria n -dimensional \vec{x} . Supongamos que tenemos una muestra aleatoria simple de tamaño m para la variable multidimensional $(\vec{X}_1, \vec{X}_2, \dots, \vec{X}_m)$. En el caso más sencillo de esta estimación, podemos plantear la función kernel $K(\vec{x})$ como un producto de funciones kernel unidimensionales $K(\vec{x}) = \prod_{j=1}^n K(x^j)$.

Definición 2.22. Dada una muestra aleatoria simple $(\vec{X}_1, \vec{X}_2, \dots, \vec{X}_m)$ donde cada \vec{X}_i es un vector de tamaño n , se define un estimador kernel multidimensional como:

$$\hat{f}_K(\vec{x}) = \frac{1}{m \prod_{j=1}^n h^j} \sum_{i=1}^m \prod_{j=1}^n K\left(\frac{x^j - X_i^j}{h^j}\right) \quad (2.21)$$

La idea subyacente en este estimador es extender las ventanas unidimensionales de los kernel a n -cubos en el espacio \mathbb{R}^n . De esta manera, se pueden aprovechar todos los resultados derivados del cálculo unidimensional. No obstante, el estimador [\(2.21\)](#) es un caso particular de situaciones más generales en las que se pueden considerar matrices H de tamaño $n \times n$ donde se contemplan amplitudes de ventanas h^{j_1, j_2} .

Definición 2.23. Dada una muestra aleatoria simple $(\vec{X}_1, \vec{X}_2, \dots, \vec{X}_m)$ donde cada \vec{X}_i es un vector de tamaño n , se define un estimador kernel multidimensional como:

$$\hat{f}_K(x^1, x^2, \dots, x^n) = \frac{1}{m \det(H)} \sum_{i=1}^m \mathcal{K}\left(H^{-1}(\vec{x} - \vec{X}_i)\right) \quad (2.22)$$

donde H es una matriz de tamaño $n \times n$.

Al suponer la matriz H diagonal en [\(2.22\)](#), se recupera el caso anterior[8]. Es importante resaltar que incluso en la versión más simplificada, el cálculo computacional necesario para realizar la estimación puede volverse muy pesado, además de que para lograr un valor del MISE similar al caso unidimensional es necesario un mayor tamaño de muestra[5].

3. Selección de ventana óptima

3.1. Introducción

En el anterior capítulo comenzamos analizando el estimador histograma para la función densidad y posteriormente mostramos como los estimadores kernel permitían corregir el problema de la arbitrariedad de la elección del punto inicial de las clases presente en el primero. Las funciones kernel añadían suavidad a la estimación de la función de densidad, pero un aspecto esencial de los mismos es la elección de la amplitud de ventana h , cuyo papel también es importante para la suavidad de las curvas estimadas. Es decir, en la suavización de la estimación de densidad por funciones kernel tenemos dos parámetros de elección: la función K y la anchura de ventana h .

En este capítulo abordaremos el problema de la elección de la amplitud de ventana asumiendo que este parámetro no tiene dependencia alguna con la posición x en la que se estima la función de densidad, es decir, se mantiene fijo para cada estimación. Es por este motivo por el que los métodos que vamos a estudiar a continuación reciben el nombre de *métodos de anchura fija*.

3.2. Aproximación práctica a la ventana óptima

El estudio que realizamos en la [Subsección 2.4.2](#) nos permitió obtener una expresión para la ventana óptima \tilde{h} (2.16) que presentaba un gran inconveniente: depende del valor real de la función de densidad $f(x)$ que desconocemos. Por este motivo, desde la introducción de los estimadores kernel en 1956 se han desarrollado numerosas técnicas para elegir un valor h que se aproxime al valor teórico \tilde{h} .

3.2.1. Elección subjetiva

El primer método que estudiamos es el menos exacto pero el más sencillo: establecer nuestro propio criterio para la elección de h . Podemos resumir este criterio en elegir una serie de valores para la amplitud de ventana $\{h_1, h_2, \dots, h_k\}$, estimar la función de densidad para cada uno de ellos y elegir aquel parámetro que, bajo nuestro punto de vista (de ahí el calificativo de *subjetivo*) mejor estime la estructura subyacente de los datos y no pierda excesiva información. Al ser completamente subjetivo, la elección de una persona puede presentar mayor sesgo y menor varianza que la que realice otra.

Por ejemplo, dada la [Figura 3.1](#), una persona puede tomar la elección de $h = 0.3$ en base a una menor varianza y un buen sesgo, pero otra puede considerar que $h = 0.5$ se ajusta mejor (suponiendo que tiene un conocimiento de la población global).

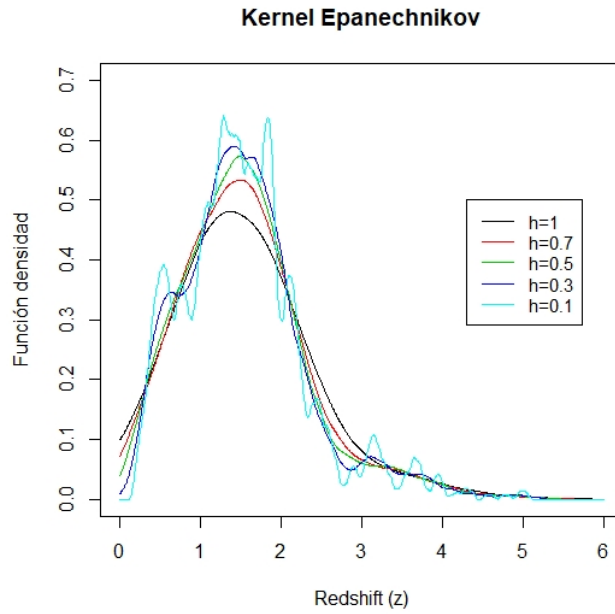


Figura 3.1: Estimación con kernel Epanechnikov para 1000 cuásares de la muestra de la campaña SDSS que estamos analizando.

3.2.2. Elección en el caso normal

Otra forma sencilla de aproximar la anchura óptima de ventana es fijar una función de densidad conocida como la función f real que desconocemos. Lo más frecuente es utilizar como función de densidad la de la distribución normal por sus importantes aplicaciones.

Proposición 3.1. Sea X una variable aleatoria que sigue una distribución normal $N(\mu, \sigma)$, entonces se verifica que:

$$R(f) = \frac{3}{8\sqrt{\pi}\sigma^5}$$

Demostración. Por hipótesis, la función de densidad real de la muestra viene dada por:

$$f_N(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Calculamos sus derivadas:

$$\begin{aligned} \frac{d}{dx} f_N(x) &= \frac{1}{\sqrt{2\pi}\sigma^2} \left(-\frac{x-\mu}{\sigma^2}\right) \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \Rightarrow \\ \Rightarrow \frac{d^2}{dx^2} f_N(x) &= \frac{1}{\sqrt{2\pi}\sigma^6} \left[\frac{(x-\mu)^2}{\sigma^2} - 1\right] \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \end{aligned}$$

Introduciendo entonces la expresión de $f_N''(x)$ en $R(f)$, obtenemos:

$$\int_{-\infty}^{\infty} (f_N''(y))^2 dy = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^6} \exp\left(-\frac{(y-\mu)^2}{\sigma^2}\right) \left[\frac{(y-\mu)^2}{\sigma^2} - 1\right]^2 dy \stackrel{(y-\mu)\sigma^{-1}=z}{=}$$

$$= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^5} \exp(-z^2) [z^2 - 1]^2 dz = \frac{3}{8\sqrt{\pi}\sigma^5}$$

□

Corolario 3.2. Sea X una variable aleatoria que sigue una distribución normal $N(\mu, \sigma)$, una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador asociado a una función kernel K . La amplitud de ventana que minimiza el MISE asintótico viene dada por:

$$\tilde{h} = \left[\frac{8\sqrt{\pi} \int_{-\infty}^{\infty} K^2(y) dy}{3V_K^2} \right]^{\frac{1}{5}} \sigma m^{-\frac{1}{5}} \quad (3.1)$$

Demostración. Las hipótesis nos permiten recuperar el resultado (2.16). Por otro lado, la función de densidad real es la normal, con lo que podemos emplear la Proposición 3.1, de lo que se deduce la expresión. □

Teniendo en cuenta este valor, podemos entonces calcular los valores de la anchura óptima (2.16) para las diferentes funciones kernel de la Tabla 2.2. Dichos resultados se muestran en la Tabla 3.4.

Tabla 3.4: Anchuras óptimas por suposición de normalidad en la función de densidad para las funciones kernel.

Función	Intervalo óptimo \tilde{h}	Valor aproximado
Epanechnikov	$(40\sqrt{\pi})^{\frac{1}{5}} m^{-\frac{1}{5}} \sigma$	$2.3449m^{-\frac{1}{5}} \sigma$
Coseno	$\left(\frac{\pi^{\frac{3}{2}}}{6[1-\frac{8}{\pi^2}]^2} \right)^{\frac{1}{5}} m^{-\frac{1}{5}} \sigma$	$2.4097m^{-\frac{1}{5}} \sigma$
Biweight	$\left(\frac{280}{3} \sqrt{\pi} \right)^{\frac{1}{5}} m^{-\frac{1}{5}} \sigma$	$2.7779m^{-\frac{1}{5}} \sigma$
Triweight	$\left(\frac{25200}{143} \sqrt{\pi} \right)^{\frac{1}{5}} m^{-\frac{1}{5}} \sigma$	$3.1545m^{-\frac{1}{5}} \sigma$
Triangular	$(64\sqrt{\pi})^{\frac{1}{5}} m^{-\frac{1}{5}} \sigma$	$2.5760m^{-\frac{1}{5}} \sigma$
Normal	$\left(\frac{4}{3} \right)^{\frac{1}{5}} m^{-\frac{1}{5}} \sigma$	$1.0592m^{-\frac{1}{5}} \sigma$
Rectangular	$(12\sqrt{\pi})^{\frac{1}{5}} m^{-\frac{1}{5}} \sigma$	$1.8431m^{-\frac{1}{5}} \sigma$

En la práctica, el valor de σ se tiene que estimar. Una forma habitual de estimar dicho valor es mediante la desviación típica de los datos muestrales, definida como:

$$\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^m (X_i - \bar{X})^2}{m-1}} \quad (3.2)$$

Sin embargo, otros autores argumentan que para estimar σ es mejor el empleo de la estimación del rango intercuartil $\hat{R}[20]$, definido como la diferencia entre el tercer cuartil y el primer cuartil $IQR = Q_3 - Q_1$.

Definición 3.3. Dada una variable aleatoria X con función de distribución $F(x)$, se definen los cuantiles

de la distribución como:

$$Q_j = \inf [x : F(x) \geq 0.25j] \quad \text{con } j \in \{1, 2, 3\} \quad (3.3)$$

Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) de dicha variable aleatoria, los cuartiles se estiman por medio de la función de distribución empírica como:

$$\hat{Q}_j = \inf [x : \hat{F}(x) \geq 0.25j] \quad (3.4)$$

Definición 3.4. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) , se define un estimador del rango intercuartil como:

$$\widehat{\text{IQR}} = \hat{Q}_3 - \hat{Q}_1 = \inf [x : \hat{F}(x) \geq 0.75] - \inf [x : \hat{F}(x) \geq 0.25] \quad (3.5)$$

Proposición 3.5. Sea X una variable aleatoria que sigue una distribución normal $N(\mu, \sigma)$. La relación entre el rango intercuartil IQR y la desviación típica σ viene dada por:

$$\text{IQR} = 1.3489796\sigma \quad (3.6)$$

Demostración. En primer lugar, podemos observar que los cuartiles de cualquier distribución del estilo $N(\mu, \sigma)$ se pueden calcular en función de los cuartiles de la normal tipificada $N(0, 1)$. Para demostrarlo basta con calcular:

$$Q_j = \inf \left[x : \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{(y-\mu)^2}{2\sigma^2} \right) dy \geq j0.25 \right] \stackrel{z=\sigma^{-1}(y-\mu)}{=} \inf \left[x : \int_{-\infty}^{\frac{x-\mu}{\sigma}} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{z^2}{2} \right) dz \geq j0.25 \right]$$

De acuerdo a este resultado, el cuartil de $N(\mu, \sigma)$ será

$$Q_j = \sigma Q_j^N + \mu$$

con Q_j^N el cuartil de la distribución normal tipifica. Para esta última, sabemos que $Q_1^N = -0.6744898$ y $Q_3^N = 0.6744898$, con lo que concluimos:

$$\text{IQR} = \sigma (0.6744898) + \mu - \sigma (-0.6744898) - \mu = 1.3489796\sigma$$

□

La Proposición 3.5 muestra que existe una relación, en el caso normal, entre el rango intercuartil y la desviación típica, pero la información que nos suministran es diferente: el rango intercuartil mide la dispersión en la zona central de distribución mientras que la desviación típica mide la dispersión global. Podemos usar dicha diferencia en los estimadores de σ : el estimador (3.2) se verá más afectado por valores muestrales situados en los extremos, problema que se soluciona al contemplar el estimador (3.5). Por todo ello, muchos autores defienden utilizar como estimador de σ el valor $\min \left\{ \hat{\sigma}, \frac{\widehat{\text{IQR}}}{1.349} \right\}$ [20].

En la [Figura 3.2](#) se muestra el ejemplo de uso de la amplitud de ventana óptima de la [Tabla 3.4](#) para nuestra muestra de cuásares. Analizando dicha figura se observa que para la ventana óptima $\tilde{h} = 0.3815$ se obtiene una mayor varianza que para $h = 0.3$. Este resultado se debe a que la variable (el redshift)

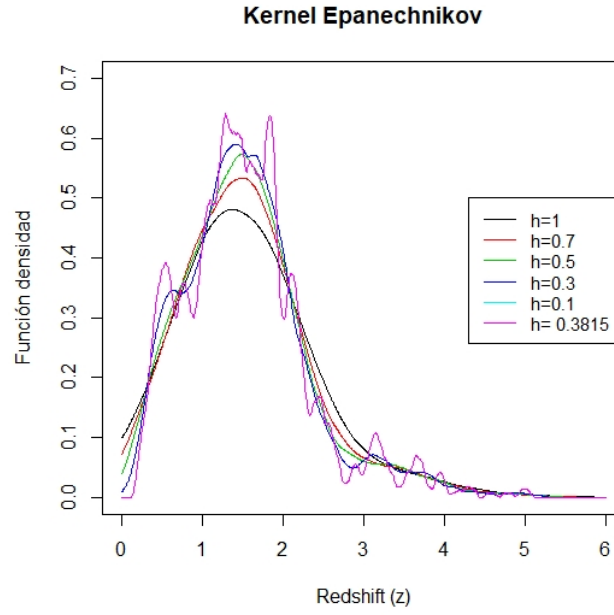


Figura 3.2: Estimación con kernel Epanechnikov para 1000 cuásares de la muestra de la campaña SDSS. La ventana óptima en el caso normal se muestra en color morado

no tiene por qué seguir una distribución normal (como es en este caso), por lo que la ventana óptima propuesta no se corresponde con la ventana óptima efectiva, aunque nos proporciona una aproximación a la misma.

3.2.3. Elección por métodos *plug-in*

Por métodos *plug-in* entendemos una serie de procedimientos encaminados a estimar el factor desconocido de la expresión de la ventana óptima (2.16), es decir, estimar el término $R(f)$.

Método 1. El primer modelo fue propuesto por Scott, Tapia y Thompson en 1977[19], y se basa en un proceso iterativo. Comenzamos introduciendo un valor arbitrario para nuestra anchura de ventana, por ejemplo h_1 , bajo el cual realizamos nuestra estimación de la función densidad que denotamos $\hat{f}_{K,1}(x)$. Utilizamos dicha estimación para calcular $R(f)$ como $R(\hat{f}_{K,1})$, que introducimos en la expresión (2.16), para obtener un nuevo valor h_2 . Realizamos el proceso nuevamente hasta obtener un valor h_{n+1} que verifique $|h_{n+1} - h_n| \leq S$, donde S es un valor que se prefija como precisión. En la iteración n -ésima tendremos:

$$\hat{R}_n = R(\hat{f}_{K,n}) = \frac{1}{m^2 h_n^6} \sum_{i=1}^m \sum_{j=1}^m \int_{-\infty}^{\infty} K''\left(\frac{x - X_i}{h_n}\right) K''\left(\frac{x - X_j}{h_n}\right) dx \quad (3.7)$$

que introducimos en (2.16) para obtener:

$$h_{n+1} = \left[\frac{\int_{-\infty}^{\infty} K^2(y) dy}{\left(\int_{-\infty}^{\infty} y^2 K(y) dy \right)^2 \left(\sum_{i=1}^m \sum_{j=1}^m \int_{-\infty}^{\infty} K''\left(\frac{x - X_i}{h_n}\right) K''\left(\frac{x - X_j}{h_n}\right) dx \right)} \right]^{\frac{1}{5}} m^{\frac{1}{5}} h_n^{\frac{6}{5}} \quad (3.8)$$

De la expresión (3.8) observamos que el trabajo computacional aumenta claramente con el tamaño de la muestra m como consecuencia del doble sumatorio presente. Por tanto, para tamaños de muestra muy grandes no debemos exigir un nivel de precisión muy alto (es decir, valores de S muy pequeños).

Ejemplo 3.6. Volvamos nuestro estudio sobre la distribución cuásares en función del redshift. Como es necesario calcular una derivada de orden 2 del kernel, ya no utilizaremos la función Epanechnikov. De acuerdo a los datos de la Tabla 2.3, utilizaremos la función Triweight, la función de mayor eficiencia para la que podemos determinar las derivadas necesarias. Calculamos entonces:

$$K(x) = \frac{35}{32} (1 - x^2)^3 I_{[-1,1]}(x) \Rightarrow K''(x) = \frac{105}{16} (1 - x^2) (5x^2 - 1) I_{[-1,1]}(x)$$

Y por otro lado:

$$\int_{-\infty}^{\infty} K''\left(\frac{x - X_i}{h}\right) K''\left(\frac{x - X_j}{h}\right) dx \underset{x - X_i = zh}{=} h \int_{-\infty}^{\infty} K''(z) K''\left(z - \frac{X_j - X_i}{h}\right) dz$$

Para simplificar los cálculos denotaremos la cantidad $\frac{X_j - X_i}{h_n} \equiv t_{ij}$, de esta manera, el cálculo de \hat{R}_n se puede expresar como:

$$\begin{aligned} \hat{R}_n &= \frac{1}{m^2 h_n^5} \sum_{i=1}^m \sum_{j=1}^m \int_{-\infty}^{\infty} K''(z) K''(z - t_{ij}) dz = \\ &= \frac{1}{m^2 h_n^5} \left(\frac{105}{16}\right)^2 \sum_{i,j=1}^m \int_{-\infty}^{\infty} (1 - z^2) (1 - (z - t_{ij})^2) (5z^2 - 1) (5(z - t_{ij})^2) I_{[-1,1]}(z) I_{[-1,1]}(z - t_{ij}) dz \end{aligned}$$

El producto de las funciones indicador se puede expresar como:

$$I_{[-1,1]}(z) I_{[-1,1]}(z - t_{ij}) = I_{[-1,1]}(z) I_{[-1+t_{ij}, 1+t_{ij}]}(z)$$

Como ambas funciones indicador están asociados a intervalos de la misma longitud ($\ell = 2$), estudiando las posibles intersecciones de los conjuntos, dicho producto toma la forma:

$$I_{[-1,1]}(z) I_{[-1+t_{ij}, 1+t_{ij}]}(z) = I_{[-1+t_{ij}, 1]}(z) I_{[0,2]}(t_{ij}) + I_{[-1, 1+t_{ij}]}(z) I_{[-2,0]}(t_{ij}) - I_{[-1,1]}(z) I_{\{0\}}(t_{ij})$$

Es necesario añadir el último término para el caso en el que $i = j$, ya que la situación $i = j$ ($t_{ij} = 0$) lo estamos contemplando dos veces. De hecho, dicho valor no dependerá de h_n , motivo por el cual en posteriores revisiones del método optaremos por omitir este término de la estimación. Retomando la integración, para cada uno de los términos en los que descomponemos el producto de intervalos obtenemos:

$$\begin{aligned} \hat{R}_n &= \frac{1}{m^2 h_n^5} \left(\frac{105}{16}\right)^2 \sum_{i,j=1}^m \left[\frac{1}{630} (-25t_{ij}^9 + 360t_{ij}^7 - 2016t_{ij}^5 + 6720t_{ij}^3 - 5760t_{ij}^2 + 512) I_{[0,2]}(t_{ij}) + \right. \\ &\quad \left. + \frac{1}{630} (25t_{ij}^9 - 360t_{ij}^7 + 2016t_{ij}^5 - 6720t_{ij}^3 - 5760t_{ij}^2 + 512) I_{[-2,0]}(t_{ij}) - \frac{64}{315} (4 - 45t_{ij}^2) I_{\{0\}}(t_{ij}) \right] \end{aligned}$$

Los resultados del método se muestran en la Figura 3.3. Fijando una precisión $S = 10^{-1}$ se obtiene una anchura de ventana $h = 0.4 \pm 0.1$ mientras que para una precisión $S = 10^{-5}$ se obtiene $h = 0.23572 \pm$

0.00001. Estos valores los podemos comparar con la anchura de ventana referida a la distribución normal (Subsección 3.2.2), que establece $h = 0.640746$. Observamos entonces que para una precisión relativamente pequeña, el método refina considerablemente la estimación referida a una distribución normal.

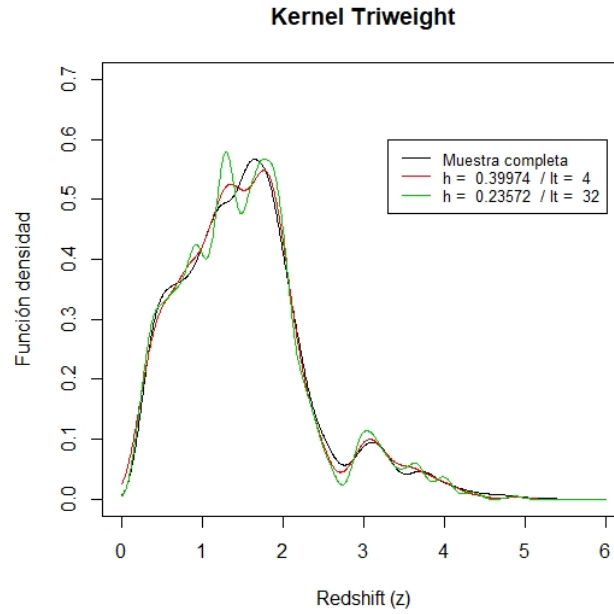
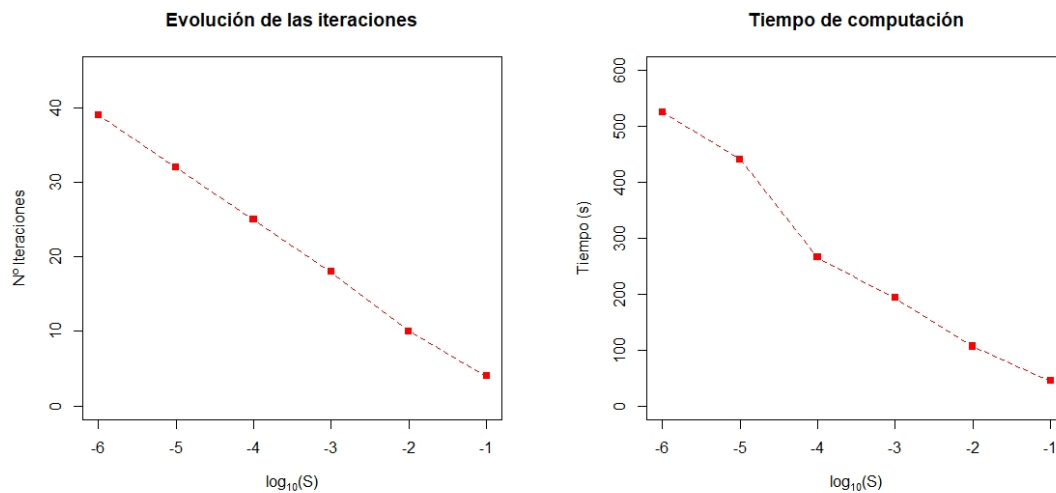


Figura 3.3: Ejemplo de la anchura de ventana obtenida por el método plug-in básico para una muestra aleatoria de 1000 cuásares. La función de densidad para la muestra completa se ha determinado con la anchura de ventana referida a la distribución normal. La curva en color rojo se corresponde con la anchura de ventana para una precisión $S = 10^{-1}$, mientras que la de color verde para una precisión $S = 10^{-5}$.



(a) Evolución de las iteraciones necesarias para la convergencia de la anchura de ventana. (b) Tiempo computacional empleado para la obtención de la ventana óptima.

Figura 3.4: Computación de la anchura de ventana óptima para una anchura inicial de $h = 5$. El parámetro S denota la precisión bajo la que se quiere la convergencia de la anchura.

Uno de los principales inconvenientes de este método, y de todos los refinamientos posteriores que explicaremos, es el trabajo computacional que requieren. La [Figura 3.4](#) (a) muestra que el número de iteraciones para una precisión $S = 10^{-1}$ es de 4, mientras que para una precisión $S = 10^{-6}$ es de 39. Sin embargo, aunque son pocas iteraciones, en cada una de ellas se realizan una gran cantidad cálculos porque hay que recalcular el valor \hat{R}_n . De esta manera, la [Figura 3.4](#) (b) muestra que para la realización de esas 4 iteraciones invertimos un tiempo de 45s, mientras que en la realización de 39 iteraciones, el tiempo invertido es de 525s.

Método 2. Un procedimiento alternativo fue propuesto por Hall y Marron en 1987[6], que utiliza la convolución de funciones para estimar el funcional $R(f)$.

Definición 3.7. Dadas dos funciones f y g integrables, se define la convolución de f con g , denotada por $f * g(t)$ como:

$$f * g(t) = \int_{-\infty}^{\infty} f(x) g(x-t) dx$$

Iniciamos la búsqueda de la ventana óptima con una estimación de la función de densidad con la anchura de intervalo genérica h , de manera que emplearíamos $\hat{f}_{K,h}$ para estimar el término $R(f)$.

Partiendo de la expresión (3.7), realizando el cambio de variable $x - X_i = hy$, podemos expresar la estimación $R(\hat{f}_{K,h})$ en términos de convolución de funciones como:

$$R(\hat{f}_{K,h}) = \frac{1}{m^2 h^5} \sum_{i=1}^m \sum_{j=1}^m K'' * K'' \left(\frac{X_j - X_i}{h} \right)$$

como los índices i y j coinciden en un total de m -veces, entonces podemos realizar el desarrollo:

$$R(\hat{f}_{K,h}) = \frac{1}{mh^5} K'' * K''(0) + \frac{1}{m^2 h^5} \sum_{i \neq j} K'' * K'' \left(\frac{X_j - X_i}{h} \right) \quad (3.9)$$

La expresión (3.9) muestra que el primer sumando no depende en absoluto de la muestra, por lo que la estimación \hat{R} se puede mejorar, simplemente omitiendo dicho término, que únicamente contribuye en su sesgo.

Hall y Marron proponen entonces como estimador:

$$\hat{R}(f) := R(\hat{f}_{K,h}) - \frac{1}{mh^5} R(K) \quad (3.10)$$

Sin embargo, la anchura de ventana \tilde{h} que optimiza el estimador $\hat{R}(f)$ es distinta de la anchura óptima \tilde{h} que se deduce en (2.16). Podemos expresar esta discrepancia en términos del cociente entre el valor de h que minimiza el MISE (de aquí en adelante denotado por $\tilde{h} \equiv h_{MISE}$) y el valor de h que se obtiene como óptimo para estimar $\hat{R}(f)$ (de aquí en adelante denotado por h_{PI}), definido como:

$$\frac{h_{PI} - h_{MISE}}{h_{MISE}} \quad (3.11)$$

En 1990, Park y Marron[12] propusieron utilizar como amplitud de ventana en (3.10) el valor h_{PI} que

optimiza el ratio que acabamos de definir. Sin embargo, este procedimiento les llevó a una relación entre ambas anchuras en la que vuelve a aparecer un término dependiente de la función de densidad real f , de manera que nos veríamos obligados a proceder como en la **Subsección 3.2.2**: aproximar la función de densidad por un modelo conocido (por ejemplo, la distribución normal); estimar los parámetros del modelo (en el caso normal, μ y σ); introducir todo ello en la expresión y obtener así el valor final de h .

Otro procedimiento distinto a estos dos anteriores es el presentado por Hall, Sheather, Jones y Marron en 1991[7] que, utilizando funciones kernel con restricciones adicionales a las cinco consideradas originalmente, consiguen un mayor orden de convergencia del sesgo, y por tanto, del MISE.

Proposición 3.8. Sean X una variable aleatoria con función de densidad $f \in \mathcal{C}^4$ y con $f^{(n)}(\infty) = f^{(n)}(-\infty) = 0 \forall n \in \mathbb{N}$, una muestra aleatoria simple suya (X_1, X_2, \dots, X_m) , una función kernel K que verifica las condiciones:

$$\begin{aligned} g) \quad & \int_{-\infty}^{\infty} y^3 K(y) dy = 0 \\ h) \quad & \int_{-\infty}^{\infty} y^4 K(y) dy \equiv V_{K,4} < \infty \end{aligned}$$

y \hat{f}_K el estimador asociado. Se verifica que el MISE toma la expresión:

$$\text{MISE}(\hat{f}_K) = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(y) dy + \frac{h^4}{4} V_K^2 R(f) - \frac{h^6}{24} V_K V_{K,4} R_3(f) + o\left(\frac{1}{mh} + h^6\right) \quad (3.12)$$

donde $R_3(f) := \int_{-\infty}^{\infty} (f'''(x))^2 dx$.

Demostración. Calculamos la esperanza del estimador, que viene dada por:

$$E(\hat{f}_K(x)) = \frac{1}{mh} \sum_{i=1}^m E\left(K\left(\frac{x - X_i}{h}\right)\right) = \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{x - y}{h}\right) f(y) dy \underset{x=y+\tilde{y}h}{=} \int_{-\infty}^{\infty} K(\tilde{y}) f(x - h\tilde{y}) d\tilde{y}$$

como por hipótesis $f \in \mathcal{C}^4$, realizamos un desarrollo de Taylor cuando $h \rightarrow 0$ hasta orden 4:

$$E(\hat{f}_K(x)) = \int_{-\infty}^{\infty} K(\tilde{y}) \left[f(x) - h\tilde{y}f'(x) + \frac{h^2\tilde{y}^2}{2}f''(x) - \frac{h^3\tilde{y}^3}{6}f'''(x) + \frac{h^4\tilde{y}^4}{24}f^{iv}(x) + o(h^4) \right] d\tilde{y}$$

y haciendo uso de las condiciones a), d) y e) de las funciones kernel, así como de las condiciones g) y h) de las hipótesis, obtenemos finalmente:

$$E(\hat{f}_K(x)) = f(x) + \frac{h^2}{2} f''(x) V_K + \frac{h^4}{24} f^{iv}(x) V_{K,4} + o(h^4)$$

Por lo tanto, el sesgo del estimador vendrá dado por:

$$\text{sesgo}(\hat{f}_K(x)) = \frac{h^2}{2} f''(x) V_K + \frac{h^4}{24} f^{iv}(x) V_{K,4} + o(h^4) \quad (3.13)$$

El cálculo de la varianza es el mismo que el presentado en la Proposición 2.14, por lo que el MSE del

estimador vendrá dado por:

$$\text{MSE}(\hat{f}_K(x)) = \frac{1}{mh} f(x) \int_{-\infty}^{\infty} K^2(y) dy + \frac{h^4}{4} (f''(x))^2 V_K^2 + \frac{h^6}{24} f''(x) f^{iv}(x) V_K V_{K,4} + o\left(h^6 + \frac{1}{mh}\right) \quad (3.14)$$

e integrando dicha expresión, obtenemos el MISE como:

$$\text{MISE} = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(y) dy + \frac{h^4}{4} V_K^2 \int_{-\infty}^{\infty} (f''(x))^2 dx + \frac{h^6}{24} V_K V_{K,4} \int_{-\infty}^{\infty} f'''(x) f^{iv}(x) dx + o\left(\frac{1}{mh} + h^6\right)$$

Por hipótesis, $f^{(n)}(\infty) = f^{(n)}(-\infty) = 0$, con lo que podemos observar entonces que:

$$\int_{-\infty}^{\infty} f'''(x) f^{iv}(x) dx = f''(x) f'''(x)|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} (f'''(x))^2 dx = - \int_{-\infty}^{\infty} (f'''(x))^2 dx$$

con lo que concluimos que:

$$\begin{aligned} \text{MISE} &= \frac{1}{mh} \int_{-\infty}^{\infty} K^2(y) dy + \frac{h^4}{4} V_K^2 \int_{-\infty}^{\infty} (f''(x))^2 dx - \frac{h^6}{24} V_K V_{K,4} \int_{-\infty}^{\infty} (f'''(x))^2 dx + o\left(\frac{1}{mh} + h^6\right) = \\ &= \frac{1}{mh} \int_{-\infty}^{\infty} K^2(y) dy + \frac{h^4}{4} V_K^2 R(f) - \frac{h^6}{24} V_K V_{K,4} R_3(f) + o\left(\frac{1}{mh} + h^6\right) \end{aligned}$$

□

La Proposición 3.8 permite considerar más términos en el desarrollo en serie gracias a la introducción de las condiciones g) y h). Dichas condiciones son una extensión de las condiciones d) y e) a los momentos de orden 3 y 4 de las funciones kernel.

Definición 3.9. Toda función kernel que verifique las condiciones

$$g) \int_{-\infty}^{\infty} y^3 K(y) dy = 0 \quad ; \quad h) \int_{-\infty}^{\infty} y^4 K(y) dy \equiv V_{K,4} < \infty$$

se denomina *función kernel de segundo grado*

Corolario 3.10. Bajo las hipótesis de la Proposición 3.8, la anchura de ventana que minimiza la parte asintótica del MISE, \tilde{h} , es una raíz de la ecuación:

$$Ax^7 - Bx^5 + C = 0$$

$$\text{con } A := \frac{V_K V_{K,4} R_3(f)}{4}, \quad B := V_K^2 R(f) \quad \text{y} \quad C := \frac{1}{m} \int_{-\infty}^{\infty} K(y) dy.$$

Demostración. De acuerdo a la Proposición 3.8, la parte asintótica del MISE viene dada por:

$$\text{MISE}(\hat{f}_K) \sim \frac{1}{mh} \int_{-\infty}^{\infty} K^2(y) dy + \frac{h^4}{4} V_K^2 R(f) - \frac{h^6}{24} V_K V_{K,4} R_3(f)$$

Para obtener los valores que optimizan dicha expresión, la derivamos:

$$\frac{d}{dh} \text{MISE}(\hat{f}_K) = -\frac{1}{mh^2} \int_{-\infty}^{\infty} K^2(y) dy + h^3 V_K^2 R(f) - \frac{h^5}{4} V_K V_{K,4} R_3(f)$$

e igualando a 0, obtenemos finalmente:

$$-\frac{1}{m\tilde{h}^2} \int_{-\infty}^{\infty} K^2(y) dy + \tilde{h}^3 V_K^2 R(f) - \frac{\tilde{h}^5}{4} V_K V_{K,4} R_3(f) = 0 \Rightarrow$$

$$\frac{1}{m} \int_{-\infty}^{\infty} K^2(y) dy - \tilde{h}^5 V_K^2 R(f) + \frac{\tilde{h}^7}{4} V_K V_{K,4} R_3(f) = 0$$

□

El Corolario 3.10 pone de relieve dos problemas que surgen al intentar mejorar el orden de convergencia del MISE:

1. A diferencia de la expresión (2.16), en este caso es imposible una expresión analítica para \tilde{h} ya que es raíz de un polinomio de grado 7 con coeficientes A , B y C no nulos.
2. Al problema inicial del desconocimiento del término $R(f)$ ahora se le añade otro término $R_3(f)$ cuya dependencia con la función de densidad real hace que lo desconozcamos también.

Método 3. El método propuesto por Hall, Sheather, Jones y Marron en 1991[7] busca solventar los dos problemas anteriores. Para ello, recupera la idea inicial de Hall y Marron, con la que podemos estimar las cantidades $R(f)$ y $R_3(f)$ mediante el uso de la función de densidad estimada $\hat{f}_K(x)$. Así, podemos definir entonces una estimación para la anchura de intervalo óptima dada por[7]:

$$\tilde{h}_H = \left[\frac{\int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 R(\hat{f}_K)} \right]^{\frac{1}{5}} m^{-\frac{1}{5}} + \frac{V_{K,4} R_3(\hat{f}_K)}{20 V_K R(\hat{f}_K)} \left[\frac{\int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 R(\hat{f}_K)} \right]^{\frac{3}{5}} m^{-\frac{3}{5}} = \hat{J}_1^{\frac{1}{5}} m^{-\frac{1}{5}} + \hat{J}_2 \hat{J}_1^{\frac{3}{5}} m^{-\frac{3}{5}} \quad (3.15)$$

$$\text{con } \hat{J}_1 := \frac{\int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 R(\hat{f}_K)} \quad \text{y} \quad \hat{J}_2 := \frac{V_{K,4} R_3(\hat{f}_K)}{20 V_K R(\hat{f}_K)}$$

Ahora comprobaremos que la anchura óptima propuesta en efecto es una buena aproximación (asintótica) para la anchura óptima. En primer lugar, calcularemos la derivada del valor aproximado asintótico del MISE:

$$\frac{d}{dh} \text{MISE} = -\frac{1}{mh^2} \int_{-\infty}^{\infty} K^2(y) dy + h^3 V_K^2 R(f) - \frac{h^5}{4} V_K V_{K,4} R_3(f)$$

Como la expresión (3.15) depende del tamaño de muestra m , podemos realizar las aproximaciones:

$$m^{-1} \tilde{h}_H^{-2} \sim \hat{J}_1^{-\frac{2}{5}} m^{-\frac{3}{5}} - 2 \hat{J}_2 m^{-1}$$

$$\tilde{h}_H^3 \sim \hat{J}_1^{\frac{3}{5}} m^{-\frac{3}{5}} + 3 \hat{J}_1 \hat{J}_2 m^{-1}$$

$$\tilde{h}_H^5 \sim \hat{J}_1 m^{-1}$$

Introduciendo estas aproximaciones en lo que sería la derivada de la estimación $\widehat{\text{MISE}}$, tendremos entonces:

$$\left. \frac{d}{dh} \widehat{\text{MISE}} \right|_{\tilde{h}_H} \sim - \left(\hat{J}_1^{-\frac{2}{5}} m^{-\frac{3}{5}} - 2 \hat{J}_2 m^{-1} \right) \hat{J}_1 V_K^2 R(\hat{f}_K) + \left(\hat{J}_1^{\frac{3}{5}} m^{-\frac{3}{5}} + 3 \hat{J}_1 \hat{J}_2 m^{-1} \right) V_K^2 R(\hat{f}_K) -$$

$$-5\hat{J}_1 m^{-1} \hat{J}_2 V_K^2 R(\hat{f}_K) = 0$$

de lo que deducimos que la estimación introducida en (3.15) para la anchura de ventana es una buena estimación asintótica.

Por último, conviene que justifiquemos el uso de desarrollos matemáticos tan sofisticados como el anterior, que conllevan a un incremento en la cantidad de términos que requieren de estimación. De acuerdo al cociente que definimos en (3.11), buscamos que ambos valores h_{PI} y h_{MISE} se acerquen lo máximo posible, es decir, que el cociente se aproxime a 0. Para el método introducido por Hall y Marron inicialmente en 1987, se obtiene que dicho ratio se comporta como $m^{-\frac{1}{10}}$ [6], lo que indica que tiende a 0 si el tamaño de muestra es muy grande, pero lo hace de una forma lenta. Sin embargo, al contemplar la estimación de (3.15), entonces se obtiene que el ratio se comporta como $m^{-\frac{1}{2}}$ [7], lo cual quiere decir que converge más rápidamente.

3.2.4. Elección por sobre suavización

En la anterior subsección, analizamos procedimientos para estimar el funcional $R(f)$, necesario para el cálculo de la ventana óptima. Ahora nos conformaremos con realizar una acotación de dicho valor ya que, de acuerdo a la expresión (2.16), una acotación inferior sobre el funcional $R(f)$, nos permite obtener una cota superior para el tamaño de ventana óptimo. Para afrontar este problema, nos basaremos en el resultado presentado por Terrell en 1990 [18].

Teorema 3.11 (Teorema de Terrell, 1990). *Sea una función de densidad f de media nula y de desviación típica la misma que presenta la función triweight, entonces se verifica:*

$$R(f) \geq R(\text{triweight}) \quad (3.16)$$

Demostración. Definamos la función diferencia entre f y la función triweight:

$$e(x) := f(x) - \frac{35}{32} (1 - x^2)^3 I_{[-1,1]}(x)$$

Como la función f y la función triweight verifican la condición b) de función de densidad, tenemos entonces:

$$\int_{-\infty}^{\infty} e(x) dx = \int_{-\infty}^{\infty} f(x) dx - \int_{-1}^1 \frac{35}{32} (1 - x^2)^3 dx = 1 - 1 = 0$$

y como por hipótesis, presentan la misma desviación típica (el mismo momento de segundo orden), también verifican:

$$\int_{-\infty}^{\infty} x^2 e(x) dx = \int_{-\infty}^{\infty} x^2 f(x) dx - \int_{-1}^1 \frac{35}{32} x^2 (1 - x^2)^3 dx = \sigma^2 - \sigma^2 = 0$$

Además, en $\mathbb{R} \setminus (-1, 1)$ la función $e(x)$ es positiva ya que coincide por construcción con $f(x)$.

Calculando la derivada segunda de e obtenemos:

$$e''(x) = f''(x) - \left[-\frac{105}{16} (1 - x^2)^2 + \frac{105}{4} x^2 (1 - x^2) \right] I_{[-1,1]}(x) = f''(x) - \frac{105}{16} [-1 + 6x^2 - 5x^4] I_{[-1,1]}(x) \Rightarrow$$

$$\begin{aligned}
f''(x) &= e''(x) + \frac{105}{16} [-1 + 6x^2 - 5x^4] I_{[-1,1]}(x) \\
\int_{-\infty}^{\infty} (f''(x))^2 dx &= \int_{-\infty}^{\infty} (e''(x))^2 dx + \int_{-\infty}^{\infty} \left(\frac{105}{16}\right)^2 [-1 + 6x^2 - 5x^4]^2 I_{[-1,1]}(x) dx + \\
&\quad + 2 \int_{-\infty}^{\infty} e''(x) \frac{105}{16} [-1 + 6x^2 - 5x^4] I_{[-1,1]}(x) dx \Rightarrow \\
R(f) &= R(e) + R(\text{triweight}) + 2 \int_{-\infty}^{\infty} e''(x) \frac{105}{16} [-1 + 6x^2 - 5x^4] I_{[-1,1]}(x) dx
\end{aligned}$$

Por construcción, tenemos:

$$\begin{aligned}
R(e) &:= \int_{-\infty}^{\infty} (e''(x))^2 dx \geq 0 \\
R(\text{triweight}) &:= \int_{-\infty}^{\infty} \left(\frac{105}{16}\right)^2 [-1 + 6x^2 - 5x^4]^2 I_{[-1,1]}(x) dx \geq 0
\end{aligned}$$

luego si demostramos que el tercer sumando es positivo, tendremos probado el resultado (3.16).

Observamos que:

$$\begin{aligned}
&\int_{-\infty}^{\infty} e''(x) \frac{105}{16} [-1 + 6x^2 - 5x^4] I_{[-1,1]}(x) dx = \frac{105}{16} \int_{-1}^1 e''(x) (-1 + 6x^2 - 5x^4) dx = \\
&= \frac{105}{16} \left[e'(x) (-1 + 6x^2 - 5x^4) \Big|_{-1}^1 - \int_{-1}^1 (12x - 20x^3) e'(x) dx \right] = \frac{105}{4} \int_{-1}^1 (5x^3 - 3x) e'(x) dx = \\
&= \frac{105}{4} \left[e(x) (5x^3 - 3x) \Big|_{-1}^1 - \int_{-1}^1 (15x^2 - 3) e(x) dx \right] = \frac{105}{2} [e(1) + e(-1)] + \frac{315}{4} \int_{-1}^1 (1 - 5x^2) e(x) dx = \\
&= \underbrace{\frac{105}{2} [e(1) + e(-1)]}_{:=A} + \underbrace{\frac{315}{4} \int_{-\infty}^{\infty} (1 - 5x^2) e(x) dx}_{:=B} + \underbrace{\int_{|x| \geq 1} (5x^2 - 1) e(x) dx}_{:=C}
\end{aligned}$$

El término B es nulo por los cálculos iniciales sobre la función $e(x)$:

$$\int_{-\infty}^{\infty} e(x) dx = \int_{-\infty}^{\infty} x^2 e(x) dx = 0 \Rightarrow B = 0$$

y los términos A y C son positivos ya que:

$$\begin{aligned}
&e(x) \geq 0 \quad \forall |x| \geq 1 \Rightarrow A \geq 0 \\
&\left\{ \begin{array}{l} (5x^2 - 1) \geq 0 \quad \forall |x| \geq \frac{1}{\sqrt{5}} \\ e(x) \geq 0 \quad \forall |x| \geq 1 \end{array} \right. \Rightarrow (5x^2 - 1) e(x) \geq 0 \quad \forall |x| \geq 1 \Rightarrow C \geq 0
\end{aligned}$$

□

Las hipótesis del Teorema de Terrell, además de ser bastante restrictivas, requerirían de un conocimiento previo de la función de densidad. No obstante, dada la forma del funcional que deseamos estimar, $R(f)$, podemos aplicar propiedades de la esperanza y de la varianza para generalizar el resultado (3.16).

Proposición 3.12. *Dada una función de densidad f se verifica que:*

$$R(f) \geq \frac{1}{243\sigma^5} R(\text{triweight}) \quad (3.17)$$

siendo σ la desviación típica asociada a f .

Demostración. Para aplicar el Teorema 3.11 a la función f debemos verificar en primer lugar que la esperanza de dicha función sea nula. La esperanza asociada a dicha función vendrá dada por:

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

por lo que desplazando la variable aleatoria como $X - E(X)$, pasamos a trabajar con una variable de esperanza nula y cuya función de densidad será $f(x + E(X))$. Como la transformación supone una traslación, se verifica:

$$R(f) = \int_{-\infty}^{\infty} (f''(x))^2 dx = \int_{-\infty}^{\infty} (f''(x + E(X)))^2 dx$$

es decir, el cambio de variable no afecta al funcional $R(f)$.

La segunda condición para aplicar el Teorema 3.11 es que la función de densidad f tenga la misma varianza que la función triweight, que es:

$$\sigma_{tri}^2 = \int_{-1}^1 x^2 \frac{35}{32} (1 - x^2)^3 dx = \frac{1}{9}$$

Realizando el cambio de variable $y = \frac{x}{3\sigma}$ la función de densidad pasará a ser $g(y) \equiv 3\sigma f(3\sigma y)$. Esta función g cumple las hipótesis del Teorema 3.11, por lo que tendremos entonces que $R(g) \geq R(\text{triweight})$. Y por último, calculamos el funcional para la función g en función de $R(f)$ como:

$$R(g) = \int_{-\infty}^{\infty} (3\sigma)^6 (f''(3\sigma x))^2 dx \underset{3\sigma x=y}{=} (3\sigma)^5 \int_{-\infty}^{\infty} (f''(y))^2 dy = 243\sigma^5 R(f)$$

de donde se obtiene el resultado. □

Corolario 3.13. *Sean X una variable aleatoria con función de densidad $f \in \mathcal{C}^2$, una muestra aleatoria simple de dicha variable (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a una función kernel K . Entonces, la amplitud de ventana que minimiza la parte asintótica del MISE verifica:*

$$\tilde{h} \leq \left[\frac{\int_{-\infty}^{\infty} K^2(y) dy}{35V_K^2} \right]^{\frac{1}{5}} 3\sigma m^{-\frac{1}{5}} := h_{OS} \quad (3.18)$$

siendo σ la desviación típica de la variable X .

Demostración. Utilizando el resultado la Proposición 3.12, e introduciendo (3.17) en el resultado del

Corolario 2.16 (2.16), obtenemos:

$$\tilde{h} \leq \left[\frac{243\sigma^5 \int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 R(triweight)} \right]^{\frac{1}{5}} m^{-\frac{1}{5}} = \left[\frac{\int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 R(triweight)} \right]^{\frac{1}{5}} 3\sigma m^{-\frac{1}{5}}$$

Y por otro lado, calculamos el funcional:

$$R(triweight) = \int_{-1}^1 \frac{105^2}{16^2} [-1 + 6x^2 - 5x^4]^2 dx = 35$$

demostrando así el resultado. \square

Nuevamente la desviación típica asociada a la función densidad real no la conocemos, por lo que tendremos que recurrir a una estimación de la misma, como discutimos en la [Subsección 3.2.2](#).

El valor obtenido en (3.18) es una cota superior para el tamaño óptimo de ventana \hat{h}_{OS} , por lo que el procedimiento se puede mejorar tomando fracciones de dicha anchura definiendo, por ejemplo una sucesión $\{\hat{h}_{OS}^i\}_{i \in \mathbb{N}}$ con $\hat{h}_{OS}^i = \hat{h}_{OS} 2^{-i}$ y elegimos el mejor intervalo. Este procedimiento es el que hemos seguido al calcular las estimaciones de la [Figura 3.5](#) donde se observa que, comparando la muestra completa y la muestra seleccionada de 1000, la mejor anchura viene dada por $h_{OS} 2^{-1} = 0.2474$.

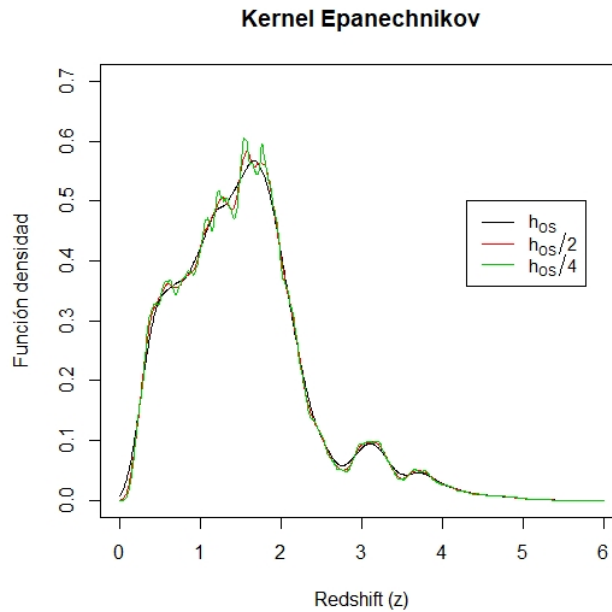


Figura 3.5: Estimación para la función de densidad con el intervalo óptimo (3.18). Para la muestra de 1000 cuásares obtenemos $h_{OS} = 0.4948$.

3.3. Elección por validación cruzada

Bajo este epígrafe englobamos una serie de procedimientos clásicos para abordar la elección de la anchura de ventana óptima \tilde{h} . Son tantos los métodos propuestos en este campo, que nos centraremos en el estudio de los siguientes:

- Validación cruzada por máxima verosimilitud.
- Validación cruzada por mínimos cuadrados
- Validación cruzada por sesgo
- Validación cruzada suave
- Validación cruzada indirecta

3.3.1. Validación cruzada por máxima verosimilitud

Como su propio nombre indica, este procedimiento considera la función de verosimilitud asociada a la muestra aleatoria simple, $L = \prod_i f(X_i)$, y busca la anchura de ventana que la maximice. Dicha función representa la credibilidad de la muestra bajo la densidad, pero como estamos realizando un análisis no paramétrico, desconocemos la forma que toma la función f , por lo que tendremos que recurrir a una estimación.

Definición 3.14. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a la función K . Se define el estimador de la función de verosimilitud como:

$$\hat{L}_{h,K} = \prod_{i=1}^m \hat{f}_{K,h}^*(X_i) = \frac{1}{(m-1)^m h^m} \prod_{i=1}^m \sum_{j \neq i} K\left(\frac{X_i - X_j}{h}\right) \quad (3.19)$$

En el estimador (3.19), la función $\hat{f}_{K,h}^*$ muestra la simplificación que hemos realizado en la expresión como consecuencia de que al evaluar en los datos de la muestra se tiene $(X_i - X_i) = 0$ que será un término constante. La ventana óptima \tilde{h}_{ML} será aquella que maximice esta verosimilitud. Algunos autores prefieren trabajar con la función $SC(h) \equiv \left(\hat{L}_{h,K}\right)^{\frac{1}{m}}$, cuyo máximo será el mismo que el de la función $\hat{L}_{h,K}$.

Proposición 3.15. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a la función $K \in \mathcal{C}^1$. La amplitud de ventana \tilde{h}_{ML} que maximiza el estimador de la función de verosimilitud asociada a la muestra verifica:

$$\tilde{h}_{ML} = \frac{1}{m} \sum_{i=1}^m \frac{\sum_{j \neq i} K'\left(\frac{X_i - X_j}{\tilde{h}_{ML}}\right) (X_j - X_i)}{\sum_{j \neq i} K\left(\frac{X_i - X_j}{\tilde{h}_{ML}}\right)} \quad (3.20)$$

Demostración. Buscamos el máximo del estimador (3.19). Teniendo en cuenta la monotonía de la función logaritmo, el trabajo es más sencillo si recurrimos a la expresión logarítmica de nuestra estimación de la función de máxima verosimilitud:

$$\ell_{h,K} = \sum_{i=1}^m \log \left[\sum_{j \neq i} K\left(\frac{X_i - X_j}{h}\right) \right] - m \log [(m-1)h] \quad (3.21)$$

puesto que los máximos de sendas funciones serán el mismo. Para obtener dicho máximo, calculamos la

derivada de $\ell_{h,K}$ e igualamos a 0:

$$\begin{aligned} \left. \frac{d}{dh} \ell_{h,K} \right|_{\tilde{h}_{ML}} &= h^{-2} \sum_{i=1}^m \frac{\sum_{j \neq i} K' \left(\frac{X_i - X_j}{h} \right) (X_j - X_i)}{\sum_{j \neq i} K \left(\frac{X_i - X_j}{h} \right)} - \frac{m}{h} \\ 0 &= \tilde{h}_{ML}^{-2} \sum_{i=1}^m \frac{\sum_{j \neq i} K' \left(\frac{X_i - X_j}{\tilde{h}_{ML}} \right) (X_j - X_i)}{\sum_{j \neq i} K \left(\frac{X_i - X_j}{\tilde{h}_{ML}} \right)} - \frac{m}{\tilde{h}_{ML}} \Rightarrow \\ \tilde{h}_{ML} &= \frac{1}{m} \sum_{i=1}^m \frac{\sum_{j \neq i} K' \left(\frac{X_i - X_j}{\tilde{h}_{ML}} \right) (X_j - X_i)}{\sum_{j \neq i} K \left(\frac{X_i - X_j}{\tilde{h}_{ML}} \right)} \end{aligned}$$

□

La expresión que acabamos de deducir no tiene dependencia en la función de densidad real pero, desgraciadamente, salvo casos muy simples, es imposible obtener una expresión analítica para la ventana óptima, y es necesario recurrir a cálculo computacional, el cual se puede volver muy pesado.

3.3.2. Validación cruzada por mínimos cuadrados

Este método considera una nueva función que también permite evaluar la calidad de la estimación realizada, el error integral cuadrático (de aquí en adelante ISE).

Definición 3.16. Dado un estimador funcional, en este caso de la función de densidad \hat{f} , el ISE del estimador se define como:

$$\text{ISE}(\hat{f}) = \int_{-\infty}^{\infty} \left(\hat{f}(x) - f(x) \right)^2 dx = \int_{-\infty}^{\infty} \hat{f}^2(x) dx - 2 \int_{-\infty}^{\infty} f(x) \hat{f}(x) dx + \int_{-\infty}^{\infty} f^2(x) dx \quad (3.22)$$

En concreto, el ISE para el estimador kernel $\hat{f}_{K,h}$ viene dado por:

$$\text{ISE}(\hat{f}_{K,h}) = \int_{-\infty}^{\infty} \left(\hat{f}_{K,h}(x) - f(x) \right)^2 dx = \int_{-\infty}^{\infty} \hat{f}_{K,h}^2(x) dx - 2 \int_{-\infty}^{\infty} f(x) \hat{f}_{K,h}(x) dx + \int_{-\infty}^{\infty} f^2(x) dx \quad (3.23)$$

El método de validación cruzada por mínimos cuadrados busca minimizar este valor. Observamos que el último término no depende en absoluto de h , por lo tanto, en nuestro objetivo de minimizar el ISE sólo tendremos que minimizar la expresión dada por:

$$\text{LS}(h) = \text{ISE} - \int_{-\infty}^{\infty} f^2(x) dx = \int_{-\infty}^{\infty} \hat{f}_{K,h}^2(x) dx - 2 \int_{-\infty}^{\infty} f(x) \hat{f}_{K,h}(x) dx = \int_{-\infty}^{\infty} \hat{f}_{K,h}^2(x) dx - 2\text{E}(\hat{f}_{K,h}) \quad (3.24)$$

Proposición 3.17. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a la función $K \in \mathcal{C}^1$. El estimador de la función $\text{LS}(h)$ viene dado por:

$$\widehat{\text{LS}}(h) = \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m K * K \left(\frac{X_j - X_i}{h} \right) - 2 \frac{1}{mh} \frac{1}{m-1} \sum_{i=1}^m \sum_{j \neq i} K \left(\frac{X_i - X_j}{h} \right) \quad (3.25)$$

Demostración. En primer lugar, podemos observar que el primer sumando de $LS(h)$ no requiere la estimación de f . Desarrollando este primer sumando obtenemos:

$$\begin{aligned} \int_{-\infty}^{\infty} \hat{f}_{K,h}^2(x) dx &= \int_{-\infty}^{\infty} \left(\frac{1}{mh} \sum_{i=1}^m K\left(\frac{x-X_i}{h}\right) \right) \left(\frac{1}{mh} \sum_{j=1}^m K\left(\frac{x-X_j}{h}\right) \right) dx = \\ &= \frac{1}{m^2 h^2} \sum_{i=1}^m \sum_{j=1}^m \int_{-\infty}^{\infty} K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right) dx \stackrel{z=(x-X_i)h^{-1}}{=} \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m \int_{-\infty}^{\infty} K(z) K\left(z - \frac{X_j - X_i}{h}\right) dz = \\ &= \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m K * K\left(\frac{X_j - X_i}{h}\right) \end{aligned}$$

La estimación se centra entonces en la esperanza de la función de densidad estimada $\hat{f}_{K,h}(x)$. Para ello, utilizaremos la propia función de densidad estimada, lo que nos conduce a la expresión:

$$\hat{E}(\hat{f}_{K,h}) = \frac{1}{mh} \frac{1}{m-1} \sum_{i=1}^m \sum_{j \neq i}^m K\left(\frac{X_i - X_j}{h}\right)$$

con lo que:

$$\begin{aligned} \widehat{LS}(h) &= \int_{-\infty}^{\infty} \hat{f}_{K,h}^2(x) dx - 2 \frac{1}{mh} \frac{1}{m-1} \sum_{i=1}^m \sum_{j \neq i}^m K\left(\frac{X_i - X_j}{h}\right) \\ &= \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m K * K\left(\frac{X_j - X_i}{h}\right) - 2 \frac{1}{mh} \frac{1}{m-1} \sum_{i=1}^m \sum_{j \neq i}^m K\left(\frac{X_i - X_j}{h}\right) \end{aligned} \quad (3.26)$$

□

Proposición 3.18. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a la función K . La amplitud de ventana \tilde{h}_{LS} que minimiza el estimador de la función $LS(h)$ verifica:

$$\tilde{h}_{LS} = \frac{\sum_{i=1}^m \sum_{j=1}^m K * K'\left(\frac{X_j - X_i}{\tilde{h}_{LS}}\right) (X_j - X_i) + \frac{2m}{m-1} \sum_{i=1}^m \sum_{j \neq i}^m K'\left(\frac{X_i - X_j}{\tilde{h}_{LS}}\right) (X_i - X_j)}{\sum_{i=1}^m \sum_{j=1}^m K * K\left(\frac{X_j - X_i}{\tilde{h}_{LS}}\right) - \frac{2m}{m-1} \sum_{i=1}^m \sum_{j \neq i}^m K\left(\frac{X_i - X_j}{\tilde{h}_{LS}}\right)} \quad (3.27)$$

Demostración. El estimador $\widehat{LS}(h)$ viene dado por (3.25). Para determinar la amplitud de ventana que lo minimiza, debemos calcular la derivada de dicho estimador respecto de h e igualar a 0. Para ello, en primer lugar conviene observar la siguiente propiedad de la convolución:

$$\begin{aligned} \frac{d}{dt} [f * g(t)] &= \frac{d}{dt} \int_{-\infty}^{\infty} f(x) g(x-t) dx = \int_{-\infty}^{\infty} \frac{d}{dt} [f(x) g(x-t)] dx = \\ &= - \int_{-\infty}^{\infty} f(x) g'(x+t) dx = -f * g'(t) \end{aligned}$$

Por lo tanto, la derivada del estimador $\widehat{LS}(h)$ será:

$$\begin{aligned} \frac{d}{dh} \widehat{LS}(h) &= -\frac{1}{m^2 h^2} \sum_{i=1}^m \sum_{j=1}^m K * K \left(\frac{X_j - X_i}{h} \right) + \frac{1}{m^2 h^3} \sum_{i=1}^m \sum_{j=1}^m K * K' \left(\frac{X_j - X_i}{h} \right) (X_j - X_i) + \\ &+ \frac{2}{m h^2 (m-1)} \sum_{i=1}^m \sum_{j \neq i} K \left(\frac{X_i - X_j}{h} \right) + \frac{2}{m h^3 (m-1)} \sum_{i=1}^m \sum_{j \neq i} K' \left(\frac{X_i - X_j}{h} \right) (X_i - X_j) \end{aligned}$$

E igualando a 0 obtenemos:

$$\begin{aligned} 0 &= -\frac{1}{m^2 \tilde{h}_{LS}^2} \sum_{i=1}^m \sum_{j=1}^m K * K \left(\frac{X_j - X_i}{\tilde{h}_{LS}} \right) + \frac{1}{m^2 \tilde{h}_{LS}^3} \sum_{i=1}^m \sum_{j=1}^m K * K' \left(\frac{X_j - X_i}{\tilde{h}_{LS}} \right) (X_j - X_i) + \\ &+ \frac{2}{m \tilde{h}_{LS}^2 (m-1)} \sum_{i=1}^m \sum_{j \neq i} K \left(\frac{X_i - X_j}{\tilde{h}_{LS}} \right) + \frac{2}{m \tilde{h}_{LS}^3 (m-1)} \sum_{i=1}^m \sum_{j \neq i} K' \left(\frac{X_i - X_j}{\tilde{h}_{LS}} \right) (X_i - X_j) \Rightarrow \\ \tilde{h}_{LS} &= \frac{\sum_{i=1}^m \sum_{j=1}^m K * K' \left(\frac{X_j - X_i}{\tilde{h}_{LS}} \right) (X_j - X_i) + \frac{2m}{m-1} \sum_{i=1}^m \sum_{j \neq i} K' \left(\frac{X_i - X_j}{\tilde{h}_{LS}} \right) (X_i - X_j)}{\sum_{i=1}^m \sum_{j=1}^m K * K \left(\frac{X_j - X_i}{\tilde{h}_{LS}} \right) - \frac{2m}{m-1} \sum_{i=1}^m \sum_{j \neq i} K \left(\frac{X_i - X_j}{\tilde{h}_{LS}} \right)} \end{aligned}$$

□

Esta expresión de la estimación de la anchura de la ventana óptima que minimiza el ISE presenta los mismos problemas que la deducida en (3.20): es necesario recurrir a cálculo computacional que se puede volver muy pesado para tamaños de muestra grandes y además presenta una gran variabilidad en función de la muestra.

3.3.3. Validación cruzada por sesgo

Este método retoma el problema de minimización de la parte asintótica del MISE (2.15). Para ello, al igual que en los métodos plug-in realizaremos una estimación del término desconocido $R(f)$, dada por la expresión (3.10), es decir, utilizaremos:

$$\hat{R}(f) = \frac{1}{m^2 h^5} \sum_{i=1}^m \sum_{j \neq i} K'' * K'' \left(\frac{X_j - X_i}{h} \right) \quad (3.28)$$

En lugar de seguir el proceso iterativo que caracterizaba los métodos plug-in, introducimos dicha expresión en la parte asintótica del MISE, de lo que se obtiene la función BC(h) que intentamos minimizar.

Definición 3.19. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a una función kernel $K \in \mathcal{C}^{(2)}$, se define como estimador de la parte asintótica del MISE la función:

$$BC(h) = \frac{V_K^2}{4m^2 h} \sum_{i=1}^m \sum_{j \neq i} K'' * K'' \left(\frac{X_j - X_i}{h} \right) + \frac{1}{mh} \int_{-\infty}^{\infty} K^2(y) dy \quad (3.29)$$

Este método combina parte del procedimiento de los métodos plug-in pero, en lugar de seguir un proceso iterativo, busca la amplitud de ventana que minimiza el estimador (3.29).

Proposición 3.20. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a una función kernel $K \in \mathcal{C}^3$. La amplitud de ventana \tilde{h}_{BC} que minimiza la función $BC(h)$ verifica la ecuación:

$$\tilde{h}_{BC} = \frac{\frac{V_K}{4m} \sum_{i=1}^m \sum_{j \neq i} K'' * K''' \left(\frac{X_j - X_i}{\tilde{h}_{BC}} \right) (X_j - X_i)}{\frac{V_K^2}{4m} \sum_{i=1}^m \sum_{j \neq i} K'' * K'' \left(\frac{X_j - X_i}{\tilde{h}_{BC}} \right) + \int_{-\infty}^{\infty} K^2(y) dy} \quad (3.30)$$

Demostración. Dada la función $BC(h)$ (3.29), buscamos la amplitud de ventana h que la minimice. Para ello, calculamos en primer lugar la derivada de la función:

$$\begin{aligned} \frac{d}{dh} BC(h) &= -\frac{V_K^2}{4m^2 h^2} \sum_{i=1}^m \sum_{j \neq i} K'' * K'' \left(\frac{X_j - X_i}{h} \right) + \frac{V_K}{4m^2 h^3} \sum_{i=1}^m \sum_{j \neq i} K'' * K''' \left(\frac{X_j - X_i}{h} \right) (X_j - X_i) \\ &\quad - \frac{1}{mh^2} \int_{-\infty}^{\infty} K^2(y) dy = \frac{V_K}{4m^2 \tilde{h}_{BC}^3} \sum_{i=1}^m \sum_{j \neq i} K'' * K''' \left(\frac{X_j - X_i}{\tilde{h}_{BC}} \right) (X_j - X_i) \\ &\quad - \frac{1}{mh^2} \left[\frac{V_K^2}{4m} \sum_{i=1}^m \sum_{j \neq i} K'' * K'' \left(\frac{X_j - X_i}{h} \right) + \int_{-\infty}^{\infty} K^2(y) dy \right] \end{aligned}$$

e igualando a 0 para obtener los puntos críticos se tiene:

$$\begin{aligned} &\frac{V_K}{4m^2 \tilde{h}_{BC}^3} \sum_{i=1}^m \sum_{j \neq i} K'' * K''' \left(\frac{X_j - X_i}{\tilde{h}_{BC}} \right) (X_j - X_i) \\ &\quad - \frac{1}{m \tilde{h}_{BC}^2} \left[\frac{V_K^2}{4m} \sum_{i=1}^m \sum_{j \neq i} K'' * K'' \left(\frac{X_j - X_i}{\tilde{h}_{BC}} \right) + \int_{-\infty}^{\infty} K^2(y) dy \right] = 0 \Rightarrow \\ &\tilde{h}_{BC} = \frac{\frac{V_K}{4m} \sum_{i=1}^m \sum_{j \neq i} K'' * K''' \left(\frac{X_j - X_i}{\tilde{h}_{BC}} \right) (X_j - X_i)}{\frac{V_K^2}{4m} \sum_{i=1}^m \sum_{j \neq i} K'' * K'' \left(\frac{X_j - X_i}{\tilde{h}_{BC}} \right) + \int_{-\infty}^{\infty} K^2(y) dy} \end{aligned}$$

□

A raíz de estos resultados podemos realizar una serie de observaciones. En primer lugar, a diferencia del método por máxima verosimilitud o por mínimos cuadrados, el método de validación cruzada por sesgo limita el uso de las funciones kernel, ya que se exige una mayor regularidad. Por otro lado, el comportamiento de la función $BC(h)$ muestra que, en determinados kernel, la amplitud óptima \tilde{h}_{BC} no se corresponde con un mínimo global, esto ocurre, por ejemplo, cuando empleamos un kernel normal.

Proposición 3.21. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado al kernel normal. Entonces se verifica que la función $BC(h)$ no presenta un mínimo global.

Demostración. Por hipótesis, la estimación se realiza con un kernel normal:

$$K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

El kernel $K \in \mathcal{C}^\infty$, por lo que podemos calcular su derivada segunda:

$$K''(x) = \frac{1}{\sqrt{2\pi}} [x^2 - 1] \exp\left(-\frac{x^2}{2}\right)$$

por lo tanto, la convolución $K'' * K''(x)$ tendrá la forma:

$$\begin{aligned} K'' * K''(x) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} [y^2 - 1] [(y-x)^2 - 1] \exp\left(-\frac{y^2}{2} - \frac{(y-x)^2}{2}\right) dy \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} [y^4 - 2y^3x + (x^2 - 1)y^2 - y^2 + 2yx - (x^2 - 1)] \exp\left(-\frac{y^2}{2} - \frac{(y-x)^2}{2}\right) dy \\ &= \frac{1}{2\pi} \exp\left(-\frac{x^2}{4}\right) \int_{-\infty}^{\infty} [y^4 - 2y^3x + (x^2 - 1)y^2 - y^2 + 2yx - (x^2 - 1)] \exp\left(-\left(y - \frac{x}{2}\right)^2\right) dy \\ &\stackrel{z=y-\frac{x}{2}}{=} \frac{\exp\left(-\frac{x^2}{4}\right)}{2\pi} \int_{-\infty}^{\infty} \left[\left(z + \frac{x}{2}\right)^4 - 2\left(z + \frac{x}{2}\right)^3 x + (x^2 - 2)\left(z + \frac{x}{2}\right)^2 + 2\left(z + \frac{x}{2}\right)x - (x^2 - 1)\right] \exp(-z^2) dz \\ &= \frac{\exp\left(-\frac{x^2}{4}\right)}{2\sqrt{\pi}} \left[\frac{1}{16} (x^4 + 12x^2 + 12) - \frac{x^2}{4} (x^2 + 6) + \frac{1}{4} (x^2 - 2)(x^2 + 2) + x^2 - (x^2 - 1) \right] \\ &= \frac{1}{32\sqrt{\pi}} \exp\left(-\frac{x^2}{4}\right) [12 - 12x^2 + x^4] \end{aligned}$$

La convolución nos permite observar que se trata de una función par, con lo que podemos calcular el límite cuando $h \rightarrow 0^+$ de:

$$\lim_{h \rightarrow 0^+} \frac{1}{h} K'' * K''\left(\frac{X_j - X_i}{h}\right) = \lim_{x \rightarrow \infty} x \frac{1}{32\sqrt{\pi}} \exp\left(-\frac{x^2}{4}\right) [12 - 12x^2 + x^4] = 0$$

lo que implica:

$$\lim_{h \rightarrow 0^+} \text{BC}(h) = \infty$$

Por otro lado, calculando el límite cuando $h \rightarrow \infty$, obtenemos:

$$\lim_{h \rightarrow \infty} \text{BC}(h) = \lim_{h \rightarrow \infty} \frac{1}{h} \left[\frac{V_K^2}{4m^2} \sum_{i=1}^m \sum_{j \neq i}^m K'' * K''(0) + \frac{1}{m} \int_{-\infty}^{\infty} K^2(y) dy \right] = 0$$

En definitiva, para que la función llegue a tener un mínimo global, será necesario que alcance valores negativos.

Dada la expresión (3.29), sumando y restando el término $i = j$, obtenemos entonces la expresión:

$$\text{BC}(h) = \frac{V_K^2}{4m^2h} \sum_{i=1}^m \sum_{j=1}^m K'' * K''\left(\frac{X_j - X_i}{h}\right) + \frac{1}{4mh} \left[4 \int_{-\infty}^{\infty} K^2(y) dy - V_K^2 K'' * K''(0) \right]$$

en el que, para muestras grandes, la parte dominante es la que acompaña el término $\frac{1}{mh}$. Para la función

kernel normal se verifica que:

$$4 \int_{-\infty}^{\infty} K^2(y) dy - V_K^2 K'' * K''(0) = \frac{2}{\sqrt{\pi}} - \frac{12}{32\sqrt{\pi}} = \frac{52}{32\sqrt{\pi}} > 0$$

de lo que concluimos que $BC(h) > 0$ para todo $h \in (0, \infty)$, es decir, no alcanzará un mínimo global. \square

Es importante resaltar que el resultado de la Proposición 3.21 no implica que no existan mínimos, pero todos ellos serán locales. Posteriormente, al evaluar estos procedimientos, observaremos que se puede seguir aplicando el método considerando el mínimo local más cercano a 0.

3.3.4. Validación cruzada suave

Este método plantea realizar la estimación directa en el término desconocido del MISE, la integral del sesgo. Recordamos que el MISE viene dado por:

$$\begin{aligned} \text{MISE}(h) &= \int_{-\infty}^{\infty} \text{Var}(\hat{f}_K(x)) dx + \int_{-\infty}^{\infty} \text{sesgo}^2(\hat{f}_K(x)) dx = \\ &= \int_{-\infty}^{\infty} \text{Var}(\hat{f}_K(x)) dx + \int_{-\infty}^{\infty} \text{E}^2[\hat{f}_K(x) - f(x)] dx \end{aligned}$$

De acuerdo a los cálculos realizados en la [Subsección 2.4.1](#), la primera integral no supone ningún problema ya que la parte asintótica dominante es conocida. El problema se presenta al analizar la segunda integral, ya que para el cálculo de la esperanza es necesario el conocimiento de la función de densidad f . El método de validación cruzada suave se basa en estimar dicho valor esperado.

Proposición 3.22. Sean X una variable aleatoria con función de densidad f , una muestra aleatoria simple de dicha variable (X_1, X_2, \dots, X_m) y \hat{f}_K el estimador kernel asociado a la función K . La integral del cuadrado del sesgo del estimador se puede estimar por:

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) \tilde{f}(y) dy - \tilde{f}(x) \right]^2 dx$$

siendo \tilde{f} un estimador de la función f .

Demostración. Nuestro objetivo es estimar el término:

$$\int_{-\infty}^{\infty} \text{sesgo}^2(\hat{f}_K(x)) dx = \int_{-\infty}^{\infty} \text{E}^2[\hat{f}_K(x) - f(x)] dx$$

Si desarrollamos el término de la esperanza obtenemos:

$$\text{E}[\hat{f}_K(x) - f(x)] = \text{E}[\hat{f}_K(x)] - f(x) = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) f(y) dy - f(x)$$

La estimación de esta última expresión mediante el uso del estimador \tilde{f} nos conduce a:

$$\tilde{\text{E}}[\hat{f}_K(x) - \tilde{f}(x)] = \int_{-\infty}^{\infty} \frac{1}{h} K\left(\frac{x-y}{h}\right) \tilde{f}(y) dy - \tilde{f}(x)$$

que al introducir en la expresión integral obtenemos:

$$\int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{h} K \left(\frac{x-y}{h} \right) \tilde{f}(y) dy - \tilde{f}(x) \right]^2 dx$$

□

El método presentado y analizado por Jones, Marron y Park en 1991[9] se basa en particularizar la estimación \tilde{f} (conocida como *estimación auxiliar* o *estimación pivote*) por medio de otro estimador kernel con una amplitud de ventana $g(h)$ dependiente de la amplitud de ventana h bajo la que se realiza la estimación principal.

Definición 3.23. Sean (X_1, X_2, \dots, X_m) una muestra aleatoria simple, $\hat{f}_{K,h}$ el estimador kernel asociado a la función K y a la amplitud de ventana h y el estimador kernel pivote $\hat{f}_{\tilde{K},g}$ asociado a la función \tilde{K} y a la amplitud de ventana $g(h)$. Un estimador de la parte asintótica del MISE de $\hat{f}_{K,h}$ viene dado por:

$$SC(h) = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{h} K \left(\frac{x-y}{h} \right) \hat{f}_{\tilde{K},g}(y) dy - \hat{f}_{\tilde{K},g}(x) \right]^2 dx \quad (3.31)$$

Al igual que en los anteriores procedimientos, la ventana óptima \tilde{h}_{SC} será aquella optimice la función $SC(h)$. Como la función es un estimador de la parte asintótica del MISE, la optimización es en verdad un problema de minimización, pero en la práctica se tienen varios problemas, ya que para cada función kernel pivote \tilde{K} y para cada función $g(h)$ se presentarán distintos resultados.

Resaltamos que la función pivote \tilde{K} no tiene porque ser igual a la función de la estimación principal K . De hecho, la estimación pivote nos permite introducir funciones kernel bastante suaves para realizar el trabajo posterior con una función más sencilla. Este el motivo, por el que se denomina que el método es *suave*. Otra forma de ver la suavidad adicional que estamos ganando es mediante el siguiente resultado:

Proposición 3.24. El estimador anteriormente definido se puede expresar de forma alternativa como:

$$SC(h) = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m \left[\left(\tilde{K} * K_{\frac{g}{h}} \right) * \left(\tilde{K} * K_{\frac{g}{h}} \right) \left(\frac{X_j - X_i}{h} \right) - 2\tilde{K} * \left(\tilde{K} * K_{\frac{g}{h}} \right)_{\frac{g}{h}} \left(\frac{X_j - X_i}{h} \right) + \frac{h}{g(h)} \tilde{K} * \tilde{K} \left(\frac{X_j - X_i}{g(h)} \right) \right] \quad (3.32)$$

con:

$$\tilde{K} * K_{\frac{g}{h}}(t) = \int_{-\infty}^{\infty} \tilde{K}(y) K \left(\frac{g(h)}{h} y - x \right) dy$$

Demostración. Por definición de estimadores kernel tenemos:

$$\hat{f}_{K,h} = \frac{1}{mh} \sum_{i=1}^m K \left(\frac{x - X_i}{h} \right) \quad ; \quad \hat{f}_{\tilde{K},h} = \frac{1}{mg(h)} \sum_{i=1}^m \tilde{K} \left(\frac{x - X_i}{g(h)} \right)$$

que al introducir en (3.31) nos permite expresar:

$$SC(h) = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \frac{1}{h} K \left(\frac{x-y}{h} \right) \frac{1}{mg(h)} \sum_{i=1}^m \tilde{K} \left(\frac{y - X_i}{g(h)} \right) dy - \hat{f}_{\tilde{K},g}(x) \right]^2 dx$$

$$-\frac{1}{mg(h)} \sum_{i=1}^m \tilde{K} \left(\frac{x - X_i}{g(h)} \right) \Bigg]^2 dx \quad (3.33)$$

Veamos que esta expresión se puede formular en términos de convoluciones.

En primer lugar podemos observar que el sumatorio se puede intercambiar con la integral como consecuencia de que es convergente al ser finito. Por otro lado, podemos realizar el cambio de variable $z = \frac{y - X_i}{g(h)}$ y gracias a la propiedad de simetría f) de las funciones kernel

$$K \left(\frac{x - y}{h} \right) = K \left(\frac{y - x}{h} \right)$$

obtenemos la expresión:

$$\begin{aligned} \text{SC}(h) &= \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \int_{-\infty}^{\infty} \left[\frac{1}{m} \sum_{i=1}^m \left[\frac{1}{h} \int_{-\infty}^{\infty} K \left(\frac{g(h)z}{h} - \frac{x - X_i}{h} \right) \tilde{K}(z) dz - \frac{1}{g(h)} \tilde{K} \left(\frac{x - X_i}{g(h)} \right) \right] \right]^2 dx \\ &= \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \underbrace{\int_{-\infty}^{\infty} \left[\frac{1}{m} \sum_{i=1}^m \left[\frac{1}{h} \tilde{K} * K_{\frac{g}{h}} \left(\frac{x - X_i}{h} \right) - \frac{1}{g(h)} \tilde{K} \left(\frac{x - X_i}{g(h)} \right) \right] \right]^2 dx}_{:=A} \end{aligned}$$

en donde definimos la convolución del enunciado, similar a la trabajada con anterioridad pero con un factor en el segundo término que hasta ahora no se presentaba porque siempre usábamos h .

Para calcular el término A , comencemos desarrollando el término cuadrático dentro de la segunda integral obtenemos:

$$\begin{aligned} &\frac{1}{m^2} \left[\sum_{i=1}^m \left[\frac{1}{h} \tilde{K} * K_{\frac{g}{h}} \left(\frac{x - X_i}{h} \right) - \frac{1}{g(h)} \tilde{K} \left(\frac{x - X_i}{g(h)} \right) \right] \right] \left[\sum_{j=1}^m \left[\frac{1}{h} \tilde{K} * K_{\frac{g}{h}} \left(\frac{x - X_j}{h} \right) - \frac{1}{g(h)} \tilde{K} \left(\frac{x - X_j}{g(h)} \right) \right] \right] = \\ &\frac{1}{m^2} \left[\sum_{i=1}^m \sum_{j=1}^m \left[\frac{1}{h^2} \left(\tilde{K} * K_{\frac{g}{h}} \left(\frac{x - X_i}{h} \right) \right) \left(\tilde{K} * K_{\frac{g}{h}} \left(\frac{x - X_j}{h} \right) \right) - \frac{2}{hg(h)} \left(\tilde{K} \left(\frac{x - X_i}{g(h)} \right) \right) \left(\tilde{K} * K_{\frac{g}{h}} \left(\frac{x - X_j}{h} \right) \right) \right. \right. \\ &\quad \left. \left. + \frac{1}{g^2(h)} \tilde{K} \left(\frac{x - X_i}{g(h)} \right) \tilde{K} \left(\frac{x - X_j}{g(h)} \right) \right] \right] \end{aligned}$$

Para realizar la integración de esta expresión, nos volvemos a encontrar con términos de convolución, por lo que realizaremos los cambios de variable $u = \frac{x - X_i}{h}$ y $v = \frac{x - X_j}{g(h)}$ con lo que llegamos a:

$$\begin{aligned} A &= \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \left[\frac{1}{h} \int_{-\infty}^{\infty} \left(\tilde{K} * K_{\frac{g}{h}}(u) \right) \left(\tilde{K} * K_{\frac{g}{h}} \left(u - \frac{X_j - X_i}{h} \right) \right) du \right. \\ &\quad \left. - \frac{2}{h} \int_{-\infty}^{\infty} \tilde{K}(v) \left(\tilde{K} * K_{\frac{g}{h}} \left(\frac{g(h)}{h} v - \frac{X_j - X_i}{h} \right) \right) dv + \frac{1}{g(h)} \int_{-\infty}^{\infty} \tilde{K}(v) \tilde{K} \left(v - \frac{X_j - X_i}{g(h)} \right) dv \right] \end{aligned}$$

que podemos expresar como:

$$A = \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m \left[\left(\tilde{K} * K_{\frac{g}{h}} \right) * \left(\tilde{K} * K_{\frac{g}{h}} \right) \left(\frac{X_j - X_i}{h} \right) - 2 \tilde{K} * \left(\tilde{K} * K_{\frac{g}{h}} \right)_{\frac{g}{h}} \left(\frac{X_j - X_i}{h} \right) \right. \\ \left. + \frac{h}{g(h)} \tilde{K} * \tilde{K} \left(\frac{X_j - X_i}{g(h)} \right) \right]$$

Por lo tanto, la función $SC(h)$ adopta la forma:

$$SC(h) = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m \left[\left(\tilde{K} * K_{\frac{g}{h}} \right) * \left(\tilde{K} * K_{\frac{g}{h}} \right) \left(\frac{X_j - X_i}{h} \right) \right. \\ \left. - 2 \tilde{K} * \left(\tilde{K} * K_{\frac{g}{h}} \right)_{\frac{g}{h}} \left(\frac{X_j - X_i}{h} \right) + \frac{h}{g(h)} \tilde{K} * \tilde{K} \left(\frac{X_j - X_i}{g(h)} \right) \right]$$

□

El método de validación cruzada por mínimos cuadrados se basaba en una estimación de la parte asintótica del ISE, mientras que este método se basa en la estimación de la parte asintótica del MISE. La diferencia entre ambos métodos es el cálculo de la esperanza, que no se realiza en el primero. Por lo tanto, si omitimos esa esperanza (estimada por la función pivote $\hat{f}_{\tilde{K},g}$), es decir, introduciendo la función delta de Dirac δ y tomando $g(h) = h$ retomamos el método de los mínimos cuadrados.

Corolario 3.25. *En las condiciones de la Definición 3.23, al reemplazar la estimación pivote por la función δ de Dirac y tomar $g(h) = h$, es decir, omitir la esperanza de la parte asintótica del MISE, se verifica:*

$$SC(h) \sim \widehat{LS(h)}$$

Demostración. Teniendo en cuenta la Proposición 3.24, y particularizando para el caso en que $g = h$ y $\tilde{K} = \delta$, se obtiene:

$$SC(h) = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m \left[(\delta * K) * (\delta * K) \left(\frac{X_j - X_i}{h} \right) \right. \\ \left. - 2 \delta * (\delta * K) \left(\frac{X_j - X_i}{h} \right) + \delta * \delta \left(\frac{X_j - X_i}{h} \right) \right]$$

En primer lugar, podemos observar que:

$$\delta * g(t) = \int_{-\infty}^{\infty} \delta(x) g(x-t) dx \stackrel{x-t=y}{=} \int_{-\infty}^{\infty} \delta(y+t) g(y) dy = g(-t)$$

y teniendo en cuenta la simetría de las funciones kernel y delta, se obtiene entonces que:

$$SC(h) = \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m \left[K * K \left(\frac{X_j - X_i}{h} \right) - 2K \left(\frac{X_j - X_i}{h} \right) + \delta \left(\frac{X_j - X_i}{h} \right) \right]$$

Realizando la aproximación bajo la cual $m \sim m-1$ y omitiendo los términos diagonales $i = j$ puesto

que no dependen de la muestra, se obtiene entonces la expresión:

$$\begin{aligned} \text{SC}(h) &\sim \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j \neq i}^m K * K \left(\frac{X_j - X_i}{h} \right) \\ &\quad + \frac{1}{m(m-1)h} \sum_{i=1}^m \sum_{j \neq i}^m \left[-2K \left(\frac{X_j - X_i}{h} \right) + \delta \left(\frac{X_j - X_i}{h} \right) \right] \end{aligned}$$

Como en el sumatorio ahora $j \neq i$, la función δ siempre tomará un valor nulo:

$$\text{SC}(h) \sim \frac{1}{mh} \int_{-\infty}^{\infty} K^2(x) dx + \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j \neq i}^m K * K \left(\frac{X_j - X_i}{h} \right) + \frac{1}{m(m-1)h} \sum_{i=1}^m \sum_{j \neq i}^m \left[-2K \left(\frac{X_j - X_i}{h} \right) \right]$$

Y como en m -veces $i = j$:

$$\begin{aligned} \sum_{i=1}^m \sum_{j=1}^m K * K \left(\frac{X_j - X_i}{h} \right) &= mK * K(0) + \sum_{i=1}^m \sum_{j \neq i}^m K * K \left(\frac{X_j - X_i}{h} \right) = \\ &= m \int_{-\infty}^{\infty} K^2(x) dx + \sum_{i=1}^m \sum_{j \neq i}^m K * K \left(\frac{X_j - X_i}{h} \right) \end{aligned}$$

Se obtiene finalmente:

$$\text{SC}(h) \sim \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m K * K \left(\frac{X_j - X_i}{h} \right) - \frac{2}{m(m-1)h} \sum_{i=1}^m \sum_{j \neq i}^m K \left(\frac{X_j - X_i}{h} \right) = \widehat{\text{LS}}(h)$$

□

3.3.5. Validación cruzada indirecta

Este método ha sido propuesto recientemente por Savchuk, Hart y Sheather en 2010[15]. Los autores, definen una familia de funciones \mathcal{K} que suponen una combinación particular de funciones kernel normales.

Definición 3.26. Sea \mathcal{K} la familia de funciones definidas por combinación de kernels normales:

$$\mathcal{K} = \left\{ (1 + \mu) \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2} \right) - \frac{\mu}{\sigma} \frac{1}{\sqrt{2\pi}} \exp \left(-\frac{x^2}{2\sigma^2} \right) : \mu \geq 0, \sigma > 0 \right\} \quad (3.34)$$

Cada elemento \mathcal{K} queda determinado por los parámetros μ y σ , por lo que se denotan como $K_{\mu, \sigma} \in \mathcal{K}$.

En la [Figura 3.6](#) se representan distintos ejemplos de funciones de la familia \mathcal{K} . Podemos observar que el kernel normal es un caso particular, en el que se toma $\mu = 0$ y $\sigma = 1$. Por otro lado, podemos observar que para determinados parámetros, las funciones $K_{\mu, \sigma}$ pueden tomar valores negativos. Es el caso de las funciones $K_{3,2}$ y $K_{3,0.5}$. Como nuestro objetivo es utilizar dichas funciones para estimar la función de densidad, debemos encontrar una subfamilia de funciones dentro de \mathcal{K} que nos sirva para este propósito.

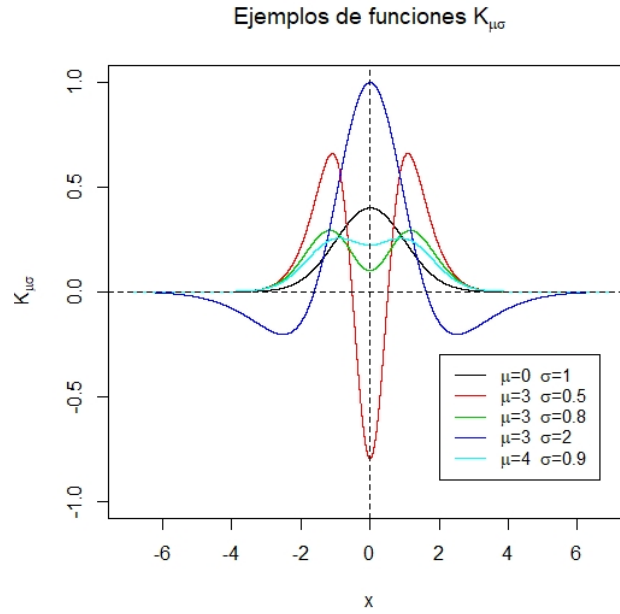


Figura 3.6: Ejemplos de funciones pertenecientes a la familia \mathcal{K} para diferentes valores de μ y σ .

Definición 3.27. Dentro de la familia de funciones \mathcal{K} definimos la subfamilia:

$$\mathcal{K}_{est} = \left\{ (1 + \mu) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) - \frac{\mu}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) : \mu \geq 0, \frac{\mu}{1 + \mu} \leq \sigma \leq 1 \right\} \quad (3.35)$$

cuyas funciones verifican que son positivas en todo⁷ \mathbb{R} .

Proposición 3.28. Sea $K_{\mu,\sigma} \in \mathcal{K}_{est}$. Se verifica que $K_{\mu,\sigma}$ es una función kernel.

Demostración. En primer lugar, demostraremos que se verifica:

$$K_{\mu,\sigma}(x) \geq 0 \quad \forall x \Leftrightarrow \frac{\mu}{1 + \mu} \leq \sigma \leq 1$$

La función $K_{\mu,\sigma}$ pertenece a la familia \mathcal{K} , luego toma la expresión:

$$K_{\mu,\sigma}(x) = (1 + \mu) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) - \frac{\mu}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

Los parámetros μ y σ son positivos, por lo que sabemos que se cumple:

$$\begin{aligned} K_{\mu,\sigma}(x) \geq 0 &\Leftrightarrow (1 + \mu) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \geq \frac{\mu}{\sigma} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \Leftrightarrow \\ (1 + \mu) &\geq \frac{\mu}{\sigma} \exp\left(-\frac{x^2}{2\sigma^2} + \frac{x^2}{2}\right) \Leftrightarrow \sigma \geq \frac{\mu}{1 + \mu} \exp\left(-\frac{x^2}{2} \left(\frac{1}{\sigma^2} - 1\right)\right) \end{aligned}$$

Si analizamos el argumento de la función exponencial, podemos observar dos casos: $\sigma > 1$ y $\sigma \leq 1$.

⁷La demostración a este resultado se presenta en la Proposición 3.28.

Consideremos que $\sigma > 1$, entonces $-\frac{x^2}{2} \left(\frac{1}{\sigma^2} - 1 \right) \geq 0 \forall x$. Por lo tanto, la función exponencial tomará valores superiores a 1 a medida que crece $|x|$. Existirá un valor $\kappa > 0$, tal que $\forall |x| \geq \kappa$:

$$\frac{\mu}{1+\mu} \exp \left(-\frac{x^2}{2} \left(\frac{1}{\sigma^2} - 1 \right) \right) > \sigma$$

luego la función $K_{\mu,\sigma}(x) < 0 \forall |x| \geq \kappa$. De lo que se concluye que necesariamente $\sigma \leq 1$.

Como necesariamente $\sigma \leq 1$, el argumento de la función exponencial verifica que $-\frac{x^2}{2} \left(\frac{1}{\sigma^2} - 1 \right) \leq 0 \forall x$. Por lo tanto, la exponencial tomará valores en $(0, 1]$. Como buscamos que la función $K_{\mu,\sigma}$ sea positiva para todo x , tomaremos la cota superior de 1 para la función exponencial de lo que deducimos:

$$\sigma \geq \frac{\mu}{1+\mu}$$

En conclusión:

$$K_{\mu,\sigma}(x) \geq 0 \forall x \Leftrightarrow \frac{\mu}{1+\mu} \leq \sigma \leq 1$$

por lo que se verifica la propiedad a) de la Definición 2.11 y motiva la elección de la subfamilia de la Definición 3.27.

Para demostrar que es una función kernel veremos que verifica las 5 propiedades restantes de la Definición 2.11. Podemos observar que toda función $K_{\mu,\sigma}$ se puede expresar como una combinación de la función kernel normal K_N dada por:

$$K_{\mu,\sigma}(x) = (1+\mu) K_N(x) - \frac{\mu}{\sigma} K_N\left(\frac{x}{\sigma}\right) \quad (3.36)$$

Teniendo en cuenta que la función K_N verifica las propiedades b), c), d), e) y f) de función kernel, obtenemos:

$$b) \int_{-\infty}^{\infty} K_{\mu,\sigma}(x) dx = (1+\mu) \int_{-\infty}^{\infty} K_N(x) dx - \frac{\mu}{\sigma} \int_{-\infty}^{\infty} K_N\left(\frac{x}{\sigma}\right) dx = (1+\mu) - \frac{\mu}{\sigma} \int_{-\infty}^{\infty} \sigma K_N(y) dy = 1$$

$$\begin{aligned} c) \exists \kappa \text{ t.q. } \int_{|x| \leq \kappa} K_{\mu,\sigma}(x) dx &= (1+\mu) \int_{|x| \leq \kappa} K_N(x) dx - \frac{\mu}{\sigma} \int_{|x| \leq \kappa} K_N\left(\frac{x}{\sigma}\right) dx = \\ &= (1+\mu) \int_{|x| \leq \kappa} K_N(x) dx - \mu \int_{|y| \leq \frac{\kappa}{\sigma}} K_N(y) dy \sim 1 + \mu - \mu = 1 \end{aligned}$$

$$d) \int_{-\infty}^{\infty} x K_{\mu,\sigma}(x) dx = (1+\mu) \int_{-\infty}^{\infty} x K_N(x) dx - \frac{\mu}{\sigma} \int_{-\infty}^{\infty} x K_N\left(\frac{x}{\sigma}\right) dx = -\mu \sigma \int_{-\infty}^{\infty} y K_N(y) dy = 0$$

$$\begin{aligned} e) \int_{-\infty}^{\infty} x^2 K_{\mu,\sigma}(x) dx &= (1+\mu) \int_{-\infty}^{\infty} x^2 K_N(x) dx - \frac{\mu}{\sigma} \int_{-\infty}^{\infty} x^2 K_N\left(\frac{x}{\sigma}\right) dx = (1+\mu) - \mu \sigma^2 \int_{-\infty}^{\infty} x K_N(y) dy \\ &= 1 + \mu - \mu \sigma^2 \leq \infty \end{aligned}$$

$$f) K_{\mu,\sigma}(-x) = (1+\mu) K_N(-x) - \frac{\mu}{\sigma} K_N\left(-\frac{x}{\sigma}\right) = (1+\mu) K_N(x) - \frac{\mu}{\sigma} K_N\left(\frac{x}{\sigma}\right) = K_{\mu,\sigma}(x)$$

□

Como se comentó anteriormente, $K_{3,2}$ y $K_{3,0.5}$ toman valores negativos justificados por la Proposición 3.28. En la función $K_{3,2}$ el parámetro $\sigma = 2 > 1$, por lo que presenta colas negativas como se justifica en la demostración de la Proposición 3.28. Por su parte, en $K_{3,0.5}$ el parámetro $\sigma = 0.5 < 0.75 = \frac{\mu}{\mu+1}$, por lo que toma valores negativos en la región central. La subfamilia de funciones \mathcal{K}_{est} será por tanto el conjunto de interés para desarrollar nuestro método, para ello veremos que todas ellas son funciones kernel.

El método de validación cruzada indirecta comienza seleccionando una función de la familia \mathcal{K}_{est} . Por la Proposición 3.28, podemos aplicar el método de la validación cruzada por mínimos cuadrados (Subsección 3.3.2), de lo que se obtiene una anchura de ventana $\tilde{h}_{LS,\mu,\sigma}$. Esta anchura es utilizada para determinar la anchura \tilde{h}_{IC} por medio de la relación existente entre las anchuras teóricas.

Proposición 3.29. *La anchura de ventana \tilde{h}_K que minimiza la parte asintótica del MISE del estimador asociado al kernel $K_{\mu,\sigma} \in \mathcal{K}_{est}$ y la anchura de ventana \tilde{h}_N que minimiza la parte asintótica del MISE del estimador asociado al kernel normal verifican la relación:*

$$\tilde{h}_K = \left[\frac{(1+\mu)^2 - \frac{\mu(1+\mu)2\sqrt{2}}{\sqrt{1+\sigma^2}} + \frac{\mu^2}{\sigma}}{(1+\mu - \mu\sigma^2)^2} \right]^{\frac{1}{5}} \tilde{h}_N \quad (3.37)$$

Demostración. De acuerdo a la expresión (2.16), la anchura de ventana óptima para el kernel normal viene dada por:

$$\tilde{h}_N = \left[\frac{\int_{-\infty}^{\infty} \frac{1}{2\pi} \exp(-x^2) dx}{\left(\int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx \right)^2 R(f)} \right]^{\frac{1}{5}} m^{-\frac{1}{5}} = \left(\frac{1}{2\sqrt{\pi}R(f)} \right)^{\frac{1}{5}} m^{-\frac{1}{5}}$$

Realizamos el mismo procedimiento con la función $K_{\mu,\sigma}(x)$, de donde obtenemos:

$$\begin{aligned} \tilde{h}_K &= \left[\frac{\int_{-\infty}^{\infty} \left(\frac{(1+\mu)^2}{2\pi} \exp(-x^2) - \frac{\mu(1+\mu)}{\sigma\pi} \exp\left(-\frac{x^2(1+\sigma^2)}{2\sigma^2}\right) + \frac{\mu^2}{2\pi\sigma^2} \exp\left(-\frac{x^2}{\sigma^2}\right) \right) dx}{\left(\int_{-\infty}^{\infty} \left(\frac{x^2(1+\mu)}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) - \frac{\mu x^2}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \right) dx \right)^2 R(f)} \right]^{\frac{1}{5}} m^{-\frac{1}{5}} = \\ &= \left[\frac{\frac{(1+\mu)^2}{2\sqrt{\pi}} - \frac{\mu(1+\mu)\sqrt{2}}{\sqrt{\pi}\sqrt{1+\sigma^2}} + \frac{\mu^2}{\sigma 2\sqrt{\pi}}}{(1+\mu - \mu\sigma^2)^2 R(f)} \right]^{\frac{1}{5}} m^{-\frac{1}{5}} = \left[\frac{(1+\mu)^2 - \frac{\mu(1+\mu)2\sqrt{2}}{\sqrt{1+\sigma^2}} + \frac{\mu^2}{\sigma}}{(1+\mu - \mu\sigma^2)^2} \right]^{\frac{1}{5}} \left(\frac{1}{2\sqrt{\pi}R(f)} \right)^{\frac{1}{5}} m^{-\frac{1}{5}} \end{aligned}$$

Lo que nos conduce a la relación:

$$\tilde{h}_K = \left[\frac{(1+\mu)^2 - \frac{\mu(1+\mu)2\sqrt{2}}{\sqrt{1+\sigma^2}} + \frac{\mu^2}{\sigma}}{(1+\mu - \mu\sigma^2)^2} \right]^{\frac{1}{5}} \tilde{h}_N$$

□

La Proposición 3.30 muestra que \tilde{h}_N y \tilde{h}_k se relacionan mediante una constante que depende exclusivamente de los parámetros μ y σ , los cuales nos permiten generar distintas funciones dentro de la subfamilia \mathcal{K}_{est} . El método de validación cruzada indirecta establece entonces que, dada la estimación de la anchura

de ventana por la validación cruzada de mínimos cuadrados, $\tilde{h}_{LS,\mu,\sigma}$, se introduce en la expresión (3.37) como \tilde{h}_K y se utiliza la anchura de ventana:

$$\tilde{h}_{IC} = \left[\frac{(1 + \mu)^2 - \frac{\mu(1+\mu)2\sqrt{2}}{\sqrt{1+\sigma^2}} + \frac{\mu^2}{\sigma}}{(1 + \mu - \mu\sigma^2)^2} \right]^{-\frac{1}{5}} \tilde{h}_{LS,\mu,\sigma} \quad (3.38)$$

Este método supone una elección de anchura de ventana para el kernel normal, pero sin necesidad de trabajar en su cálculo con dicha función, motivo por el que recibe el calificativo de *indirecto*.

3.3.6. Evaluación de los métodos de validación cruzada

Como hemos apuntado en los distintos apartados, los métodos de validación cruzada proporcionan estimaciones para la anchura de ventana óptima, pero no es posible obtener una forma explícita de dicho valor como se deduce de las expresiones (3.20), (3.27) o (3.30). Por lo tanto, la mejor forma de calcular dicha anchura es buscando computacionalmente los valores que optimicen las diferentes expresiones.

Omitiremos de este estudio los métodos de validación cruzada suave e indirecta: el primero, porque supone una mayor complicación numérica y computacional que el método de validación cruzada por mínimos cuadrados; el segundo, porque sólo se presenta para una determinada subfamilia de funciones kernel, y replica los procedimientos de los otros métodos.

Dadas las expresiones de los funcionales $\ell_{h,K}$ (3.21), $\widehat{LS}(h)$ (3.25) y $BC(h)$ (3.29), observamos que para la aplicación de los dos primeros métodos podemos utilizar cualquiera de las funciones kernel listadas en la Tabla 2.2. Sin embargo, para la aplicación el último es necesario recurrir a funciones kernel con suficiente regularidad, motivo por el cual restringiremos nuestro estudio a los kernel normal y triweight.

En la Figura 3.7 se muestra el cálculo de las funciones de validación cruzada para una muestra de 500 cuásares. Observamos que tanto el método por máxima verosimilitud como el método por mínimos cuadrados no presentan ningún inconveniente: es posible determinar los máximos globales (en el caso del método por máxima verosimilitud) y los mínimos globales (en el caso del método por mínimos cuadrados).

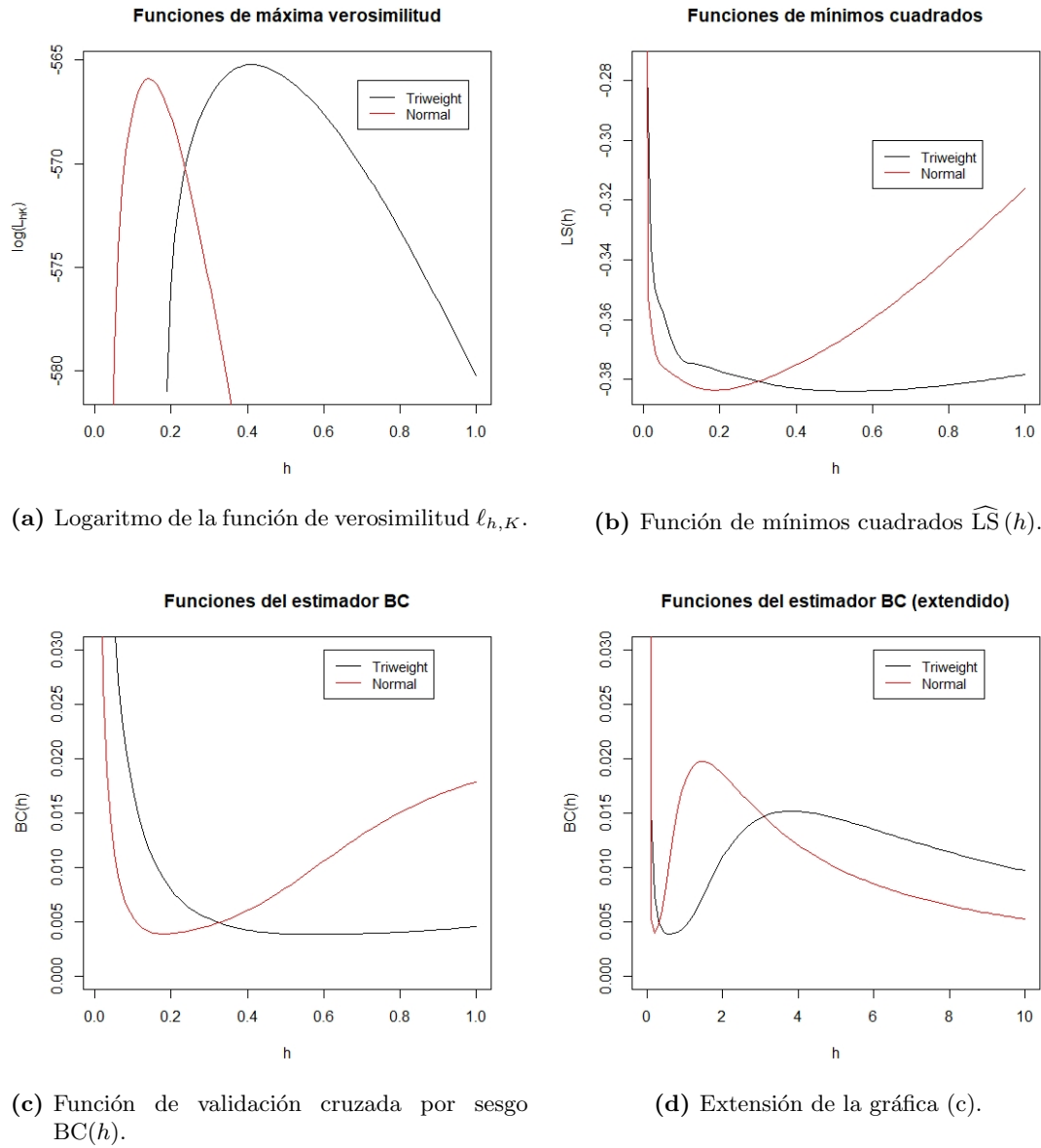
No ocurre lo mismo con el método de validación cruzada por sesgo. Ya demostramos que en el caso de la función kernel normal, no es posible determinar un mínimo global como consecuencia del comportamiento la función $BC(h)$ en la Proposición 3.21. La demostración del resultado se basaba en que la función normal verifica la propiedad:

$$4 \int_{-\infty}^{\infty} K^2(y) dy - V_K^2 K'' * K''(0) > 0$$

de lo que se deducía el problema de la no existencia de un mínimo global. El caso de la función kernel normal ya lo demostramos, y el comportamiento que observamos en la Figura 3.7 (d), sugiere que también se presenta para la función triweight. En efecto, evaluando la expresión (3.39), obtenemos que:

$$4 \int_{-\infty}^{\infty} K^2(y) dy - V_K^2 K'' * K''(0) = 4 \frac{350}{429} - \frac{4}{81} \frac{64}{315} \sim 3.25 > 0$$

de donde se deduce el comportamiento de la gráfica.

(a) Logaritmo de la función de verosimilitud $\ell_{h,K}$.(b) Función de mínimos cuadrados $\widehat{LS}(h)$.

(c) Función de validación cruzada por sesgo BC(h).

(d) Extensión de la gráfica (c).

Figura 3.7: Funciones de validación cruzada para una muestra aleatoria de 500 cuásares extraídos de la muestra completa de la SDSS.

Corolario 3.30. La función $BC(h)$ asociada al estimador de una función kernel $K \in \mathcal{C}^{(2)}$ y que verifica:

$$4 \int_{-\infty}^{\infty} K^2(y) dy - V_K^2 K'' * K''(0) > 0 \quad (3.39)$$

no presenta un mínimo global respecto de h .

Demostración. Los pasos finales de la demostración de la Proposición 3.21 justifican este resultado. \square

Sin embargo, la [Figura 3.7 \(c\)](#) muestra que pese a no existir un mínimo global, sí que se observa un mínimo local, el cual se corresponde con el valor óptimo de la anchura de ventana que determina el método.

En la [Tabla 3.5](#) se muestran los valores obtenidos de la anchura de ventana para los diferentes métodos de validación cruzada en comparación con los otros métodos analizados hasta ahora. Por último, la [Figura 3.8](#) muestra las estimaciones realizadas con la anchura obtenida en cada método.

Tabla 3.5: Selección de anchuras de ventana por diferentes procedimientos. Se ha fijado una precisión de 10^{-2} para todos los cálculos.

Método de selección	Kernel Triweight	Kernel normal
Validación Cruzada Máxima Verosimilitud	0.41	0.14
Validación Cruzada Mínimos Cuadrados	0.55	0.19
Validación Cruzada Sesgo	0.58	0.18
Plug-in	0.46	0.15
Referida a la normal	0.74	0.25
Sobre suavización	0.80	0.27

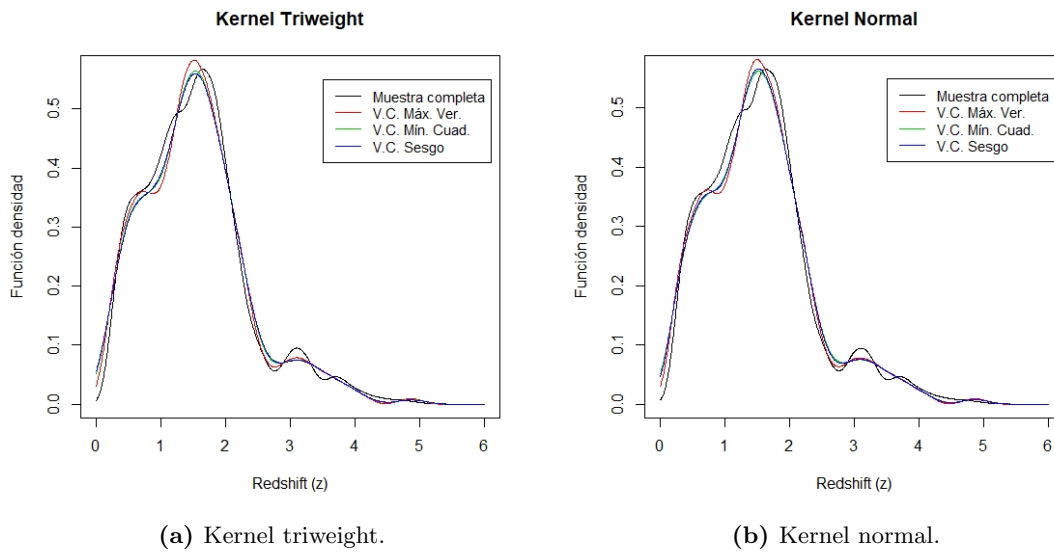


Figura 3.8: Estimación de la función de densidad para una muestra de 500 cuásares, bajo diferentes métodos de validación cruzada y bajo diferentes kernel.

3.4. Procedimientos bootstrap

3.4.1. Introducción y procedimiento básico

Si bien los procedimientos *bootstrap* se introdujeron con la finalidad de aproximar la distribución de estadísticos, se han generalizado a la selección de ciertas anchuras de ventana[2]. Puesto que la anchura de ventana óptima es aquella que minimiza el MISE, el procedimiento bootstrap busca estimar dicho error, calculando lo que se conoce como su *forma bootstrap* $MISE_B[2]$.

Definición 3.31. Sean (X_1, X_2, \dots, X_m) una muestra aleatoria simple, \hat{f}_g la estimación de la función de densidad de la muestra en base a una anchura de ventana pivote g , $(X_1^B, X_2^B, \dots, X_m^B)$ una muestra bootstrap de la muestra inicial y $\hat{f}_{B,h}$ la estimación de la función de densidad de la muestra bootstrap en base a una anchura de ventana h . Se define el MISE bootstrap como:

$$\text{MISE}_B(h) = E_B \left[\int_{-\infty}^{\infty} \left(\hat{f}_{B,h}(x) - \hat{f}_g(x) \right)^2 dx \right] \quad (3.40)$$

donde E_B denota la esperanza referida a la muestra bootstrap $(X_1^B, X_2^B, \dots, X_m^B)$, que se toma de la muestra inicial.

Las diferentes posibilidades en la elección de la anchura pivote g y en la generación de la muestra bootstrap $(X_1^B, X_2^B, \dots, X_m^B)$ nos permiten definir diferentes procedimientos.

3.4.2. Bootstrap suave sin anchura pivote

El primer procedimiento bootstrap contemplado realiza una estimación pivote con el mismo kernel y la misma anchura de ventana h .

Método 4. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) , estimamos la función de densidad $\hat{f}_{K,h}$ como:

$$\hat{f}_{K,h}(x) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - X_i}{h}\right)$$

Para generar la muestra bootstrap, utilizamos dicha estimación de la función de densidad. Para ello calculamos la función de distribución como:

$$\hat{F}(x) = \int_{-\infty}^x \hat{f}_{K,h}(y) dy = \frac{1}{mh} \int_{-\infty}^x \sum_{i=1}^m K\left(\frac{y - X_i}{h}\right) dy$$

y, haciendo uso de la inversa de la función de distribución $\hat{F}(z)$, y generando valores aleatorios con una distribución uniforme $U_{[0,1]}$, podremos generar la muestra bootstrap como:

$$(X_1^B, X_2^B, \dots, X_m^B) = \left(\hat{F}^{-1}(U_{[0,1]}), \hat{F}^{-1}(U_{[0,1]}), \dots, \hat{F}^{-1}(U_{[0,1]}) \right)$$

Con lo que la estimación bootstrap vendrá dada entonces por:

$$\hat{f}_{B,h}(x) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - X_i^B}{h}\right)$$

De esta manera, se busca minimizar la función:

$$\text{MISE}_B(h) = \frac{1}{m^2 h^2} E_B \left[\int_{-\infty}^{\infty} \left[\sum_{i=1}^m \left[K\left(\frac{x - X_i^B}{h}\right) - K\left(\frac{x - X_i}{h}\right) \right] \right]^2 dx \right]$$

y se toma como anchura de ventana \tilde{h}_{BSW} , el valor que la minimiza.

3.4.3. Bootstrap suave con anchura pivote

El procedimiento a seguir es análogo al presentado en la subsección anterior, pero ahora la estimación inicial de la función de densidad se realiza con una anchura de ventana $g \neq h$. Una primera posibilidad que se nos presenta es tomar la anchura pivote como aquella obtenida por el método de validación cruzada por mínimos cuadrados $g = \tilde{h}_{LS}$. La segunda posibilidad consiste en introducir $g = g(h)$, y realizar todo el desarrollo teórico del MISE para obtener una expresión que lo minimice.

La primera posibilidad presenta una gran inconveniente: como se deduce del desarrollo de la [Subsección 3.3.2](#): no es posible obtener un valor explícito de la anchura \tilde{h}_{LS} que muestre su dependencia con h . La segunda alternativa requiere de un trabajo bastante exhaustivo de las expresiones y se encuentra explicado en el artículo de Cao, Cuevas y Manteiga[2].

Método 5. Otra posibilidad de trabajo es realizar el procedimiento bootstrap M -veces, para así tomar el valor esperado de $\text{MISE}_B(h)$, y mediante cálculos numéricos determinar bajo que valor de h se obtiene la mejor acotación inferior para la función. Es decir, se generan M muestras $(X_1^{B,j}, X_2^{B,j}, \dots, X_m^{B,j})$ con $j \in \{1, 2, \dots, M\}$, a partir de la función de densidad $\hat{f}_{K, \tilde{h}_{LS}}(x)$. Para cada j se realiza la estimación $\hat{f}_{B,h}^j(x)$, y el MISE se aproxima entonces por:

$$\widehat{\text{MISE}}_B(h) = \frac{1}{M} \sum_{j=1}^M \left[\int_{-\infty}^{\infty} \left(\frac{1}{mh} \sum_{i=1}^m K\left(\frac{x - X_i^{B,j}}{h}\right) - \frac{1}{m\tilde{h}_{LS}} \sum_{i=1}^m K\left(\frac{x - X_i}{\tilde{h}_{LS}}\right) \right)^2 dx \right]$$

y se busca la anchura de ventana que minimiza esta función.

3.4.4. Bootstrap no-suave

Los procedimientos contemplados hasta ahora se enmarcan dentro de los denominados *procedimientos bootstrap suaves* por el hecho de que para obtener la muestra bootstrap $(X_1^B, X_2^B, \dots, X_m^B)$ realizamos una primera estimación de la función de densidad. En el *procedimiento bootstrap no-suave*, la muestra bootstrap se toma directamente de la función de distribución empírica, es decir, se trabaja con un funcional de distribución no suave.

Método 6. Dada la muestra inicial (X_1, X_2, \dots, X_m) , calculamos la función de distribución empírica por (1.3). De esta manera, podemos calcular la inversa de dicha función $\hat{F}^{-1}(x)$ de una forma sencilla como:

$$\hat{F}^{-1}(x) = \inf \left\{ y \in \mathbb{R} ; \hat{F}(y) \geq x \right\} = \min \left\{ X \in (X_1, X_2, \dots, X_m) ; \frac{\#X_i \leq X}{m} \geq x \right\}$$

generando la muestra bootstrap usando m valores aleatorios de la distribución uniforme $U_{[0,1]}$.

Ahora bien, dado que la muestra bootstrap la obtenemos directamente de la distribución empírica, estamos aumentando considerablemente el sesgo frente al método suave con estimación pivote propuesto con anterioridad. Por este motivo, plantear un modelo sin anchura pivote ($h = g$) conduciría a obtener un valor de h que no necesariamente minimiza la expresión del $\text{MISE}_B(h)$.

Método 7. Hall propone tomar una muestra bootstrap de menor tamaño $(X_1^B, X_2^B, \dots, X_t^B)$ con $t < m$ y utilizar una anchura pivote g que se relacione con la anchura de ventana h como $h = gt^{-\frac{1}{5}}m^{\frac{1}{5}}$ [2]. Así,

la función a optimizar será:

$$\text{MISE}_B(g) = E_B \left[\int_{-\infty}^{\infty} \left(\frac{1}{m^{\frac{1}{5}} t^{\frac{4}{5}} g} \sum_{j=1}^t K \left(\frac{x - X_j^B}{gm^{\frac{1}{5}} t^{-\frac{1}{5}}} \right) - \frac{1}{mg} \sum_{i=1}^m K \left(\frac{x - X_i}{g} \right) \right)^2 dx \right]$$

de forma que se busca la anchura pivote \tilde{g} que minimiza esta expresión.

3.4.5. Evaluación de los procedimientos bootstrap

Aplicaremos los procedimientos bootstrap analizados con detalle en las anteriores subsecciones al ejemplo de la distribución de cuásares en la campaña SDSS. Utilizaremos un estimador basado en un kernel normal, para el que obtenemos una serie de resultados que nos simplifican el cálculo computacional de los métodos bootstrap.

Proposición 3.32. *Dado el método bootstrap de Hall (método 7), si la estimación se realiza con un kernel normal, entonces la expresión del MISE_B toma la forma:*

$$\begin{aligned} \text{MISE}_B(g) = \frac{1}{2\sqrt{\pi}g} & \left[\frac{1}{ta} + \sum_{i=1}^m \sum_{j=1}^m \left[\frac{(t-1)}{m^2 ta} \exp \left(-\frac{(X_j - X_i)^2}{4g^2 a^2} \right) \right. \right. \\ & \left. \left. - \frac{2}{m^2} \sqrt{\frac{2}{1+a^2}} \exp \left(-\frac{(X_j - X_i)^2}{2g^2(1+a^2)} \right) + \frac{1}{m^2} \exp \left(-\frac{(X_j - X_i)^2}{4g^2} \right) \right] \right] \end{aligned} \quad (3.41)$$

con $a = m^{\frac{1}{5}} t^{-\frac{1}{5}}$.

Demostración. De acuerdo a la expresión del MISE_B del método 7, podemos desarrollar el integrando como:

$$\begin{aligned} \left[\frac{1}{tga} \sum_{i=1}^t K \left(\frac{x - X_i^B}{ga} \right) - \frac{1}{mg} \sum_{i=1}^m K \left(\frac{x - X_i}{g} \right) \right]^2 &= \frac{1}{t^2 a^2 g^2} \sum_{i=1}^t \sum_{j=1}^t K \left(\frac{x - X_i^B}{ga} \right) K \left(\frac{x - X_j^B}{ga} \right) \\ &- \frac{2}{g^2 tma} \sum_{i=1}^t \sum_{j=1}^m K \left(\frac{x - X_i^B}{ga} \right) K \left(\frac{x - X_j}{g} \right) + \frac{1}{m^2 g^2} \sum_{i=1}^m \sum_{j=1}^m K \left(\frac{x - X_i}{g} \right) K \left(\frac{x - X_j}{g} \right) \end{aligned}$$

donde expresamos el factor a como la relación entre g y h dada por $h = ga$.

Ahora pasaremos a calcular su esperanza respecto al procedimiento bootstrap. Podemos observar que el último sumando no tiene ninguna dependencia con la muestra bootstrap, luego su valor será constante. En el caso del segundo término, sólo un factor depende de la muestra bootstrap, de manera que tendremos que calcularlo. Como la muestra bootstrap se extrae directamente de la distribución empírica, entonces podemos observar que de forma general:

$$E_B [g(X^B)] = \frac{1}{m} \sum_{i=1}^m g(X_i)$$

de manera que obtenemos:

$$E_B \left[K \left(\frac{x - X_i^B}{ga} \right) \right] = \frac{1}{m} \sum_{k=1}^m K \left(\frac{x - X_k^B}{ga} \right)$$

En el caso del primer sumando, al evaluar la esperanza del primer término debemos distinguir dos casos: cuando $i = j$ y cuando $i \neq j$, de manera que entonces se tiene:

$$\begin{aligned} E_B \left[\sum_{i=1}^t \sum_{j=1}^t K \left(\frac{x - X_i^B}{ga} \right) K \left(\frac{x - X_j^B}{ga} \right) \right] &= \sum_{i=1}^t E_B \left[K^2 \left(\frac{x - X_i^B}{ga} \right) \right] + (t-1) \sum_{i=1}^t \left[E_B \left[K \left(\frac{x - X_i^B}{ga} \right) \right] \right]^2 \\ &= \frac{t}{m} \sum_{k=1}^m K^2 \left(\frac{x - X_k}{ga} \right) + \frac{t(t-1)}{m^2} \sum_{k=1}^m \sum_{r=1}^m K \left(\frac{x - X_k}{ga} \right) K \left(\frac{x - X_r}{ga} \right) \end{aligned}$$

Por lo tanto, el MISE bootstrap será de la forma:

$$\begin{aligned} \text{MISE}_B(g) &= \frac{1}{tmg^2a^2} \sum_{k=1}^m \int_{-\infty}^{\infty} K^2 \left(\frac{x - X_k}{ga} \right) dx + \frac{(t-1)}{m^2tg^2a^2} \sum_{k=1}^m \sum_{r=1}^m \int_{-\infty}^{\infty} K \left(\frac{x - X_k}{ga} \right) K \left(\frac{x - X_r}{ga} \right) dx \\ &\quad - \frac{2}{g^2am^2} \sum_{k=1}^m \sum_{r=1}^m \int_{-\infty}^{\infty} K \left(\frac{x - X_k}{ga} \right) K \left(\frac{x - X_r}{g} \right) dx + \frac{1}{g^2m^2} \sum_{k=1}^m \sum_{r=1}^m \int_{-\infty}^{\infty} K \left(\frac{x - X_k}{g} \right) K \left(\frac{x - X_r}{g} \right) dx \end{aligned}$$

Siguiendo la notación de la Proposición 3.24, esta expresión en forma de convoluciones es:

$$\begin{aligned} \text{MISE}_B(g) &= \frac{1}{gta} \int_{-\infty}^{\infty} K^2(y) dy + \frac{(t-1)}{gtm^2a} \sum_{k=1}^m \sum_{r=1}^m K * K \left(\frac{X_k - X_r}{ga} \right) \\ &\quad - \frac{2}{gam^2} \sum_{k=1}^m \sum_{r=1}^m K * K_{\frac{1}{a}} \left(\frac{X_k - X_r}{ga} \right) + \frac{1}{gm^2} \sum_{k=1}^m \sum_{r=1}^m K * K \left(\frac{X_k - X_r}{g} \right) \end{aligned}$$

Por lo tanto, sólo tendremos que calcular las convoluciones:

$$\begin{aligned} K * K(t) &= \int_{-\infty}^{\infty} K(x) K(x-t) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left(-\frac{x^2}{2} \right) \exp \left(-\frac{(x-t)^2}{2} \right) dx \\ &= \frac{1}{2\pi} \exp \left(-\frac{t^2}{4} \right) \int_{-\infty}^{\infty} \exp \left(-\left(x - \frac{t}{2} \right)^2 \right) dx = \frac{1}{2\sqrt{\pi}} \exp \left(-\frac{t^2}{4} \right) \end{aligned}$$

y:

$$\begin{aligned} K * K_{\frac{1}{a}}(t) &= \int_{-\infty}^{\infty} K(x) K(a^{-1}x - t) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp \left(-\frac{x^2}{2} \right) \exp \left(-\frac{(a^{-1}x - t)^2}{2} \right) dx \\ &= \frac{1}{2\pi} \exp \left(-\frac{t^2}{2(1+a^{-2})} \right) \int_{-\infty}^{\infty} \exp \left(-\left(\sqrt{\frac{1+a^{-2}}{2}} x - \frac{a^{-1}t}{\sqrt{2}\sqrt{1+a^{-2}}} \right)^2 \right) dx \\ &= \frac{1}{\sqrt{2\pi(1+a^{-2})}} \exp \left(-\frac{t^2}{2(1+a^{-2})} \right) \end{aligned}$$

□

Proposición 3.33. *Dado el método bootstrap con anchura pivote (método 5), si la estimación se realiza con un kernel normal, entonces la expresión del MISE_B toma la forma:*

$$\begin{aligned} \text{MISE}_B(h) = \frac{1}{M} \sum_{k=1}^M \left[\frac{1}{2m^2\sqrt{\pi}} \sum_{i=1}^m \sum_{j=1}^m \left[\frac{1}{h} \exp \left(-\frac{(X_{B,j}^k - X_{B,i}^k)^2}{4h^2} \right) - \right. \right. \\ \left. \left. - \frac{2\sqrt{2}}{\sqrt{h^2 + h_{LS}^2}} \exp \left(-\frac{(X_{B,j}^k - X_i)^2}{2(h^2 + h_{LS}^2)} \right) + \frac{1}{h_{LS}} \exp \left(-\frac{(X_j - X_i)^2}{4h_{LS}^2} \right) \right] \right] \end{aligned} \quad (3.42)$$

Demostración. Procediendo de forma análoga a la Proposición 3.32, al desarrollar el cuadrado del integrando obtenemos:

$$\begin{aligned} \left[\frac{1}{mh} \sum_{i=1}^m K \left(\frac{x - X_i^B}{h} \right) - \frac{1}{mg} \sum_{i=1}^m K \left(\frac{x - X_i}{h_{LS}} \right) \right]^2 = \frac{1}{m^2h^2} \sum_{i=1}^m \sum_{j=1}^m K \left(\frac{x - X_i^B}{h} \right) K \left(\frac{x - X_j^B}{h} \right) - \\ - \frac{2}{hm^2h_{LS}} \sum_{i=1}^m \sum_{j=1}^m K \left(\frac{x - X_i^B}{h} \right) K \left(\frac{x - X_j}{h_{LS}} \right) + \frac{1}{m^2h_{LS}^2} \sum_{i=1}^m \sum_{j=1}^m K \left(\frac{x - X_i}{h_{LS}} \right) K \left(\frac{x - X_j}{h_{LS}} \right) \end{aligned}$$

que, al integrar, podemos expresar en términos de convoluciones como:

$$\begin{aligned} \int_{-\infty}^{\infty} \left[\frac{1}{mh} \sum_{i=1}^m K \left(\frac{x - X_i^B}{h} \right) - \frac{1}{mg} \sum_{i=1}^m K \left(\frac{x - X_i}{h_{LS}} \right) \right]^2 dx = \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m \left[\frac{1}{h} K * K \left(\frac{X_j^B - X_i^B}{h} \right) - \right. \\ \left. - \frac{2}{h} K * K_{\frac{h_{LS}}{h}} \left(\frac{X_j - X_i^B}{h} \right) + \frac{1}{h_{LS}} K * K \left(\frac{X_j - X_i}{h_{LS}} \right) \right] \end{aligned}$$

Empleando los cálculos de convoluciones de la Proposición 3.32, e introduciendo el resultado en la expresión del método 6, se obtiene el resultado. □

Proposición 3.34. *Dado el método bootstrap sin anchura pivote (método 4), si la estimación se realiza con un kernel normal, entonces la expresión del MISE_B toma la forma:*

$$\begin{aligned} \text{MISE}_B(h) = \frac{1}{2hm^2\sqrt{2\pi}} \left[\sqrt{2}m + \sum_{i=1}^m \sum_{j=1}^m \left[\frac{(m-1)}{m} \exp \left(-\frac{(X_j - X_i)^2}{8h^2} \right) - \right. \right. \\ \left. \left. - \frac{4}{\sqrt{3}} \exp \left(-\frac{(X_j - X_i)^2}{6h^2} \right) + \sqrt{2} \exp \left(-\frac{(X_j - X_i)^2}{4h^2} \right) \right] \right] \end{aligned} \quad (3.43)$$

Demostración. Partiendo de la expresión del MISE del método 5, desarrollamos el integrando como:

$$\left[\frac{1}{mh} \sum_{i=1}^m K \left(\frac{x - X_i^B}{h} \right) - \frac{1}{mg} \sum_{i=1}^m K \left(\frac{x - X_i}{h} \right) \right]^2 = \frac{1}{m^2h^2} \sum_{i=1}^m \sum_{j=1}^m K \left(\frac{x - X_i^B}{h} \right) K \left(\frac{x - X_j^B}{h} \right) -$$

$$- \frac{2}{m^2 h^2 \sum_{i=1}^m \sum_{j=1}^m K\left(\frac{x-X_i^B}{h}\right) K\left(\frac{x-X_j}{h}\right) + \frac{1}{m^2 h} \sum_{i=1}^m \sum_{j=1}^m K\left(\frac{x-X_i}{h}\right) K\left(\frac{x-X_j}{h}\right)}$$

El último término no depende de la muestra bootstrap, luego sólo tenemos que calcular la esperanza de los otros dos términos. Para el segundo término, podemos observar que sólo un factor depende de la muestra bootstrap cuya función de densidad es $f(x) = \frac{1}{mh} \sum_{i=1}^m K\left(\frac{x-X_i}{h}\right)$, luego tendremos:

$$\begin{aligned} E_B \left[K\left(\frac{x-X_i^B}{h}\right) \right] &= \frac{1}{mh} \sum_{k=1}^m \int_{-\infty}^{\infty} K\left(\frac{x-y}{h}\right) K\left(\frac{y-X_k}{h}\right) dy = \\ &= \frac{1}{m} \sum_{k=1}^m K * K\left(\frac{x-X_k}{h}\right) \end{aligned}$$

y, siguiendo el razonamiento de la Proposición 3.32, la esperanza del primer término será:

$$\begin{aligned} E_B \left[\sum_{i=1}^m \sum_{j=1}^m K\left(\frac{x-X_i^B}{h}\right) K\left(\frac{x-X_j^B}{h}\right) \right] &= \sum_{i=1}^m E_B \left[K^2\left(\frac{x-X_i^B}{h}\right) \right] + (m-1) \sum_{i=1}^m \left[E_B \left[K\left(\frac{x-X_i^B}{h}\right) \right] \right]^2 \\ &= \sum_{k=1}^m K * K^2\left(\frac{x-X_k}{h}\right) + \frac{(m-1)}{m} \sum_{k=1}^m \sum_{r=1}^m K * K\left(\frac{x-X_k}{h}\right) K * K\left(\frac{x-X_r}{h}\right) \end{aligned}$$

Por lo tanto, la expresión final del MISE será:

$$\begin{aligned} \text{MISE}_B(h) &= \frac{1}{m^2 h^2} \left[\sum_{k=1}^m \frac{1}{h^2} \int_{-\infty}^{\infty} K * K^2\left(\frac{x-X_k}{h}\right) dx + \sum_{k=1}^m \sum_{r=1}^m \left[\int_{-\infty}^{\infty} K * K\left(\frac{x-X_k}{h}\right) K * K\left(\frac{x-X_r}{h}\right) dx \right. \right. \\ &\quad \left. \left. - 2 \int_{-\infty}^{\infty} K * K\left(\frac{x-X_k}{h}\right) K\left(\frac{x-X_r}{h}\right) dx + \int_{-\infty}^{\infty} K\left(\frac{x-X_k}{h}\right) K\left(\frac{x-X_r}{h}\right) dx \right] \right] \end{aligned}$$

que podemos expresar en términos de convoluciones como:

$$\begin{aligned} \text{MISE}_B(h) &= \frac{1}{m^2} \left[\sum_{k=1}^m \frac{1}{h^2} \int_{-\infty}^{\infty} K * K^2\left(\frac{x-X_k}{h}\right) dx + \frac{1}{h} \sum_{k=1}^m \sum_{r=1}^m \left[(K * K) * (K * K)\left(\frac{X_r - X_s}{h}\right) - \right. \right. \\ &\quad \left. \left. - 2K * (K * K)\left(\frac{X_r - X_k}{h}\right) + K * K\left(\frac{X_r - X_k}{h}\right) \right] \right] \end{aligned}$$

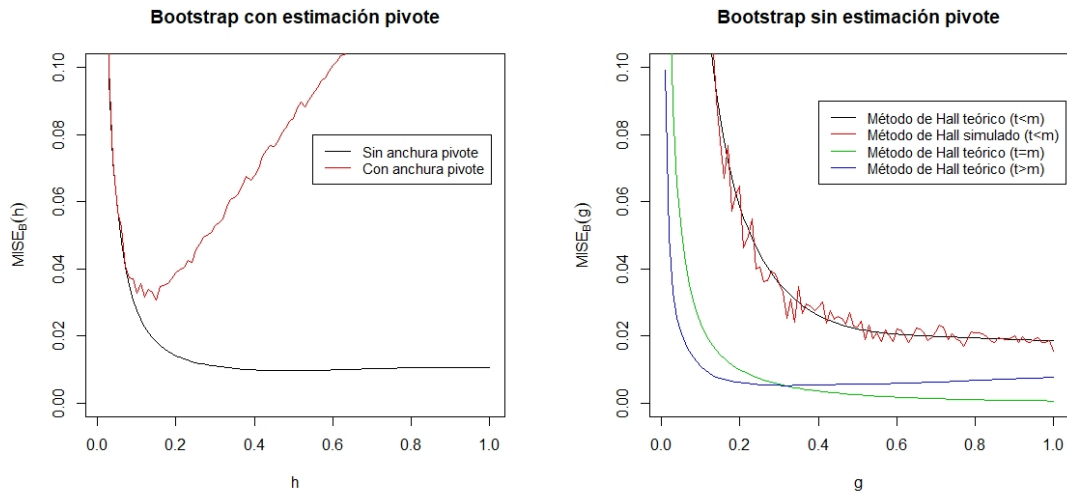
Aprovechando el cálculo de algunas convoluciones de la Proposición 3.32, sólo tendremos que calcular las convoluciones:

$$\begin{aligned} K * (K * K)(t) &= \frac{1}{2\pi} \frac{1}{\sqrt{2}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) \exp\left(-\frac{(x-t)^2}{4}\right) dx = \\ &= \frac{1}{2\pi} \frac{1}{\sqrt{2}} \exp\left(-\frac{t^2}{6}\right) \int_{-\infty}^{\infty} \exp\left(-\left(\sqrt{\frac{3}{4}}x - \frac{t}{\sqrt{12}}\right)^2\right) dx = \frac{1}{\sqrt{6\pi}} \exp\left(-\frac{t^2}{6}\right) \\ (K * K) * (K * K)(t) &= \frac{1}{4\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{4}\right) \exp\left(-\frac{(x-t)^2}{4}\right) dx = \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{4\pi} \exp\left(-\frac{t^2}{4}\right) \int_{-\infty}^{\infty} \exp\left(-\left(\frac{x}{\sqrt{2}} - \frac{t}{2\sqrt{2}}\right)^2\right) dx = \frac{1}{2\sqrt{2}\pi} \exp\left(-\frac{t^2}{8}\right) \\
K * K^2(t) &= \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{-\infty}^{\infty} \exp\left(-\frac{x^2}{2}\right) \exp\left(-(x-t)^2\right) dx = \\
&= \frac{1}{(2\pi)^{\frac{3}{2}}} \exp\left(-\frac{t^2}{3}\right) \int_{-\infty}^{\infty} \exp\left(-\left(\sqrt{\frac{3}{2}}x - \frac{\sqrt{2}t}{\sqrt{3}}\right)^2\right) dx = \frac{1}{2\pi\sqrt{3}} \exp\left(-\frac{t^2}{3}\right)
\end{aligned}$$

Introduciendo estas expresiones en el MISE, se concluye el resultado de la proposición. \square

De acuerdo a los resultados (3.41) y (3.43), podemos evaluar los métodos bootstrap 4 y 7 sin necesidad de generar muestras bootstraps, lo que agiliza el tiempo de computación. Sin embargo, para aplicar el método 6 sí es necesario generar las muestras bootstrap, y además se debe hacer un total de M -veces. Como seleccionamos una muestra aleatoria de tamaño $m = 100$, para agilizar el cálculo tomaremos como $M = 50$. Otro resultado importante que podemos analizar a priori es la dependencia de las estimaciones del MISE con la amplitud de ventana. Por las expresiones (3.41) y (3.43), podemos observar que $\text{MISE}_B \rightarrow 0$ cuando $h, g \rightarrow \infty$. Puesto que el MISE es una cantidad positiva, esto quiere decir que no se tiene un mínimo global para dichas funciones, sino que tendremos que realizar la búsqueda de mínimos locales.



(a) MISE bootstrap con estimación pivote (métodos 4 y 5).

(b) MISE bootstrap sin estimación pivote (métodos 6 y 7).

Figura 3.9: Estimación del MISE bootstrap para una muestra aleatoria simple de 100 cuásares mediante el uso de kernel normal.

En la Figura 3.9 (a) podemos observar como el método 6 proporciona un claro mínimo global, mientras que el método 7 nos proporciona en el intervalo representado $(0, 1)$ un mínimo local. Los resultados que se muestran en la Tabla 3.6 muestran cómo el método sin anchura pivote proporciona una amplitud bastante mayor que el resto de métodos, mientras que el método con anchura pivote corrige el valor pequeño que se obtendría de aplicar exclusivamente la validación cruzada por mínimos cuadrados y nos proporciona un valor más cercano al proporcionado por otras aproximaciones.

Tabla 3.6: Selección de anchuras de ventana por procedimientos bootstrap así como por elección en el caso normal y por sobre suavización. Se ha utilizado un kernel normal y se ha fijado una precisión de 10^{-2} para todos los cálculos.

Método de selección	Kernel normal
Elección referida a la normal	0.25
Elección por sobre suavización	0.27
Validación Cruzada Mínimos Cuadrados	0.08
Método Bootstrap sin anchura pivote	0.48
Método Bootstrap con anchura pivote	0.15
Método Bootstrap sin estimación pivote ($t > m$)	0.32
Método Bootstrap sin estimación pivote ($t \leq m$)	-

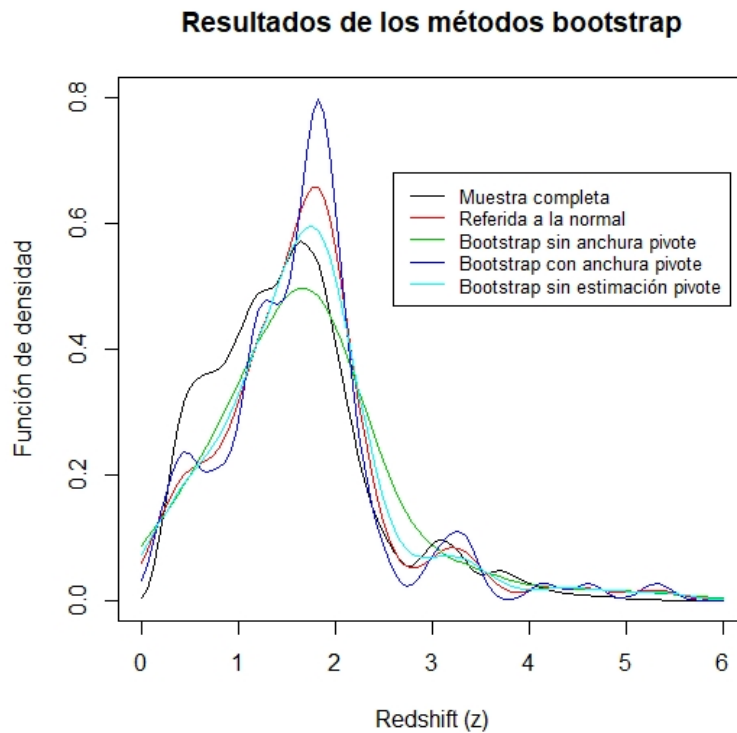


Figura 3.10: Estimación con kernel normal para 100 cuásares de la muestra de la campaña SDSS bajo los diferentes métodos bootstrap.

Al analizar la [Figura 3.9](#) (b) observamos que el método 7 no es aplicable al ejemplo que estamos considerando. De acuerdo a dicho método propuesto por Hall, deberíamos encontrar un mínimo local cercano al punto $t^{-\frac{1}{5}}$, sin embargo, la función presenta un comportamiento monótono en la región. Una simulación directa del método confirma la inexistencia de dicho mínimo (teniendo en cuenta las fluctuaciones ocasionadas de estimar la esperanza bootstrap mediante una repetición del método). El mismo problema se

presenta si consideramos el método 6. Sin embargo, es importante destacar que al reconsiderar el método de Hall, pero ahora tomando una muestra bootstrap de mayor tamaño $t > m$, entonces el método nos proporciona una amplitud de ventana que incluso mejora la del método bootstrap sin anchura pivote. En los artículos publicados respecto a este método no se hace mención a la problemática aquí presentada, aunque muchos muestran que el método propuesto por Hall sólo proporciona buenos resultados en ciertos tipos de funciones de densidad[2].

La [Figura 3.10](#) muestra los resultados de la estimación con las anchuras de ventana obtenidas en cada método. La estimación proporcionada por el método bootstrap con anchura pivote es la que más dependencia presenta de la muestra aleatoria simple (ya que usa una anchura de ventana más pequeña), y nos indica cómo nuestra muestra aleatoria presenta una importante fracción de valores en la cola derecha. Al tener una muestra más pequeña de lo que hemos trabajado con anterioridad, es posible que esta sea una de las razones por las cuales hemos tenido problemas para implementar el método de Hall.

4. Otras mejoras en la estimación

4.1. Suavidad adaptada

Todos los procedimientos que hemos analizado hasta ahora para la estimación de la función densidad buscaban una aproximación a la anchura de ventana óptima deducida en (2.16), suponiendo que es fija en todos los puntos. Ahora nos centraremos en procedimientos de estimación que permiten la variación de la anchura con la posición. En concreto, analizaremos dos métodos clásicos: el método del k -vecino más cercano y el método kernel adaptativo.

4.1.1. Método del k -vecino más cercano

Este método se basa en considerar en cada punto $x \in \mathbb{R}$ una ventana que incluya los k puntos más cercanos de la muestra aleatoria simple (X_1, X_2, \dots, X_m) .

Definición 4.1. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) , un punto $x \in \mathbb{R}$ y $k \in \{2, 3, \dots, m\}$ se define la distancia entre el punto x y el k -ésimo vecino más cercano $d_k(x)$ como:

$$d_k(x) := \min\{|x - X_{k(x)}| \mid (\#\{x - X_i \mid |x - X_i| \leq |x - X_{k(x)}|\} \geq k)\} \quad (4.1)$$

El valor muestral $X_{k(x)}$ recibe el nombre de k -ésimo vecino más cercano a x .

Definición 4.2. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) y $k \in \{2, 3, \dots, m\}$, se define el estimador del k -vecino más cercano como la función:

$$\hat{f}_{kv}(x) = \frac{k}{2md_k(x)} \quad (4.2)$$

La idea que subyace en la definición (4.2) es la de la frecuencia muestral. Si la función de densidad viene dada por \hat{f}_{kv} , al generar m valores aleatorios de la variable, es de esperar que en el intervalo $[x - d_k(x), x + d_k(x)]$ nos encontremos con k valores, es decir $k \sim m2d_k(x)\hat{f}_{kv}(x)$.

Analizar las propiedades del estimador \hat{f}_{kv} (4.2) pasa por estudiar la función $d_k(x)$. Para ello, puesto que dicha función depende de las posiciones de los valores muestrales, trabajaremos con la muestra ordenada dada por $(X_{(1)}, X_{(2)}, \dots, X_{(m)})$.

Proposición 4.3. Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) y $k \in \{2, 3, \dots, m\}$, la función $d_k(x)$ es continua en todo \mathbb{R} y presenta problemas de derivabilidad en los puntos de la forma $\frac{X_{(j)} + X_{(j+k-1)}}{2}$ $\forall j \in \{1, 2, \dots, m\}$.

Demostración. Fijemos $k \in \{2, 3, \dots, m\}$. Cuando el punto $x \leq X_{(1)}$ la función $d_k(x)$ toma la forma:

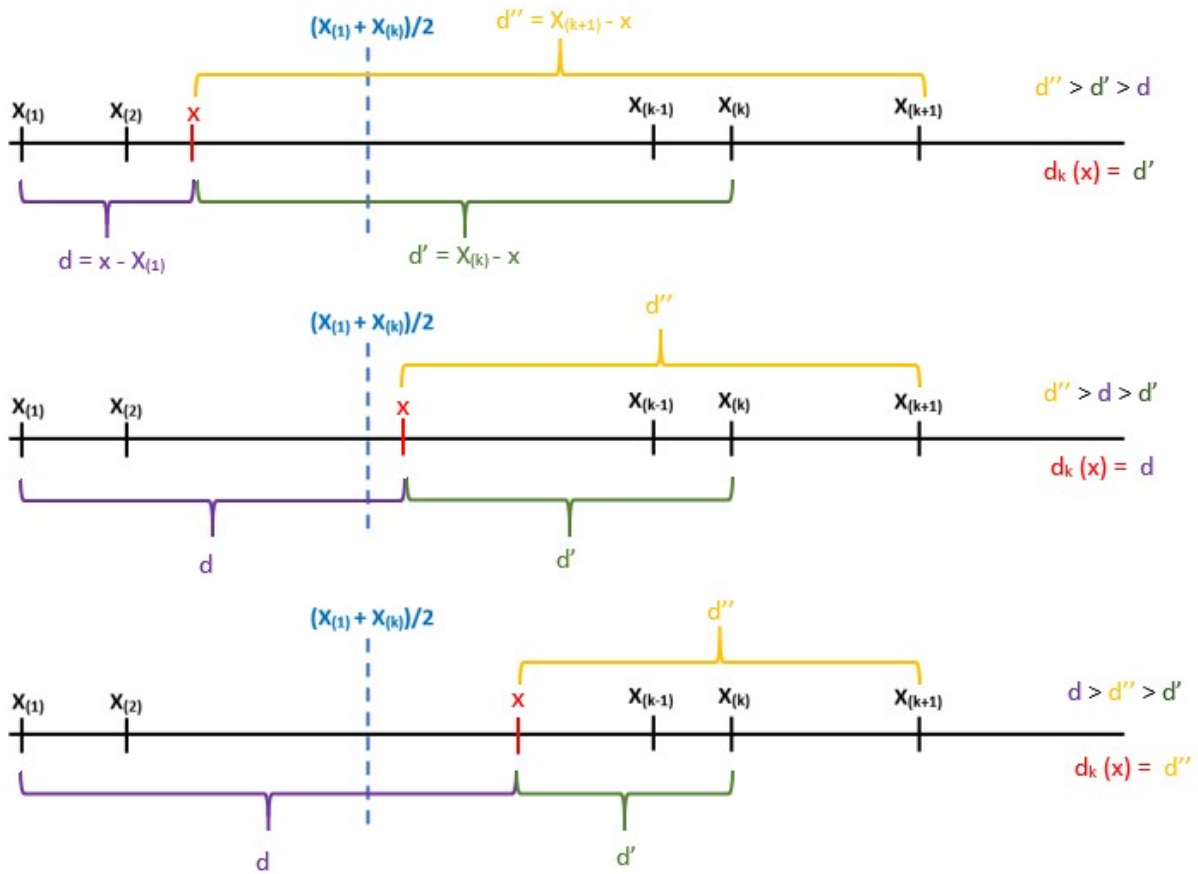
$$d_k(x) = X_{(k)} - x$$

que es claramente continua. De forma análoga, si $x \geq X_{(m)}$ la función $d_k(x)$ será:

$$d_k(x) = x - X_{(m-k+1)}$$

que también es continua. Estas dos situaciones ya nos muestran que la función $d_k(x)$ es de la forma $x - X_k$ o $X_k - x$, luego su derivada puede ser 1 o -1: en los puntos $x \leq X_{(1)}$ la derivada es -1 mientras que en los puntos $x \geq X_{(m)}$ es 1.

Luego para analizar las posibles discontinuidades consideramos puntos x en el intervalo $(X_{(1)}, X_{(m)})$. Consideremos las siguientes situaciones:



Supongamos ahora que el punto $x \in (X_{(1)}, X_{(k)})$. En dicho intervalo tenemos exactamente k valores muestrales. De acuerdo al esquema planteado, cuando $x \in \left(X_{(1)}, \frac{X_{(1)} + X_{(k)}}{2}\right)$ la función $d_k(x)$ será de la forma:

$$d_k(x) = X_{(k)} - x$$

pero cuando x toma un valor superior a $\frac{X_{(1)}+X_{(k)}}{2}$ la función $d_k(x)$ toma la forma:

$$d_k(x) = x - X_{(1)}$$

hasta que se verifique que $x - X_{(1)} > X_{(k+1)} - x$, es decir, a partir del punto $\frac{X_{(1)}+X_{(k+1)}}{2}$ la distancia $d_k(x)$ pasará a tomar la forma:

$$d_k(x) = X_{(k+1)} - x$$

en donde se presentará de nuevo la misma situación pero en el intervalo $(X_{(2)}, X_{(k+1)})$.

Luego, en el caso más general en el que $x \in (X_{(j)}, X_{(j+k-1)})$. Tenemos que distinguir una serie de posibilidades:

1. Si $\frac{X_{(j-1)}+X_{(j+k-1)}}{2} > X_{(j)}$ y $\frac{X_{(j)}+X_{(j+k)}}{2} < X_{(j+k-1)}$ entonces:

$$d_k(x) = \begin{cases} x - X_{(j-1)} & \text{si } x \in \left(X_{(j)}, \frac{X_{(j-1)}+X_{(j+k-1)}}{2}\right) \\ X_{(j+k-1)} - x & \text{si } x \in \left[\frac{X_{(j-1)}+X_{(j+k-1)}}{2}, \frac{X_{(j)}+X_{(j+k-1)}}{2}\right) \\ x - X_{(j)} & \text{si } x \in \left[\frac{X_{(j)}+X_{(j+k-1)}}{2}, \frac{X_{(j)}+X_{(j+k)}}{2}\right) \\ X_{(j+k)} - x & \text{si } x \in \left[\frac{X_{(j)}+X_{(j+k)}}{2}, X_{(j+k-1)}\right) \end{cases}$$

2. Si $\frac{X_{(j-1)}+X_{(j+k-1)}}{2} > X_{(j)}$ y $\frac{X_{(j)}+X_{(j+k)}}{2} > X_{(j+k-1)}$ entonces:

$$d_k(x) = \begin{cases} x - X_{(j-1)} & \text{si } x \in \left(X_{(j)}, \frac{X_{(j-1)}+X_{(j+k-1)}}{2}\right) \\ X_{(j+k-1)} - x & \text{si } x \in \left[\frac{X_{(j-1)}+X_{(j+k-1)}}{2}, \frac{X_{(j)}+X_{(j+k-1)}}{2}\right) \\ x - X_{(j)} & \text{si } x \in \left[\frac{X_{(j)}+X_{(j+k-1)}}{2}, X_{(j+k-1)}\right) \end{cases}$$

3. Si $\frac{X_{(j-1)}+X_{(j+k-1)}}{2} < X_{(j)}$ y $\frac{X_{(j)}+X_{(j+k)}}{2} < X_{(j+k-1)}$ entonces:

$$d_k(x) = \begin{cases} X_{(j+k-1)} - x & \text{si } x \in \left(X_{(j)}, \frac{X_{(j)}+X_{(j+k-1)}}{2}\right) \\ x - X_{(j)} & \text{si } x \in \left[\frac{X_{(j)}+X_{(j+k-1)}}{2}, \frac{X_{(j)}+X_{(j+k)}}{2}\right) \\ X_{(j+k)} - x & \text{si } x \in \left[\frac{X_{(j)}+X_{(j+k)}}{2}, X_{(j+k-1)}\right) \end{cases}$$

4. Si $\frac{X_{(j-1)}+X_{(j+k-1)}}{2} < X_{(j)}$ y $\frac{X_{(j)}+X_{(j+k)}}{2} > X_{(j+k-1)}$ entonces:

$$d_k(x) = \begin{cases} X_{(j+k-1)} - x & \text{si } x \in \left(X_{(j)}, \frac{X_{(j)}+X_{(j+k-1)}}{2}\right) \\ x - X_{(j)} & \text{si } x \in \left[\frac{X_{(j)}+X_{(j+k-1)}}{2}, X_{(j+k-1)}\right) \end{cases}$$

En los cuatro casos, d_k es continua en todos sus puntos. Sin embargo, en general, no es derivable en los puntos $\frac{X_{(j)}+X_{(j+k-1)}}{2}$ y, en algunas ocasiones, en los puntos $\frac{X_{(j)}+X_{(j+k)}}{2}$. \square

Por otro lado, no sólo la función d_k nos limita el uso del estimador para el cálculo de las derivadas de la función de densidad, sino que en verdad el estimador del k -vecino más cercano no nos proporciona una función de densidad.

Proposición 4.4. *Dada una muestra aleatoria simple (X_1, X_2, \dots, X_m) y $k \in \{2, 3, \dots, m\}$. El estimador del k -vecino más cercano no es una función de densidad.*

Demostración. Las funciones de densidad cumplen las propiedades a) y b) de (1.2). El estimador del k -vecino más cercano viene dado por:

$$\hat{f}_{kv}(x) = \frac{k}{2md_k(x)}$$

y por la Definición 4.1 $d_k(x) > 0 \forall x \in \mathbb{R}$ por lo que el estimador cumple la propiedad a).

Estudiemus si cumple la propiedad b). La integral la podemos descomponer en 3 intervalos $(-\infty, X_{(1)})$, $(X_{(1)}, X_{(m)})$ y $(X_{(m)}, \infty)$, de manera que:

$$\int_{-\infty}^{\infty} \hat{f}_{kv}(x) dx = \int_{-\infty}^{X_{(1)}} \frac{k}{2md_k(x)} dx + \int_{X_{(1)}}^{X_{(m)}} \frac{k}{2md_k(x)} dx + \int_{X_{(m)}}^{\infty} \frac{k}{2md_k(x)} dx$$

Por la Proposición 4.3, en el intervalo $(X_{(1)}, X_{(m)})$ la función $\frac{k}{2md_k}$ será continua y estrictamente positiva, luego tomará un valor positivo y finito A :

$$\int_{X_{(1)}}^{X_{(m)}} \frac{k}{2md_k(x)} dx = A$$

Por otro lado, en el intervalo $(-\infty, X_{(1)})$ tenemos $d_k(x) = X_{(k)} - x$ y en $(X_{(m)}, \infty)$ tenemos $d_k(x) = x - X_{(m+1-k)}$ de manera que:

$$\begin{aligned} \int_{-\infty}^{X_{(1)}} \frac{k}{2md_k(x)} dx &= \int_{-\infty}^{X_{(1)}} \frac{k}{2m(X_{(k)} - x)} dx = \frac{k}{2m} \ln(X_{(k)} - x) \Big|_{X_{(1)}}^{-\infty} = \infty \\ \int_{X_{(m)}}^{\infty} \frac{k}{2md_k(x)} dx &= \int_{X_{(m)}}^{\infty} \frac{k}{2m(x - X_{(m+1-k)})} dx = \frac{k}{2m} \ln(x - X_{(m+1-k)}) \Big|_{X_{(m)}}^{\infty} = \infty \end{aligned}$$

En definitiva:

$$\int_{-\infty}^{\infty} \hat{f}_{kv}(x) dx = A + \frac{k}{2m} \ln(X_{(k)} - x) \Big|_{X_{(1)}}^{-\infty} + \frac{k}{2m} \ln(x - X_{(m+1-k)}) \Big|_{X_{(m)}}^{\infty} = \infty$$

luego no se cumple la propiedad b) y, por tanto, no es función de densidad. \square

Como consecuencia de este resultado, el estimador del k -vecino más cercano se utiliza restringido al intervalo muestral $[X_{(1)}, X_{(m)}]$ y para garantizar que cumpla las condiciones de función de densidad se normaliza:

$$\tilde{f}_{kv} = \frac{\hat{f}_{kv}(x)}{A} I_{[X_{(1)}, X_{(m)}]}(x) \quad (4.3)$$

De esta manera, solucionamos los problemas de divergencia que se presentarían en los extremos, pero perdemos una gran cantidad de información en la estimación.

El estimador (4.2) es análogo al estimador naïve (2.5), pero el papel de h ahora lo juega la distancia al k -ésimo vecino más cercano. De hecho, observamos que:

$$\hat{f}(x) = \frac{k}{2md_k(x)} = \frac{1}{2md_k(x)} \sum_{i=1}^m I_{[x-d_k(x), x+d_k(x)]}(X_i) = \frac{1}{2md_k(x)} \sum_{i=1}^m I_{[-1,1]}\left(\frac{x-X_i}{d_k(x)}\right) \quad (4.4)$$

Por lo tanto, es intuitivo plantearse una estimación más suave por medio del uso de funciones kernel de la misma manera que se generalizó en la Sección 2.4 el estimador naïve.

Definición 4.5. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) , $k \in \{2, 3, \dots, m\}$ y una función kernel K . Se define el estimador del k -vecino más cercano con kernel K como la función:

$$\hat{f}_{kv,K}(x) = \frac{1}{md_k(x)} \sum_{i=1}^m K\left(\frac{x-X_i}{d_k(x)}\right) \quad (4.5)$$

El estimador anteriormente definido es un método de estimación de densidad con anchura de ventana variable $d_k(x)$. Además la suavidad del método se puede variar, no sólo por la elección de diferentes funciones kernel, sino también bajo la simple elección del valor k que se tome para el método del vecino más cercano.

Como el parámetro k es libre, cabe preguntarse si existe algún valor para el que se optimice el método. Para ello, procedemos de manera análoga a las anteriores subsecciones, es decir, buscamos minimizar el MISE, para lo que tendremos que calcular en primer lugar el sesgo y la varianza del estimador.

Proposición 4.6. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) , $k \in \{2, 3, \dots, m\}$, una función kernel K y $\hat{f}_{kv,K}$ el estimador del k -vecino más cercano con kernel K . El sesgo y la varianza en el punto x verifican:

$$\text{sesgo}\left(\hat{f}_{kv,K}(x)\right) = \frac{1}{8} \frac{f''(x)}{f(x)} \left(\frac{k}{m}\right)^2 \int_{-\infty}^{\infty} y^2 K(y) dy + o\left(\left(\frac{k}{m}\right)^2\right) \quad (4.6)$$

$$\text{Var}\left(\hat{f}_{kv,K}(x)\right) = \frac{2}{k} (f(x))^2 \int_{-\infty}^{\infty} K^2(y) dy + o\left(\frac{1}{k}\right) \quad (4.7)$$

Demostración. Los cálculos requeridos para obtener los resultados son bastante extensos y se encuentran detallados en el trabajo presentado por Mack y Rosenblatt en 1979[11]. \square

Si comparamos estos dos desarrollos con los realizados para la estimación kernel básica (2.12) y (2.13), observamos que el papel de h lo juega en cierta medida el factor $\frac{k}{m}$, hecho que se deduce del orden de los

desarrollos. Es más, una comparación directa entre estas cuatro expresiones conduce inequívocamente a la relación:

$$\frac{2}{k} f(x) = \frac{1}{mh} \Rightarrow \frac{k}{m} = 2hf(x) \quad (4.8)$$

pero añade una dependencia puntual, resultado coherente puesto que estamos considerando un método con anchura variable. La expresión (4.8) implica que el valor óptimo de k , es decir, el orden del vecino que vamos a elegir va a depender de cada punto $x \in \mathbb{R}$. Sin embargo, también muestra la imposibilidad de determinar el óptimo valor de k , puesto que en el caso del método del k -ésimo vecino es necesario conocer directamente la función de densidad.

Corolario 4.7. Sean una muestra aleatoria simple (X_1, X_2, \dots, X_m) , $k \in \{2, 3, \dots, m\}$, una función kernel K y $\hat{f}_{k,K}$ el estimador del k -vecino más cercano con kernel K . El valor \tilde{k} que minimiza la parte asintótica del MISE del estimador viene dado por:

$$\tilde{k} = \left[\frac{32 \int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 \int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx} \right]^{\frac{1}{5}} m^{-\frac{4}{5}} \quad (4.9)$$

Demostración. De acuerdo a los resultados de la Proposición 4.6, la parte asintótica del MISE viene dado por:

$$\text{MISE}(\hat{f}_{k,K}) \sim \frac{1}{64} \frac{k^4}{m^4} V_K^2 \int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx + \frac{2}{k} \int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy$$

Calculamos los extremos relativos. Para ello derivamos la anterior expresión respecto del parámetro que deseamos optimizar, k :

$$\frac{d}{dk} \text{MISE}(\hat{f}_{k,K}) = \frac{1}{16} \frac{k^3}{m^4} V_K^2 \int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx - \frac{2}{k^2} \int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy$$

e igualando a 0 obtenemos:

$$\frac{1}{16} \frac{\tilde{k}^3}{m^4} V_K^2 \int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx - \frac{2}{\tilde{k}^2} \int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy = 0 \Rightarrow$$

$$\tilde{k} = \left[\frac{32 \int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy}{V_K^2 \int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx} \right]^{\frac{1}{5}} m^{-\frac{4}{5}}$$

Calculamos la derivada segunda para comprobar que dicho valor minimiza la parte asintótica del MISE:

$$\frac{d^2}{dk^2} \text{MISE}(\hat{f}_{k,K}) = \frac{3}{16} \frac{k^2}{m^4} V_K^2 \int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx + \frac{4}{k^3} \int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy$$

y evaluando en \tilde{k} :

$$\frac{3}{16} \frac{\tilde{k}^2}{m^4} V_K^2 \int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx + \frac{4}{\tilde{k}^3} \int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy$$

$$= \frac{9}{2} V_K^{\frac{6}{5}} \left[\int_{-\infty}^{\infty} f^2(x) dx \int_{-\infty}^{\infty} K^2(y) dy \right]^{\frac{2}{5}} \left[\int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx \right]^{\frac{3}{5}} \geq 0$$

□

El resultado (4.9) muestra como el método complica considerablemente la estimación óptima. En los métodos de anchura fija, el problema residía en estimar el funcional $R(f) \equiv \int_{-\infty}^{\infty} (f''(x))^2 dx$, pero ahora tenemos que estimar dos funcionales: $\int_{-\infty}^{\infty} f^2(x) dx$ y $\int_{-\infty}^{\infty} \left(\frac{f''(x)}{f(x)} \right)^2 dx$.

4.1.2. Método kernel adaptado

La idea básica de este método es combinar la versatilidad del método del k-vecino más cercano con la suavidad que se obtiene del método kernel. Al introducir una anchura de ventana variable, en los puntos en los que se presenta una menor probabilidad será más conveniente utilizar una anchura mayor. A continuación, desarrollaremos el método presentado por Silverman en 1986[17].

Método 8. Este método se basa en construir en un primer paso una estimación inicial de la función de densidad para tener una idea de las zonas de mayor y menor densidad; establecer entonces una secuencia de anchuras de ventana que tengan en cuenta la distribución inicial y con ellos construir el estimador adaptado.

- El primer paso consiste en seleccionar un estimador inicial \hat{f}_p , denominado *estimador pivote*, que satisfaga $\hat{f}_p(X_i) > 0$ para todo $X_i \in (X_1, X_2, \dots, X_m)$. Esta estimación inicial se puede realizar tanto por el método del k-vecino más cercano como por el método kernel básico, en el que la anchura de ventana se elige entre cualquier método, pero frecuentemente se usa el referido a la distribución normal por simplicidad en el cálculo y así evitar el tiempo de trabajo computacional que requieren los métodos plug-in o los métodos de validación cruzada.
- El segundo paso determina las anchuras de banda para cada posición muestral X_i . La forma de trabajo consiste en considerar dichas anchuras (h_1, h_2, \dots, h_m) como un reescalamiento en cada posición de una anchura genérica h por un factor λ_i , de manera que $h_i = \lambda_i h$ para todo $i \in \{1, 2, \dots, m\}$. Silverman propone calcular dichos factores de escala como[17]:

$$\lambda_i = \left(\frac{\hat{f}_p(X_i)}{\mu_p} \right)^{-\alpha} \quad \text{con} \quad \mu_g = \exp \left(\frac{1}{m} \sum_{i=1}^m \log [\hat{f}_p(X_i)] \right) \quad y \quad \alpha \in [0, 1] \quad (4.10)$$

Observamos que por imposición en el paso uno, los factores de escala cumplen $\lambda_i > 0$. Además, al introducir el factor μ_g , la media geométrica, garantizamos entonces que:

$$\begin{aligned} \frac{1}{m} \sum_{i=1}^m \log(\lambda_i) &= \frac{1}{m} \sum_{i=1}^m \log \left(\left(\frac{\hat{f}_p(X_i)}{\mu_p} \right)^{-\alpha} \right) = -\alpha \left[\frac{1}{m} \sum_{i=1}^m \log [\hat{f}_p(X_i)] - \frac{1}{m} \sum_{i=1}^m \log(\mu_g) \right] = \\ &= -\alpha \left[\frac{1}{m} \sum_{i=1}^m \log [\hat{f}_p(X_i)] - \log(\mu_g) \right] = 0 \end{aligned}$$

es decir, la media geométrica de los λ_i es 1. Por otro lado, el exponente $\alpha \in [0, 1]$ nos garantiza que para las zonas de mayor peso la anchura de ventana sea menor que en las otras zonas. Dicho exponente α recibe el nombre de *parámetro de sensibilidad*: al tomar $\alpha = 0$ recuperamos el método clásico de estimación kernel con anchura fija, mientras que fijando $\alpha = 1$ tendremos entonces una selección de anchuras muy dependiente de la estimación inicial pivote.

- Por último, con la colección de factores de escala $\{\lambda_i\}_{i=1}^m$ se estima la función de densidad, mediante el uso de una función kernel K como:

$$\hat{f}_{K,adp}(x) = \frac{1}{mh} \sum_{i=1}^m \frac{1}{\lambda_i} K\left(\frac{x - X_i}{\lambda_i h}\right) \quad (4.11)$$

Las ventajas que presenta este método frente a todos los que hemos estudiado hasta ahora son bastantes numerosas. En primer lugar, mantiene un parámetro h fijo que se puede utilizar para regular en cierta medida la suavidad de la función de densidad junto con la elección de la función Kernel, los únicos parámetros disponibles en los métodos con anchura de ventana fija. Por otro lado, presenta la ventaja de introducir un método de anchura variable, en el que las anchuras se seleccionan de acuerdo a una estimación inicial de la densidad, y no de directamente de la muestra como realiza el método del k-vecino más cercano.

Puesto que ahora disponemos de tres parámetros para la suavización de la función de densidad, el kernel K , la anchura h y el parámetro de sensibilidad α , centraremos nuestros esfuerzos en la selección óptima del último parámetro, ya que los dos primeros se han estudiado con anterioridad. Para ello, calcularemos el sesgo del estimador (4.11).

Proposición 4.8. Sean X una variable aleatoria con función de densidad $f \in \mathcal{C}^4$, (X_1, X_2, \dots, X_m) una muestra aleatoria de dicha variable, una función kernel $K \in \mathcal{C}^4$ y $\hat{f}_{K,adp}$ el estimador obtenido de aplicar el Método 8 con $\alpha = \frac{1}{2}$ y una estimación pivote sin errores $\hat{f}_p \approx f$. Entonces el sesgo del estimador verifica:

$$\text{sesgo}\left(\hat{f}_{K,adp}(x)\right) = A(x) h^4 \mu_g^2 + o(h^4) \quad (4.12)$$

Demostración. De forma general, podemos observar que la esperanza del estimador del Método 7 viene dada por:

$$\mathbb{E}\left(\hat{f}_{K,adp}(x)\right) = \frac{1}{m} \sum_{i=1}^m \mathbb{E}\left(\frac{1}{\lambda_i} K\left(\frac{x - X_i}{\lambda_i h}\right)\right) = \frac{1}{h} \int_{-\infty}^{\infty} \left(\frac{\hat{f}_p(y)}{\mu_g}\right)^\alpha K\left(\frac{x - y}{h} \left(\frac{\hat{f}_p(y)}{\mu_g}\right)^\alpha\right) f(y) dy$$

y como por hipótesis la estimación pivote inicial se ha calculado sin errores, es decir, $\hat{f}_p(x) \approx f(x)$, realizando el cambio de variable $x = y + zh\mu_g^\alpha$, pasamos a trabajar con:

$$\mathbb{E}\left(\hat{f}_{K,adp}(x)\right) = \int_{-\infty}^{\infty} f^{\alpha+1}(x - zh\mu_g^\alpha) K(zf^\alpha(x - zh\mu_g^\alpha)) dz$$

Ahora particularizamos para $\alpha = \frac{1}{2}$. Como por hipótesis $f \in \mathcal{C}^4$, calculamos los desarrollos en serie

cuando $h \rightarrow 0$, en primer lugar del término:

$$\begin{aligned} f^{\frac{3}{2}}\left(x - zh\mu_g^{\frac{1}{2}}\right) &= f^{\frac{3}{2}}(x) - \frac{3}{2}f^{\frac{1}{2}}(x)f'(x)zh\mu_g^{\frac{1}{2}} + \frac{3}{4}\left[\frac{1}{2}\frac{(f'(x))^2}{f^{\frac{1}{2}}(x)} + f^{\frac{1}{2}}(x)f''(x)\right]z^2h^2\mu_g \\ &\quad - \frac{1}{4}\left[\frac{3}{2}\frac{f'(x)f''(x)}{f^{\frac{1}{2}}(x)} - \frac{1}{4}\frac{(f'(x))^3}{f^{\frac{3}{2}}(x)} + f^{\frac{1}{2}}(x)f'''(x)\right]z^3h^3\mu_g^{\frac{3}{2}} \\ &+ \frac{1}{16}\left[\frac{3}{2}\frac{(f''(x))^2}{f^{\frac{1}{2}}(x)} + 2\frac{f'(x)f'''(x)}{f^{\frac{1}{2}}(x)} - \frac{3}{2}\frac{(f'(x))^2f''(x)}{f^{\frac{3}{2}}(x)} + \frac{3}{8}\frac{(f'(x))^4}{f^{\frac{5}{2}}(x)} + f^{\frac{1}{2}}(x)f^{(iv)}(x)\right]z^4h^4\mu_g^2 + o(h^4) \end{aligned}$$

y en segundo lugar del término:

$$\begin{aligned} K\left(zf^{\frac{1}{2}}\left(x - zh\mu_g^{\frac{1}{2}}\right)\right) &= K\left(zf^{\frac{1}{2}}(x)\right) - \frac{1}{2}\frac{f'(x)}{f^{\frac{1}{2}}(x)}K'\left(zf^{\frac{1}{2}}(x)\right)z^2h\mu_g^{\frac{1}{2}} + \frac{1}{4}\left[\frac{f''(x)}{f^{\frac{1}{2}}(x)}K'\left(zf^{\frac{1}{2}}(x)\right) - \right. \\ &\quad \left.\frac{1}{2}\frac{(f'(x))^2}{f^{\frac{3}{2}}(x)}K'\left(zf^{\frac{1}{2}}(x)\right) + \frac{1}{2}\frac{(f'(x))^2}{f(x)}K''\left(zf^{\frac{1}{2}}(x)\right)z\right]z^3h^2\mu_g - \frac{z^4h^3\mu_g^{\frac{3}{2}}}{12}\left[\left(\frac{f'''(x)}{f^{\frac{1}{2}}(x)} - \frac{3}{2}\frac{f'(x)f''(x)}{f^{\frac{3}{2}}(x)} + \right.\right. \\ &\quad \left.\left.\frac{3}{4}\frac{(f'(x))^3}{f^{\frac{5}{2}}(x)}\right)K'\left(zf^{\frac{1}{2}}(x)\right) + \left(\frac{3}{2}\frac{f'(x)f''(x)}{f(x)} - \frac{3}{4}\frac{(f'(x))^3}{f^2(x)}\right)zK''\left(zf^{\frac{1}{2}}(x)\right) + \frac{1}{4}\frac{(f'(x))^3}{f^{\frac{3}{2}}(x)}z^2K'''(zf^{\frac{1}{2}}(x))\right] \\ &\quad + \frac{1}{48}\left[\left(\frac{f^{(iv)}(x)}{f^{\frac{1}{2}}(x)} - 2\frac{f'(x)f'''(x)}{f^{\frac{3}{2}}(x)} - \frac{3}{2}\frac{(f''(x))^2}{f^{\frac{3}{2}}(x)} + \frac{9}{2}\frac{(f'(x))^2f''(x)}{f^{\frac{5}{2}}(x)} - \frac{15}{8}\frac{(f'(x))^4}{f^{\frac{7}{2}}(x)}\right)K'\left(zf^{\frac{1}{2}}(x)\right) \right. \\ &\quad \left. + \left(\frac{3}{2}\frac{(f''(x))^2}{f(x)} + 2\frac{f'(x)f'''(x)}{f(x)} - \frac{9}{2}\frac{(f'(x))^2f''(x)}{f^2(x)} + \frac{9}{4}\frac{(f'(x))^4}{f^3(x)}\right)zK''\left(zf^{\frac{1}{2}}(x)\right) \right. \\ &\quad \left. + \left(\frac{3}{2}\frac{(f'(x))^2f''(x)}{f^{\frac{3}{2}}(x)} - \frac{3}{4}\frac{(f'(x))^4}{f^{\frac{5}{2}}(x)}\right)z^2K'''(zf^{\frac{1}{2}}(x)) + \frac{1}{8}\frac{(f'(x))^4}{f^2(x)}z^3K^{(iv)}(zf^{\frac{1}{2}}(x))\right]z^5h^4\mu_g^2 + o(h^4) \end{aligned}$$

Multiplicando ambas expresiones y quedándonos con los términos hasta un orden $o(h^4)$, los agrupamos por órdenes:

$$\begin{aligned} o(h^0) &: f^{\frac{3}{2}}(x)K\left(zf^{\frac{1}{2}}(x)\right) \\ o(h^1) &: -\frac{3}{2}f^{\frac{1}{2}}(x)f'(x)zh\mu_g^{\frac{1}{2}}K\left(zf^{\frac{1}{2}}(x)\right) - \frac{1}{2}f'(x)f(x)K'\left(zf^{\frac{1}{2}}(x)\right)z^2h\mu_g^{\frac{1}{2}} \\ o(h^2) &: \frac{1}{4}\left[f(x)f''(x)K'\left(zf^{\frac{1}{2}}(x)\right) - \frac{1}{2}(f'(x))^2K'\left(zf^{\frac{1}{2}}(x)\right) + \frac{1}{2}(f'(x))^2f^{\frac{1}{2}}(x)K''\left(zf^{\frac{1}{2}}(x)\right)\right] \\ &\quad + \frac{3}{4}\left[\frac{1}{2}\frac{(f'(x))^2}{f^{\frac{1}{2}}(x)} + f^{\frac{1}{2}}(x)f''(x)\right]K\left(zf^{\frac{1}{2}}(x)\right)z^2h^2\mu_g + \frac{3}{4}(f'(x))^2K'\left(zf^{\frac{1}{2}}(x)\right)z^3h^2\mu_g \\ &= \frac{3}{4}\frac{d}{dx}\left[f'(x)f^{\frac{1}{2}}(x)K\left(zf^{\frac{1}{2}}(x)\right)\right]z^2h^2\mu_g + \frac{1}{4}\frac{d}{dx}\left[f'(x)f(x)zK'\left(zf^{\frac{1}{2}}(x)\right)\right]z^2h^2\mu_g \\ &= \frac{1}{2}\frac{d^2}{dx^2}\left[f^{\frac{3}{2}}(x)K\left(zf^{\frac{1}{2}}(x)\right)\right]z^2h^2\mu_g \end{aligned}$$

$$\begin{aligned}
o(h^3) &: -\frac{1}{12} \left[f(x) f'''(x) - \frac{3}{2} f'(x) f''(x) + \frac{3}{4} \frac{(f'(x))^3}{f^{\frac{1}{2}}(x)} \right] K' \left(z f^{\frac{1}{2}}(x) \right) z^4 h^3 \mu_g^{\frac{3}{2}} \\
&- \frac{1}{12} \left(\frac{3}{2} f^{\frac{1}{2}}(x) f'(x) f''(x) - \frac{3}{4} \frac{(f'(x))^3}{f^{\frac{1}{2}}(x)} \right) K'' \left(z f^{\frac{1}{2}}(x) \right) z^5 h^3 \mu_g^{\frac{3}{2}} - \frac{1}{48} (f'(x))^3 K''' \left(z f^{\frac{1}{2}}(x) \right) z^6 h^3 \mu_g^{\frac{3}{2}} \\
&- \frac{1}{4} \left[\frac{3}{2} \frac{f'(x) f''(x)}{f^{\frac{1}{2}}(x)} - \frac{1}{4} \frac{(f'(x))^3}{f^{\frac{3}{2}}(x)} + f^{\frac{1}{2}}(x) f'''(x) \right] K \left(z f^{\frac{1}{2}}(x) \right) z^3 h^3 \mu_g^{\frac{3}{2}} \\
&- \frac{3}{8} \left(2 f'(x) f''(x) + \frac{1}{2} \frac{(f'(x))^3}{f(x)} \right) K' \left(z f^{\frac{1}{2}}(x) \right) z^4 h^3 \mu_g^{\frac{3}{2}} - \frac{3}{16} \frac{(f'(x))^3}{f^{\frac{1}{2}}(x)} K'' \left(z f^{\frac{1}{2}}(x) \right) z^5 h^3 \mu_g^{\frac{3}{2}}
\end{aligned}$$

El término de orden $o(h^4)$ no lo desarrollaremos, ya que sólo nos interesa demostrar que los anteriores términos se anulan para obtener el resultado.

Para el cálculo de la esperanza, tendremos que integrar estas expresiones. Teniendo en cuenta que la función kernel K verifica la propiedad f) (es simétrica), las respectivas derivadas $K', K'', K''', K^{(iv)}$ se irán alternando como funciones impares (derivadas $(2p-1)$ -ésimas) y pares (derivadas $2p$ -ésimas). Por lo tanto al calcular la integral, los términos de la forma $K^{(2p-1)} x^{2q}$ y $K^{(2p)} x^{2q-1}$ darán lugar a factores nulos al tratarse de funciones impares. Bajo estas observaciones las integrales de los términos $o(h_1)$ y $o(h_3)$ son nulas, de lo que deducimos que:

$$\begin{aligned}
E \left(\hat{f}_{K,adp}(x) \right) &= \int_{-\infty}^{\infty} f^{\frac{3}{2}} \left(x - zh\mu_g^{\frac{1}{2}} \right) K \left(z f^{\frac{1}{2}} \left(x - zh\mu_g^{\frac{1}{2}} \right) \right) dz = \int_{-\infty}^{\infty} f^{\frac{3}{2}}(x) K \left(z f^{\frac{1}{2}}(x) \right) dz + \\
&+ \int_{-\infty}^{\infty} \frac{1}{2} \frac{d^2}{dx^2} \left[f^{\frac{3}{2}}(x) K \left(z f^{\frac{1}{2}}(x) \right) \right] z^2 h^2 \mu_g dz + A(x) h^4 \mu_g^2 + o(h^4)
\end{aligned}$$

Realizando el cambio de variable $z f^{\frac{1}{2}}(x) = \tilde{z}$, pasaremos a trabajar entonces con:

$$E \left(\hat{f}_{K,adp}(x) \right) = f(x) \int_{-\infty}^{\infty} K(\tilde{z}) d\tilde{z} + \frac{1}{2} \frac{d^2}{dx^2} \left(\int_{-\infty}^{\infty} \tilde{z}^2 K(\tilde{z}) d\tilde{z} \right) + A(x) h^4 \mu_g^2 + o(h^4)$$

Aplicando las propiedades b) y e) de las funciones kernel, y como la segunda integral será un valor finito V_k que depende únicamente del kernel y no de la posición x , se tiene que:

$$E \left(\hat{f}_{K,adp}(x) \right) = f(x) + A(x) h^4 \mu_g^2 + o(h^4)$$

En conclusión, el sesgo (1.4) vendrá dado por:

$$\text{sesgo} \left(\hat{f}_{K,adp}(x) \right) = E \left(\hat{f}_{K,adp}(x) \right) - f(x) = A(x) h^4 \mu_g^2 + o(h^4)$$

□

Por la implicación de la Proposición 4.8, Silverman propone utilizar como parámetro de sensibilidad $\alpha = \frac{1}{2}$; con este exponente, conseguimos que el estimador sea asintóticamente insesgado de una forma más rápida, $o(h^4)$, de lo que se obtenía con el método básico de ventana fija, $o(h^2)$. Otra justificación,

más intuitiva, es que al seleccionar este valor estamos considerando la situación media entre el método de la anchura fija y el método del k -vecino más cercano, una combinación de ambos métodos.

4.2. Evaluación de los métodos de suavidad adaptada

En la [Sección 4.1](#) hemos presentado los principales métodos de estimación con amplitud de ventana variable. A continuación, aplicaremos estos métodos a nuestro ejemplo: la muestra de cuásares de la campaña SDSS.

4.2.1. Aplicación del método del k -vecino más cercano

En primer lugar, es importante destacar que al implementar el método del k -vecino más cercano, es necesario recurrir a la utilización de bucles anidados lo que aumenta el tiempo de computación. Por este motivo, para la evaluación de este método nos restringiremos ahora a tomar muestras aleatorias simples de menor tamaño (en concreto $m = 100$).

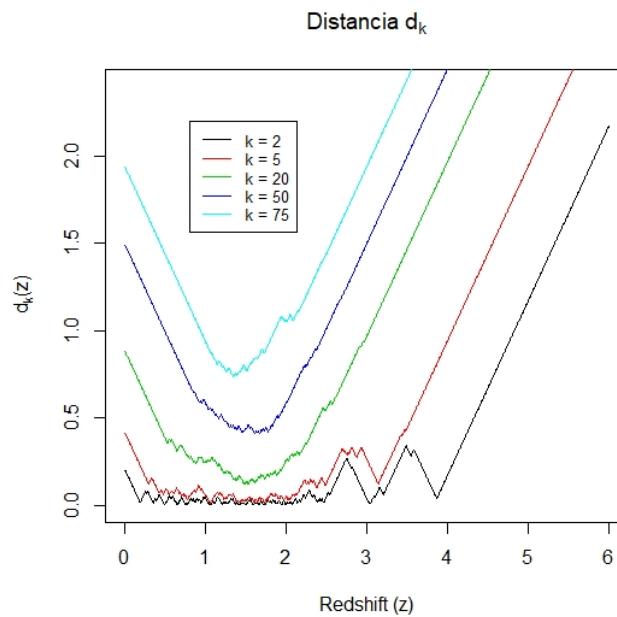
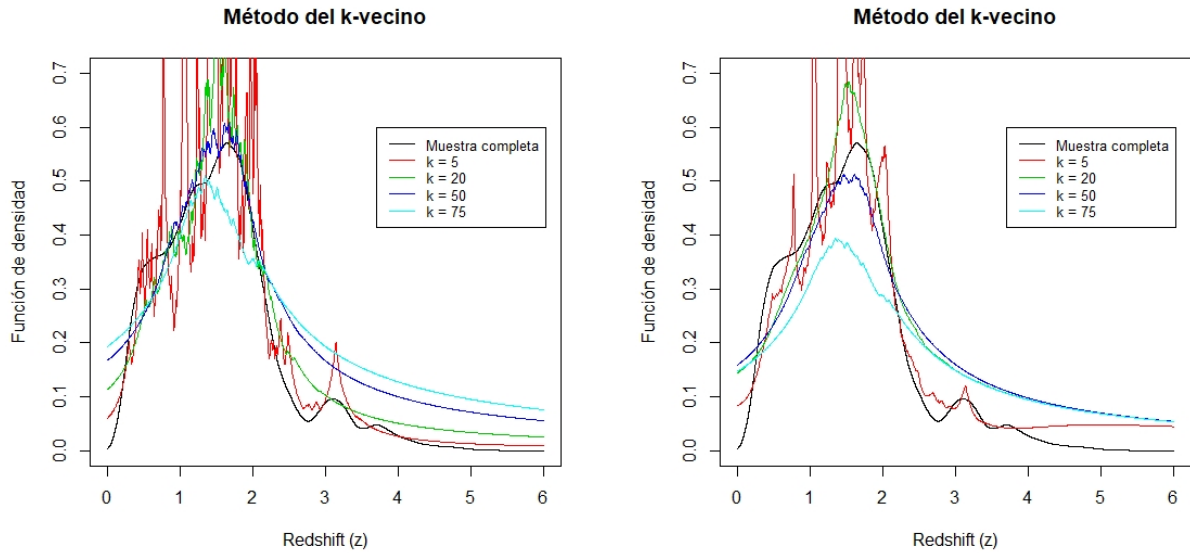


Figura 4.1: Comportamiento de la distancia d_k (4.1), dada una muestra aleatoria de 100 cuásares en la campaña SDSS.

En la [Figura 4.1](#) podemos observar la función $d_k(x)$ (4.1) para diferentes valores de k . Esta gráfica refleja los resultados de la Proposición 4.3: la función es claramente continua, pero presenta numerosos cambios de pendiente que indican la no derivabilidad de la misma. De hecho, podemos resaltar el papel que juega el número k . Cuando fijamos un k pequeño (por ejemplo $k = 2$), la distancia d_k presenta valores pequeños en el intervalo $(0, 2)$, lo que nos indica que gran parte de las muestras se localizan en dicha zona, mientras que el aumento a partir de $z = 4$ muestra la ausencia de muestras en esa zona. Estos resultados son más extremos al fijar un valor de k más grande (por ejemplo $k = 75$). Como la muestra

tiene un tamaño de 100, la distancia tomará valores más pequeños en un intervalo más reducido, puesto que se requiere de una mayor presencia de valores muestrales.



(a) Estimación de la función densidad con kernel rectangular.

(b) Estimación de la función densidad con kernel normal

Figura 4.2: Estimación de la función de densidad para una muestra aleatoria simple de 100 cuásares

En la [Figura 4.2](#) (a) presentamos la estimación directa por (4.2) o, equivalentemente, la estimación dada por (4.5) con un kernel rectangular. Valores pequeños k suponen un aumento de la varianza, puesto que la estimación mostrará una gran dependencia con la muestra aleatoria. Sin embargo, al aumentar el valor de k perdemos gran parte de información en las zonas donde hay menos muestras. Por otro lado, con independencia del valor k , podemos observar como la función estimada sobreestima la densidad en las zonas más alejadas, como consecuencia de la Proposición 4.4. En la [Figura 4.2](#) (b) realizamos la estimación con un kernel normal. El comportamiento que observamos es análogo al estudiado con anterioridad, si bien es cierto que hemos ganado cierta suavidad respecto a la anterior estimación.

El problema de la elección del entero k óptimo surge de forma análoga a la anchura de ventana óptima, idea que se refuerza a raíz del resultado (4.8). En la [Figura 4.3](#) comparamos la estimación realizada por el método del k -vecino más cercano, la estimación realizada con una amplitud de ventana fija (referida a la distribución normal) y la función de densidad a estimar. De acuerdo a la relación (4.8), observamos que el valor óptimo de k depende de h y de la función de densidad en el punto x . Tomando como valor de h óptimo el que nos proporciona la elección en el caso normal, y para evitar la dependencia en la posición, tomando el máximo valor de la función de densidad, obtenemos el valor $k = 32$. Comparando el método del 32-vecino más cercano y el método con anchura fija podemos observar que si bien el método variable estima bien en las zonas con mayor peso, no realiza un correcto ajuste en el resto de zonas. Por ejemplo, los máximos de la densidad situados en $z \sim 3.1$ y $z \sim 3.8$ son omitidos por completo en la estimación con anchura variable, pero se muestran en la estimación con anchura fija.

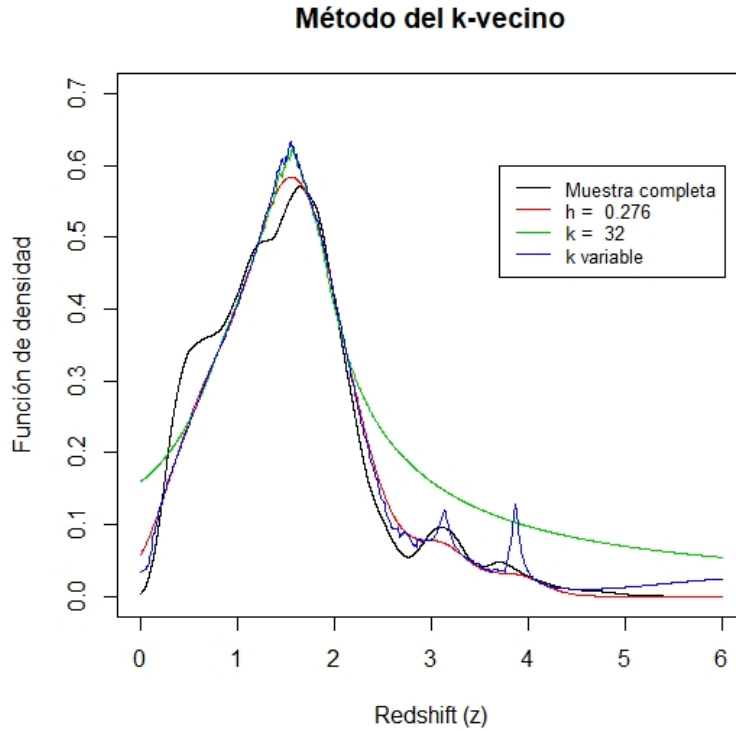


Figura 4.3: Comparativa de los métodos de anchura fija y variable, ambos con kernel normal, para una muestra aleatoria simple de 100 cuásares.

El problema que presenta el método con $k = 32$ reside en que es un número bueno para estimar la densidad en las zonas con mayor peso, pero no en las zonas de menor peso. Este resultado lo podíamos anticipar con la relación (4.8): el valor óptimo, suponiendo conocida la amplitud fija óptima \tilde{h} , dependerá de la posición por el factor $f(x)$, luego la mejor estimación será por medio de diferentes valores de k que dependerán de la posición. Este es precisamente el método que subyace a la estimación en color azul, donde hemos seleccionado la amplitud de ventana h óptima por (3.2). Observamos que mantiene el buen ajuste en las zonas de mayor peso, y en las zonas de menor peso recuperamos los máximos de la distribución.

4.2.2. Aplicación del método kernel adaptado

Del estudio presentado en la Subsección 4.1.2 concluimos que el método kernel adaptado tiene varios parámetros que podemos modificar para buscar la mejor estimación de la función de densidad: la estimación pivote inicial, la anchura de ventana genérica h y el parámetro de sensibilidad α .

Comencemos estudiando la influencia de la estimación pivote inicial. En la Figura 4.4 (a) se muestra el uso del método kernel adaptado con una estimación pivote dada por un kernel triangular y la anchura seleccionada por el método de sobre suavización. En comparación con la Figura 4.4 (b), donde hemos modificado la anchura de ventana pivote por un valor mucho mayor, podemos observar que la estimación es esencialmente la misma para todos los valores de α . De hecho, en la Figura 4.4 (c) consideramos una estimación pivote dada por el estimador naïve y una anchura de ventana grande, y seguimos obteniendo

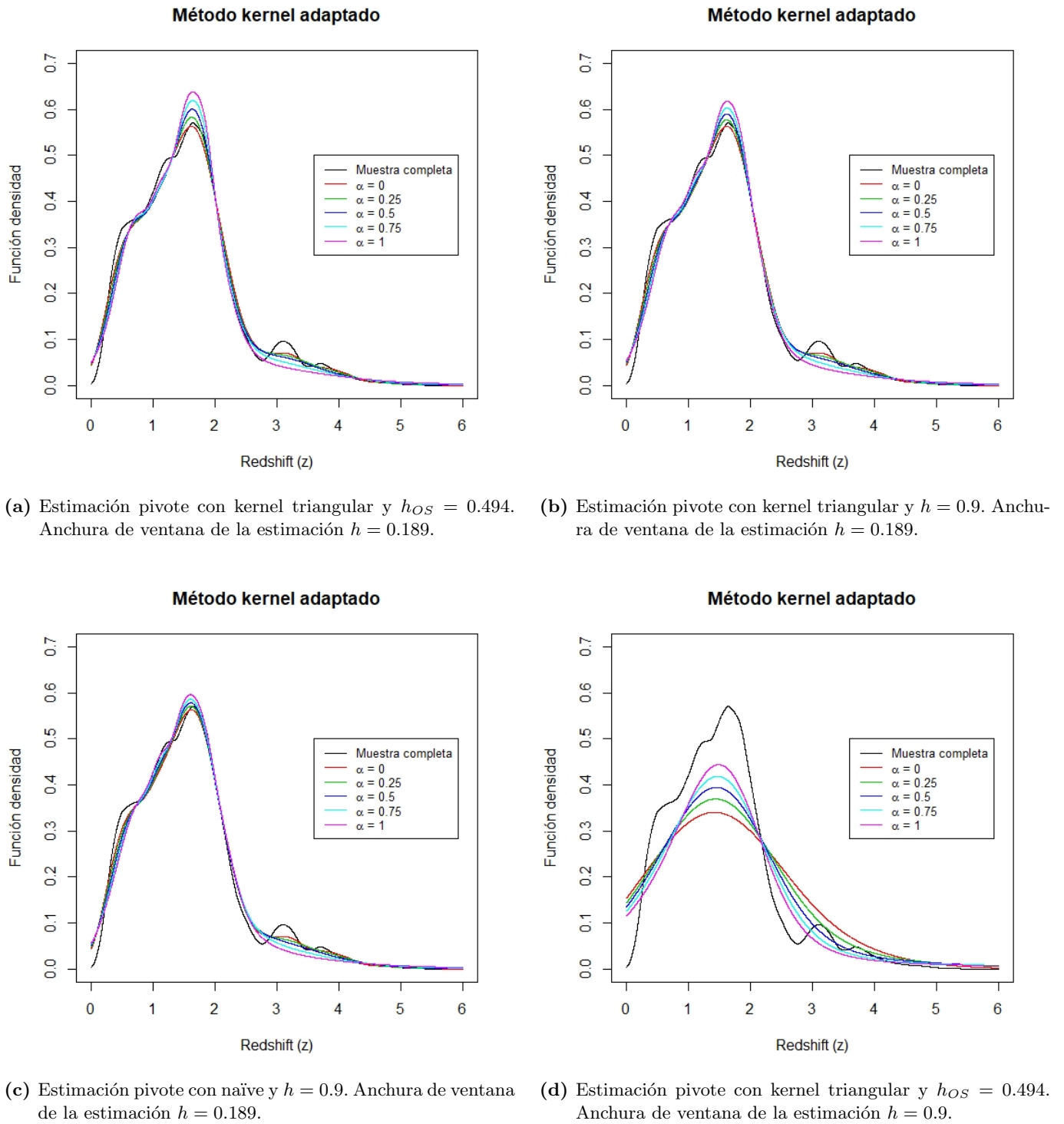


Figura 4.4: Método kernel adaptado, con función kernel normal, para una muestra de 1000 cuásares de la campaña SDSS.

los mismos resultados. Esto nos permite concluir que la estimación pivote no juega un papel determinante en la estimación inicial.

Por el contrario, la anchura de ventana h si juega un papel importante la estimación final. En la

Figura 4.4 (a) hemos seleccionado la anchura de ventana óptima referida a la distribución normal, cuyo valor es $h_N = 0.189$, mientras que en la Figura 4.4 (d) seleccionamos una anchura considerablemente mayor, $h = 0.9$. Como es de esperar, al introducir una anchura de ventana muy grande se pierde gran parte de la información en la estimación, lo mismo que ocurre en los métodos con anchura fija. Podemos concluir que con el método kernel adaptado se obtiene una buena estimación de la función de densidad siempre que se seleccione una anchura de ventana razonable. Por este motivo, lo más aconsejable es recurrir a alguna de las aproximaciones comentadas en el capítulo anterior de la anchura de ventana óptima fija.

Por otro lado, el parámetro de sensibilidad α nos permite justificar las estimaciones que obtenemos referidas a la función de densidad de la muestra completa. Cuando seleccionamos $\alpha = 0$ estamos considerando de nuevo los métodos con anchura fija, y podemos observar que realizan una buena estimación en las zonas de menor densidad. Al utilizar $\alpha = 0.5$, valor necesario para una mayor reducción del sesgo de acuerdo a la Proposición 4.8, la estimación parece fallar en las zonas donde el método con anchura fija se ajusta mejor. Esto se debe a la muestra aleatoria simple: la mayoría de los datos muestrales caen en la zona $(0.1, 2.4)$, motivo por el cual en dicho intervalo se selecciona una anchura menor, mientras que en el resto de zonas la anchura será mayor, por lo que se pierde información.

Anexo de programas en R

A continuación presentamos el código empleado en R para la construcción de todos los ejemplos basados en la muestra de cuásares.

A.1. Código del Ejemplo 3.6

```
load("../SDSS_cuasares/qso.Rdata"); #Lectura de los datos
mc <- qso$z; #Selección del redshift (parámetro a estudiar)
ma <- sample(mc, 1000, replace = TRUE); #Construcción de una m.a.s. de tamaño 1000
hn <- ( ((25200/143)*sqrt( pi ))^(1/5) )*(length(mc)^(-(1/5)))*sd(mc); #Anchura de ventana bajo
    la que se representarán los datos globales
S <- c(1e-1,1e-2,1e-3,1e-4,1e-5,1e-6); #Precisión método
R <- c(); #Vector auxiliar para el cálculo
Raux <- c(); #Vector auxiliar para el cálculo
h <- c(); #Vector para almacenar las ventas obtenidas
It <- c(); #Vector para guardar el número de iteraciones requeridas para la convergencia
temp <- c(); #Vector para guardar el tiempo empleado en la computación
#Forma explícita de las convoluciones
conv1 <- function(x) ((32*sqrt( pi ))^(-1))*(12 - 12*x^2 + x^4 )*exp( -(x^2)/4 ); #Convolución
    del kernel normal
conv3 <- function(x) ((105/16)^2)*((1/630)*(-25*x^9 + 360*x^7 - 2016*x^5 + 6720*x^3 - 5760*x^2
    + 512)*(x >= 0)*(x <= 2) + (1/630)*(25*x^9 -360*x^7+2016*x^5 - 6720*x^3 -5760*x^2 + 512)*(x
    <= 0)*(x >= -2) - (64/315)*(4 - 45*x^2)*(x == 0) ); #Convolución del kernel triweight
#Comienzo del método (se barren las diferentes precisiones)
for (k in 1:length(S)) {
    It[k] <- 0; #Iteración a 0
    h0 <- 0; #Ventana inicial para el método
    h1 <- 5; #Ventana inicial para el método
    s.time <- Sys.time() #Comienzo de la medida del tiempo
    while ( abs(h0 - h1) > S[k] ) {
        It[k] <- It[k] + 1; #Cálculo de iteraciones
        h0 <- h1; #Selección de la anchura previa
        for (i in 1:length(ma)) {
            for (j in 1:length(ma)) {
```

```

        Raux[j] <- conv3( (ma[j]-ma[i])/h0 ); #Función convolución para
        cada valor
    }
    R[i] <- sum(Raux); #Primer sumatorio
    Raux <- c(); #Valores obtenidos del primer sumatorio para cada índice del
    segundo
}
Rt <- sum(R); #Segundo sumatorio
R <- c(); #Se resetea el vector auxiliar
h1 <- ( ( 9450/(Rt*143) )^(1/5) )*( length(ma)^(1/5) )*h0; #Cálculo de la nueva
anchura
}
f.time <- Sys.time() #Tiempo de finalización del método
temp[k] <- f.time - s.time; #Tiempo de computación
h[k] <- h1; #Anchura de ventana para la que converge el método
}
#Cálculo de las estimaciones kernel con Epanechnikov
ept <- function(x) (1/(length(mc)*hn))*sum( (35/32)*(( 1 - ((mc-x)/hn)^2 )^3)*( abs((mc-x)/hn)
    <= 1 ) ); eptv <- Vectorize(ept);
#Muestra de 1000 cuásares para los valores asociados a precisión 10-1 y 10-5
ep1 <- function(x) (1/(length(ma)*h[1]))*sum( (35/32)*(( 1 - ((ma-x)/h[1])^2 )^3)*( abs((ma-x)/
    h[1]) <= 1 ) ); ep1v <- Vectorize(ep1);
ep2 <- function(x) (1/(length(ma)*h[5]))*sum( (35/32)*(( 1 - ((ma-x)/h[5])^2 )^3)*( abs((ma-x)/
    h[5]) <= 1 ) ); ep2v <- Vectorize(ep2);
x <- seq(0,6,length=1e3); #Array para la representación gráfica
plot(x,eptv(x),type='l',main='Kernel_Triweight',ylab='Función_densidad',xlab='Redshift_(z)',col
    =1,ylim=c(0,0.7)); lines(x,ep1v(x),col=2) ; lines(x,ep2v(x),col=3)
legend(3.3,0.6,c('Muestra_completa',paste('h_=',round(h[1],5),'_/_It_=',It[1] ),paste('h_=',
    round(h[5],5),'_/_It_=',It[5] )),col=c(1,2,3),lty=c(1,1,1),cex=0.8)
#Representación de las iteraciones y tiempo vs la precisión requerida
plot(c(-1,-2,-3,-4,-5,-6),It,type='p',pch=15,main='Evolución_de_las_iteraciones',ylab='Nº_
    Iteraciones',xlab=expression(paste( 'log' [10] , '(S)')),col=2,ylim=c(0,45)); lines(c
    (-1,-2,-3,-4,-5,-6),It, col=2, lty=2)
temp1[1] <- temp[1]; temp1[2] <- 60*temp[2]; temp1[3] <- 60*temp[3]; temp1[4] <- 60*temp[4];
temp1[5] <- 60*temp[5]; temp1[6] <- 60*temp[6];
plot(c(-1,-2,-3,-4,-5,-6),temp1,type='p',pch=15,main='Tiempo_de_computación',ylab='Tiempo_(s)',
    xlab=expression(paste( 'log' [10] , '(S)')),col=2,ylim=c(0,600)); lines(c(-1,-2,-3,-4,-5,-6),
    temp1, col=2, lty=2)

```

A.2. Código del estudio de la Subsección 3.3.6

```

load("../SDSS_cuasares/qso.Rdata"); #Lectura de los datos
mc <- qso$z; #Selección del parámetro a analizar
ma <- sample(mc, 500, replace = TRUE); m <- length(ma); #Construcción de una m.a.s de tamaño

```

```

500
Nor <- function(x) (1/sqrt(2*pi))*exp(-(x^2)/2); #Kernel normal
Triw <- function(x) (35/32)*((1-x^2)^3)*(x >= -1)*(x <= 1); #Kernel triweight
#Cálculo de convoluciones
convn <- function(x) (1/(2*sqrt(pi)))*exp(-(x^2)/4 );
convt <- function(x) (-35/1757184)*((x-2)^7)*(5*x^6 + 70*x^5 + 404*x^4 + 1176*x^3 +1616*x^2 +
  1120*x +320)*(x >= 0)*(x <= 2) + (35/1757184)*((x+2)^7)*(5*x^6 - 70*x^5 + 404*x^4 - 1176*x
  ^3 +1616*x^2 - 1120*x +320)*(x <= 0)*(x >= -2) - (350/429)*(x == 0);
convn2 <- function(x) (1/(32*sqrt(pi)))*((12 - 12*x^2 + x^4 )*exp( -(x^2)/4 ));
convt2 <- function(x) ((105/16)^2)*((1/630)*(-25*x^9 + 360*x^7 - 2016*x^5 + 6720*x^3 - 5760*x^2
  + 512)*(x >= 0)*(x <= 2) + (1/630)*(25*x^9 -360*x^7 +2016*x^5 - 6720*x^3 -5760*x^2 + 512)*(
  x <= 0)*(x >= -2) - (64/315)*(4 - 45*x^2)*(x == 0) );
NMV <- c(); #Función Maxima verosimilitud K triweight
NMV2 <- c(); #Función Maxima verosimilitud K normal
NLS <- c(); #Función LS Minimos cuadrados K triweight
NLS2 <- c(); #Función LS Minimos cuadrados K normal
NBC <- c(); #Función BC K triweight
NBC2 <- c(); #Función BC K normal
h <- seq(0.01,1,length=1e2); #Anchuras de ventana para el kernel triweight
h2 <- seq(0.001,1,length=1e2); #Anchuras de ventana para el kernel normal
#Cálculo función BC(h) para triweight
for (k in 1:length(h)){
  L <- c();
  for (i in 1:length(ma)) {
    Laux <- c();
    for (j in 1:length(ma)) {
      Laux[j] <- convt2( (ma[i]-ma[j])/h[k] )*(i != j);
    }
    L[i] <- sum(Laux);
  }
  NBC[k] <- sum(L)*(((1/9)^2)/(4*(m^2)*h[k])) + (1/(m*h[k]))*(350/429);
}
#Cálculo función BC(h) para normal
for (k in 1:length(h2)){
  L <- c();
  for (i in 1:length(ma)) {
    Laux <- c();
    for (j in 1:length(ma)) {
      Laux[j] <- convn2( (ma[i]-ma[j])/h2[k] )*(i != j);
    }
    L[i] <- sum(Laux);
  }
  NBC2[k] <- sum(L)*(1/(4*(m^2)*h2[k])) + (1/(m*h2[k]))*(1/(2*sqrt(pi)));
}
#Cálculo función LS(h) para triweight
for (k in 1:length(h)){
  L <- c();

```

```

    for (i in 1:length(ma)) {
      Laux <- c();
      for (j in 1:length(ma)) {
        Laux[j] <- (1/m)*convt( (ma[i]-ma[j])/h[k] ) - (2/(m-1))*Triw( (ma[i]-ma[
          j])/h[k] )*(i != j);
      }
      L[i] <- sum(Laux);
    }
    NLS[k] <- (1/(m*h[k]))*sum(L)
  }
#Cálculo función LS(h) para normal
for (k in 1:length(h2)){
  L <- c();
  for (i in 1:length(ma)) {
    Laux <- c();
    for (j in 1:length(ma)) {
      Laux[j] <- (1/m)*convn( (ma[i]-ma[j])/h2[k] ) - (2/(m-1))*Nor( (ma[i]-ma[
        j])/h2[k] )*(i != j);
    }
    L[i] <- sum(Laux);
  }
  NLS2[k] <- (1/(m*h2[k]))*sum(L)
}
#Cálculo función verosimilitud(h) para triweight
for (k in 1:length(h)){
  L <- c();
  for (i in 1:length(ma)) {
    Laux <- c();
    for (j in 1:length(ma)) {
      Laux[j] <- Triw( (ma[i]-ma[j])/h[k] )*(i != j);
    }
    L[i] <- log(sum(Laux));
  }
  NMV[k] <- sum(L) - m*log( (m-1)*h[k] );
}
#Cálculo función verosimilitud(h) para normal
for (k in 1:length(h2)){
  L <- c();
  for (i in 1:length(ma)) {
    Laux <- c();
    for (j in 1:length(ma)) {
      Laux[j] <- Nor( (ma[i]-ma[j])/h2[k] )*(i != j);
    }
    L[i] <- log(sum(Laux));
  }
  NMV2[k] <- sum(L) - m*log( (m-1)*h2[k] );
}
}

```

```

#Método Plug-In con precisión  $10^{-2}$  para el kernel normal
S <- c(1e-2); R <- c(); Raux <- c(); hr <- c(); It <- c(); temp <- c();
conv1 <- function(x) ((32*sqrt( pi ))^(-1))*(12 - 12*x^2 + x^4 )*exp( -(x^2)/4 );
conv2 <- function(x) (9/4)*((2 - x)*( x >= 0)*( x <= 2 ) + (2 + x)*( x <= 0 )*(x >= -2) - 2*( x
== 0));
conv3 <- function(x) ((105/16)^2)*((1/630)*(-25*x^9 + 360*x^7 - 2016*x^5 + 6720*x^3 - 5760*x^2
+ 512)*(x >= 0)*(x <= 2) + (1/630)*(25*x^9 -360*x^7+2016*x^5 - 6720*x^3 -5760*x^2 + 512)*(x
<= 0)*(x >= -2) - (64/315)*(4 - 45*x^2)*(x == 0) );
for (k in 1:length(S)) {
  It[k] <- 0; h0 <- 0; h1 <- 5;
  s.time <- Sys.time()
  while ( abs(h0 - h1) > S[k] ) {
    It[k] <- It[k] + 1; h0 <- h1;
    for (i in 1:length(ma)) {
      for (j in 1:length(ma)) {
        Raux[j] <- convn2( (ma[j]-ma[i])/h0 );
      }
      R[i] <- sum(Raux); Raux <- c();
    }
    Rt <- sum(R); R <- c(); h1 <- ( ( 1/(Rt*2*sqrt(pi)) )^(1/5) )*( length(ma)^(1/5)
)*h0;
  }
  f.time <- Sys.time()
  temp[k] <- f.time - s.time; hr[k] <- h1;
}

#Método Plug-In con precisión  $10^{-2}$  para kernel triweight
S2 <- c(1e-2); R2 <- c(); Raux2 <- c(); hr2 <- c(); It2 <- c(); temp2 <- c();
for (k in 1:length(S2)) {
  It2[k] <- 0;
  h02 <- 0;
  h12 <- 5;
  s.time <- Sys.time()
  while ( abs(h02 - h12) > S2[k] ) {
    It2[k] <- It2[k] + 1;
    h02 <- h12;
    for (i in 1:length(ma)) {
      for (j in 1:length(ma)) {
        Raux2[j] <- convt2( (ma[j]-ma[i])/h02 );
      }
      R2[i] <- sum(Raux2); Raux2 <- c();
    }
    R2t <- sum(R2); R2 <- c(); h12 <- ( ( 9450/(R2t*143) )^(1/5) )*( length(ma)^(1/
5) )*h02;
  }
  f.time <- Sys.time()
  temp2[k] <- f.time - s.time; hr2[k] <- h12;
}

```



```

#Representación gráfica de las funciones verosimilitud
plot(h,NMV, type = 'l',main='Funciones de máxima verosimilitud',xlab='h',ylab=expression(paste('
  log(', 'L' [h] [,] [K], ')))'); lines(h2,NMV2,col=2)
legend(0.69,-566,c('Triweight','Normal'),col=c(1,2),lty=c(1,1),cex=0.9)
h[which.max(NMV)] ; h2[which.max(NMV2)]; #Anchuras que las optimizan h(triweight), h2(normal)
#Representación de las funciones de mínimos cuadrados
plot(h,NLS, type = 'l',main='Funciones de mínimos cuadrados',xlab='h',ylab=expression(paste(
  widehat(LS), '(h)'))); lines(h2,NLS2,col=2)
legend(0.6,-0.3,c('Triweight','Normal'),col=c(1,2),lty=c(1,1),cex=0.9)
h[which.min(NLS)]; h2[which.min(NLS2)] ; #Anchuras que las optimizan h(triweight), h2(normal)
#Representación de las funciones BC(h)
plot(h,NBC, type = 'l',main='Funciones del estimador BC',xlab='h',ylab='BC(h)',ylim=c(0,0.03)) ;
  lines(h2,NBC2,col=2)
legend(0.6,0.03,c('Triweight','Normal'),col=c(1,2),lty=c(1,1),cex=0.9)
#Zoom en la zona deseada para el cálculo de las anchuras óptimas de las funciones BC(h)
ah <- seq(0.01,10,length=1e2); ah2 <- seq(0.001,10,length=1e2);
ANBC <- c(); ANBC2 <- c();
for (k in 1:length(ah)){
  L <- c();
  for (i in 1:length(ma)) {
    Laux <- c();
    for (j in 1:length(ma)) {
      Laux[j] <- convt2( (ma[i]-ma[j])/ah[k] )*(i != j);
    }
    L[i] <- sum(Laux);
  }
  ANBC[k] <- sum(L)*(((1/9)^2)/(4*(m^2)*ah[k])) + (1/(m*ah[k]))*(350/429);
}
for (k in 1:length(ah2)){
  L <- c();
  for (i in 1:length(ma)) {
    Laux <- c();
    for (j in 1:length(ma)) {
      Laux[j] <- convn2( (ma[i]-ma[j])/ah2[k] )*(i != j);
    }
    L[i] <- sum(Laux);
  }
  ANBC2[k] <- sum(L)*(1/(4*(m^2)*ah2[k])) + (1/(m*ah2[k]))*(1/(2*sqrt(pi)));
}
plot(ah,ANBC, type = 'l',main='Funciones del estimador BC (extendido)',xlab='h',ylab='BC(h)',
  ylim=c(0,0.03)); lines(ah2,ANBC2,col=2)
legend(6,0.03,c('Triweight','Normal'),col=c(1,2),lty=c(1,1),cex=0.9)
hosn <- ( (1/(70*sqrt(pi)))^(1/5) )*3*sd(ma)*m^(-(1/5)); #Sobresuavización K normal
host <- ( (810/429)^(1/5) )*3*sd(ma)*m^(-(1/5)); #Sobresuavización K ntriweight
hnn <- ( (4/3)^(1/5) )*sd(ma)*m^(-(1/5)); #Referida a la normal, K normal
hnt <- ( ( (25200/143)*sqrt(pi) )^(1/5) )*sd(ma)*m^(-(1/5)); #Referida a la normal, K triweight
htnt <- ( ( (25200/143)*sqrt(pi) )^(1/5) )*sd(mc)*(length(mc)^(-(1/5))); #Kernel triweight (

```

```

global)
htnn <- ( (4/3)^(1/5) ) * sd(mc) * (length(mc)^(-(1/5))); #Kernel normal (global)
#Construcción y vectorización de las estimaciones para los métodos de validación cruzada
t1 <- function(x) (1/(m*ht[1])) * sum( (35/32) * ((1 - ((x - ma)/ht[1])^2)^3) * (abs((x - ma)/ht
  [1]) <= 1) );
t2 <- function(x) (1/(m*ht[2])) * sum( (35/32) * ((1 - ((x - ma)/ht[2])^2)^3) * (abs((x - ma)/ht
  [2]) <= 1) );
t3 <- function(x) (1/(m*ht[3])) * sum( (35/32) * ((1 - ((x - ma)/ht[3])^2)^3) * (abs((x - ma)/ht
  [3]) <= 1) );
n1 <- function(x) (1/(m*hn[1])) * sum( (1/sqrt(2*pi)) * exp( - (((x - ma)/hn[1])^2) / 2 ) );
n2 <- function(x) (1/(m*hn[2])) * sum( (1/sqrt(2*pi)) * exp( - (((x - ma)/hn[2])^2) / 2 ) );
n3 <- function(x) (1/(m*hn[3])) * sum( (1/sqrt(2*pi)) * exp( - (((x - ma)/hn[3])^2) / 2 ) );
tt <- function(x) (1/(length(mc)*htnt)) * sum( (35/32) * ((1 - ((x - mc)/htnt)^2)^3) * (abs((x - mc
  )/htnt) <= 1) );
nn <- function(x) (1/(length(mc)*htnn)) * sum( (1/sqrt(2*pi)) * exp( - (((x - mc)/htnn)^2) / 2 ) );
t1v <- Vectorize(t1); t2v <- Vectorize(t2); t3v <- Vectorize(t3); ttv <- Vectorize(tt);
n1v <- Vectorize(n1); n2v <- Vectorize(n2); n3v <- Vectorize(n3); nnv <- Vectorize(nn);
x <- seq(0,6,length=1e3); #Array para la representación gráfica
plot(x,ttv(x),type='l',main='Kernel_Triweight',xlab='Redshift_(z)', ylab='Función_densidad');
  lines(x,t1v(x),col=2); lines(x,t2v(x),col=3); lines(x,t3v(x),col=4);
legend(3.5,0.55,c('Muestra_completa','V.C._Máx._Ver.','V.C._Mín._Cuad.','V.C._Sesgo'),col=c
  (1,2,3,4),lty=c(1,1,1,1),cex=0.9)
plot(x,nnv(x),type='l',main='Kernel_Normal',xlab='Redshift_(z)', ylab='Función_densidad');
  lines(x,n1v(x),col=2); lines(x,n2v(x),col=3); lines(x,n3v(x),col=4);
legend(3.5,0.55,c('Muestra_completa','V.C._Máx._Ver.','V.C._Mín._Cuad.','V.C._Sesgo'),col=c
  (1,2,3,4),lty=c(1,1,1,1),cex=0.9)

```

A.3. Código del estudio de la Subsección 3.4.4

```

load("../SDSS_cuasares/qso.Rdata"); #Lectura de los datos
mc <- qso$z; #Selección del redshift (parámetro a estudiar)
ma <- sample(mc, 100, replace = TRUE); m <- length(ma); #Construcción de una m.a.s. de tamaño
  100
### DATOS GLOBALES ###
haux <- min(sd(mc), (quantile(mc,0.75)-quantile(mc,0.25))/1.34 ); #Anchura referida a la
  distribución normal
hn <- ( (40*sqrt( pi ))^(1/5) ) * (length(mc)^(-(1/5))) * haux;
fep <- function(x) (1/(length(mc)*hn)) * sum( (3/4) * (1 - ((mc-x)/hn)^2) * (abs((mc-x)/hn) <= 1)
  ); fepv <- Vectorize(fep);
### CÁLCULO PREVIO DE LA VENTANA POR MÍNIMOS CUADRADOS ###
convn <- function(x) (1/(2*sqrt(pi))) * exp(-(x^2)/4 ); #Normal
#Cálculo función LS(h) para normal y epanechnikov
kex <- function(x) (1/sqrt(2*pi)) * exp( - (x^2)/2 ); #Kernel normal
NLSn <- c(); #Almacenamiento valores NLS

```

```

h2 <- seq(0.01,1,length=1e2); #Precisión 1e-2
for (k in 1:length(h2)){
  Ln <- c();
  for (i in 1:m) {
    Lauxn <- c();
    for (j in 1:m) {
      Lauxn[j] <- (1/m)*convn( (ma[i]-ma[j])/h2[k] ) - (2/(m-1))*kex( (ma[i]-ma
        [j])/h2[k] )*(i != j);
    }
    Ln[i] <- sum(Lauxn);
  }
  NLSn[k] <- (1/(m*h2[k]))*sum(Ln)
}
hLSn <- h2[which.min(NLSn)]; #Anchura por mínimos cuadrados (normal)
##BOOTSTRAP SIN ANCHURA PIVOTE y MÉTODO CAO,CUEVAS,MANTEIGA (CCM) ##
#Tamaño t del método CCM
t <- round(sqrt(m)); a <- (m/t)^(1/5); t1 <- m; a1 <- (m/t1)^(1/5); t2 <- 5*m; a2 <- (m/t2)^(1/
  5);
#Diferentes valores para la anchura de ventana
h <- seq(0.01,1,length=1e2); #Precisión 1e-2 para métodos bootstrap con pivote
g <- seq(0.01,1,length=1e2) ; #Precisión 1e-2 para método Hall
#Función bajo la que se define el MISE (sin anchura pivote)
auxMISE1n <- function(x) ( (m-1)/m ) *exp( - (x^2)/8 ) - (4/sqrt(3))*exp( - (x^2)/6 ) + sqrt(2)*
  exp( - (x^2)/4 ) ;
auxMISE3n <- function(x) ((t - 1)/(a*t*m^2))* (1/(2*sqrt(pi)))*exp( - (x^2)/(4*a^2) ) - (2/(m^2)
  )*( 1/(sqrt( 2*pi*(1+a^2) ) ) ) *exp( - (x^2)/(2*(1+a^2)) ) + exp( - (x^2)/4 )*(1/(m^2))* (1/
  (2*sqrt(pi))) ;
auxMISE3n1 <- function(x) ((t1 - 1)/(a1*t1*m^2))* (1/(2*sqrt(pi)))*exp( - (x^2)/(4*a1^2) ) - (2/
  (m^2))* ( 1/(sqrt( 2*pi*(1+a1^2) ) ) ) *exp( - (x^2)/(2*(1+a1^2)) ) + exp( - (x^2)/4 )*(1/(m
  ^2))* (1/(2*sqrt(pi))) ;
auxMISE3n3 <- function(x) ((t2 - 1)/(a2*t2*m^2))* (1/(2*sqrt(pi)))*exp( - (x^2)/(4*a2^2) ) - (2/
  (m^2))* ( 1/(sqrt( 2*pi*(1+a2^2) ) ) ) *exp( - (x^2)/(2*(1+a2^2)) ) + exp( - (x^2)/4 )*(1/(m
  ^2))* (1/(2*sqrt(pi))) ;
#Vector de almacenamiento de los valores del MISE para cada h
#Kernel normal
MISEB1n <- c(); #Sin anchura pivote
MISEB3n <- c(); #Metodo Hall (t < m)
MISEB3n1 <- c(); #Metodo Hall (t=m)
MISEB3n3 <- c(); #Metodo Hall (t > m)
for (i in 1:length(h)){
  AUX11n <- c(); AUX31n <- c(); AUX31n1 <- c(); AUX31n3 <- c();
  for (j in 1:m) {
    AUX12n <- c(); AUX32n <- c(); AUX32n1 <- c(); AUX32n3 <- c();
    for (k in 1:m) {
      AUX12n[k] <- auxMISE1n( (ma[k]-ma[j])/h[i] ); AUX32n[k] <- auxMISE3n( (ma
        [k]-ma[j])/g[i] ); AUX32n1[k] <- auxMISE3n1( (ma[k]-ma[j])/g[i] );
      AUX32n3[k] <- auxMISE3n3( (ma[k]-ma[j])/g[i] );
    }
  }
}

```

```

    }
    AUX11n[j] <- sum( AUX12n ); AUX31n[j] <- sum( AUX32n ); AUX31n1[j] <- sum(
      AUX32n1 ); AUX31n3[j] <- sum( AUX32n3 );
  }
  MISEB1n[i] <- (1/( (m^2)*h[i]*sqrt(8*pi ) ))*( sqrt(2)*m + sum( AUX11n ) ); MISEB3n[i]
    <- (1/g[i])*( (1/(2*sqrt(pi)))*(1/(t*a))+ sum( AUX31n ) ) ); MISEB3n1[i] <- (1/g[i])*
    ( (1/(2*sqrt(pi)))*(1/(t1*a1))+ sum( AUX31n1 ) ) ); MISEB3n3[i] <- (1/g[i])*( (1/(2*
      sqrt(pi)))*(1/(t2*a2))+ sum( AUX31n3 ) ) );
}

## MÉTODO BOOTSTRAP CON ANCHURA PIVOTE ##
xp <- seq(0,6,length=1e2);
MISEB2n <- c();
M <- 50; #Repetición del método Bootstrap
Fpiv2 <- c(); #Método con anchura pivote, función de distribución
for (k in 1:length(xp)) {
  Fpiv2[k] <- (1/m)*sum(pnorm( (xp[k] - ma)/hLSn ))
}

for (j in 1:length(h)) {
  AUX2n <- c();
  for (n in 1:M) {
    #Generación muestras bootstrap
    y1 <- runif(m,0,1); #Muestras tamaño m
    mb2 <- c(); #Almacenamiento muestra bootstrap (Con anchura pivote)
    for (s in 1:length(y1)) {
      mb2[s] <- min( xp[ which( Fpiv2 >= y1[s] ) ] );
    }
    ##MISE MÉTODO CON ANCHURA PIVOTE
    AUX21n <- c();
    for (s in 1:m) {
      AUX22n <- c();
      for (k in 1:m) {
        AUX22n[k] <- (1/(2*sqrt(pi)*h[j]*m^2))*exp( -((mb2[k]-mb2[s])^2)/
          (4*h[j]^2) ) - (2/(sqrt(2*pi)*( 1 + (hLSn/h[j])^2 ) ) *h[j]*m^2
          ) *exp( - ((mb2[k] - ma[s])^2)/(2*(h[j]^2+hLSn^2)) ) + (1/(2*
            sqrt(pi)*hLSn*m^2))*exp( -((ma[k]-ma[s])^2)/(4*hLSn^2) );
      }
      AUX21n[s] <- sum(AUX22n);
    }
    AUX2n[n] <- sum( AUX21n );
  }
  MISEB2n[j] <- (1/M)*sum( AUX2n );
}

MISEB3n2 <- c();
## MÉTODO BOOTSTRAP NO SUAVE SIMULADO
for (j in 1:length(g)) {
  AUX3n2 <- c(); #Función auxiliar
  for (n in 1:M) {

```

```

mb3 <- sample(ma,t,replace=TRUE); #Muestra bootstrap sin suavización
AUX31n21 <- c();
for (k in 1:t) {
  AUX32n21 <- c();
  for (s in 1:t) {
    AUX32n21[s] <- (1/(2*sqrt(pi)*g[j]*a*t^2))*exp( -((mb3[k]-mb3[s])
      ^2)/(4*(g[j]*a)^2) )
  }
  AUX31n21[k] <- sum(AUX32n21);
}
AUX31n22 <- c();
for (k in 1:m) {
  AUX32n22 <- c();
  for (s in 1:m) {
    AUX32n22[s] <- (1/(2*sqrt(pi)*g[j]*m^2))*exp( -((ma[k]-ma[s])^2)/
      (4*(g[j])^2) )
  }
  AUX31n22[k] <- sum(AUX32n22);
}
AUX31n23 <- c();
for (k in 1:t) {
  AUX32n23 <- c();
  for (s in 1:m) {
    AUX32n23[s] <- - (2/(sqrt(2*pi*( 1 + a^2 ) ) *g[j]*m*t))*exp( - ((
      mb3[k] - ma[s])^2)/(2*(g[j]^2+(g[j]*a)^2)) )
  }
  AUX31n23[k] <- sum(AUX32n23);
}
AUX3n2[n] <- sum( AUX31n23 ) + sum( AUX31n22 ) + sum( AUX31n21 );
}
MISEB3n2[j] <- (1/M)*sum( AUX3n2 );
}

#Ventanas obtenidas kernel normal
hSPn <- h[ which.min(MISEB1n) ]; #Anchura Bootstrap sin anchura pivote
hCPn <- h[ which.min(MISEB2n) ]; #Anchura Bootstrap con anchura pivote
gCCMn <- g[ which.min(MISEB3n) ]; #Anchura Bootstrap Método CCM (t < m)
gCCMn1 <- g[ which.min(MISEB3n1) ]; #Anchura Bootstrap Método CCM (t = m)
gCCMn3 <- g[ which.min(MISEB3n3) ]; #Anchura Bootstrap Método CCM (t > m)
hSauxn <- min(sd(ma), (quantile(ma,0.75)-quantile(ma,0.25))/1.34 );
hSSn <- ((1/(2*sqrt(pi)*35))^(1/5))*3*hSauxn*(m^(-(1/5))); #Anchura por sobre suavizamiento
hNn <- ((4/3)^(1/5))*hSauxn*(m^(-(1/5))); #Anchura referida a la distribución normal
plot(g,MISEB1n,type='l',col=1,main='Bootstrap con estimación pivote',xlab='h',ylab=expression(
  MISE [B] (h) ),ylim=c(0,0.1)); lines(h,MISEB2n,col=2);
legend(0.55,0.08,c('Sin anchura pivote','Con anchura pivote'),col=c(1,2),lty=c(1,1),cex=0.9)
plot(g,MISEB3n,type='l',col=1,main='Bootstrap sin estimación pivote',xlab='g',ylab=expression(
  MISE [B] (g) ),ylim=c(0,0.1)); lines(g,MISEB3n2,col=2); lines(g,MISEB3n1,col=3); lines(g,
  MISEB3n3,col=4)

```

```

legend(0.4,0.09,c('Método de Hall teórico (t<m)', 'Método de Hall simulado (t<m)', 'Método de
Hall teórico (t=m)', 'Método de Hall teórico (t>m)'), col=c(1,2,3,4), lty=c(1,1,1,1), cex=0.9)
#Estimaciones de la función de densidad con las anchuras
fkex1 <- function(x) (1/(length(ma)*hNn))*sum( (1/sqrt(2*pi)*exp( -(((ma-x)/hNn)^2)/2 ) ) );
fkex1v <- Vectorize(fkex1);
fkex2 <- function(x) (1/(length(ma)*hSPn))*sum( (1/sqrt(2*pi)*exp( -(((ma-x)/hSPn)^2)/2 ) ) );
fkex2v <- Vectorize(fkex2);
fkex3 <- function(x) (1/(length(ma)*hCPn))*sum( (1/sqrt(2*pi)*exp( -(((ma-x)/hCPn)^2)/2 ) ) );
fkex3v <- Vectorize(fkex3);
fkex4 <- function(x) (1/(length(ma)*gCCMn3))*sum( (1/sqrt(2*pi)*exp( -(((ma-x)/gCCMn3)^2)/2 ) ) );
fkex4v <- Vectorize(fkex4);
plot(xp,fepv(xp),type='l',main='Resultados de los métodos bootstrap', col=1, ylab='Función de
densidad', xlab='Redshift(z)', ylim=c(0,0.8)); lines(xp,fkex1v(xp), col=2); lines(xp,fkex2v(
xp), col=3); lines(xp,fkex3v(xp), col=4); lines(xp,fkex4v(xp), col=5);
legend(2.6,0.68,c('Muestra completa', 'Referida a la normal', 'Bootstrap sin anchura pivote',
Bootstrap con anchura pivote', 'Bootstrap sin estimación pivote'), col=c(1,2,3,4,5), lty=c
(1,1,1,1,1), cex=0.8)

```

A.4. Código del estudio de la Subsección 4.2.1

```

load("../SDSS_cuasares/qso.Rdata"); #Lectura de los datos
mc <- qso$z; #Selección del redshift (parámetro a estudiar)
ma <- sample(mc, 100, replace = TRUE); #Construcción de una m.a.s. de tamaño 100
### DATOS GLOBALES ###
haux <- min(sd(mc), (quantile(mc,0.75)-quantile(mc,0.25))/1.34 ); #Anchura referida a la
distribución normal
hn <- ( (40*sqrt( pi ))^(1/5) )*(length(mc)^(-(1/5)))*haux;
fep <- function(x) (1/(length(mc)*hn))*sum( (3/4)* ( 1 - ((mc-x)/hn)^2 )*( abs((mc-x)/hn) <= 1 )
); fepv <- Vectorize(fep);
k <- c(2,5,20,50,75); #Valores de k
### MÉTODO DEL K-VECINO MÁS CERCANO ###
##Construcción de la distancia d_k
x1 <- seq(0,6,length=1e4); x <- seq(0,6,length=1e3);
d1 <- c(); d2 <- c(); d3 <- c(); d4 <- c(); d5 <- c(); #Almacenamiento de distancias
for (j in 1:length(x)) {
  daux1 <- abs( x[j] - ma ); #Construcción de las distancias
  daux2 <- c();
  for (m in 1:length(ma)) {
    daux2[m] <- sum( 1*(daux1 <= abs( x[j] - ma[m] ) ) );
  }
  d1[j] <- min( daux1[which( daux2 >= k[1] )] ); d2[j] <- min( daux1[which( daux2 >= k[2]
)] ); d3[j] <- min( daux1[which( daux2 >= k[3] )] ); d4[j] <- min( daux1[which(
daux2 >= k[4] )] ); d5[j] <- min( daux1[which( daux2 >= k[5] )] );
}
}

```

```

plot(x,d1,type='l', main=expression(paste("Distancia",d[k])), ylab=expression(paste(d[k],"(z)"
)), xlab='Redshift',ylim=c(0,2.4)); lines(x,d2,col=2); lines(x,d3,col=3); lines(x,d4,
col=4); lines(x,d5,col=5);
legend(0.8,2.2,c('k=2','k=5','k=20','k=50','k=75'), col=c(1,2,3,4,5), lty=c
(1,1,1,1,1), cex=0.8 );
##Construcción del estimador sin kernel
fnk1 <- function(y) k[1]/(2*length(ma)*y); fnk2 <- function(y) k[2]/(2*length(ma)*y); fnk3 <-
function(y) k[3]/(2*length(ma)*y); fnk4 <- function(y) k[4]/(2*length(ma)*y); fnk5 <-
function(y) k[5]/(2*length(ma)*y);
fnk1v <- Vectorize(fnk1); fnk2v <- Vectorize(fnk2); fnk3v <- Vectorize(fnk3); fnk4v <-
Vectorize(fnk4); fnk5v <- Vectorize(fnk5);
#Representación gráfica
plot(x1,fepv(x1),type='l',main='Método del k-vecino',ylab='Función de densidad',xlab='Redshift',
(z)',ylim=c(0,0.7)); lines(x,fnk2v(d2),col=2); lines(x,fnk3v(d3),col=3); lines(x,fnk4v(d4),
col=4); lines(x,fnk5v(d5),col=5);
legend(3.7,0.6,c('Muestra completa','k=5','k=20','k=50','k=75'), col=c(1,2,3,4,5), lty=
c(1,1,1,1,1), cex=0.8 );
##Construcción de la estimación kernel
kex <- function(z) (1/sqrt(2*pi))*exp( - (z^2))/2 ); kexv <- Vectorize(kex);
#Función auxiliar para recorrer distancias
sel2 <- function(y) d2[which( x == y )] ; sel3 <- function(y) d3[which( x == y )] ; sel4 <-
function(y) d4[which( x == y )] ; sel5 <- function(y) d5[which( x == y )] ;
#Construcción de las funciones densidad
fk2 <- function(y) {
  faux <- c(); #Vector de apoyo en el cálculo
  for (j in 1:length(ma)){
    faux[j] <- (1/ (length(ma)*sel2(y)) )*kex( (y-ma[j])/sel2(y) );
  }
  sum( faux );
}
fk3 <- function(y) {
  faux <- c(); #Vector de apoyo en el cálculo
  for (j in 1:length(ma)){
    faux[j] <- (1/ (length(ma)*sel3(y)) )*kex( (y-ma[j])/sel3(y) );
  }
  sum( faux );
}
fk4 <- function(y) {
  faux <- c(); #Vector de apoyo en el cálculo
  for (j in 1:length(ma)){
    faux[j] <- (1/ (length(ma)*sel4(y)) )*kex( (y-ma[j])/sel4(y) );
  }
  sum( faux );
}
fk5 <- function(y) {
  faux <- c(); #Vector de apoyo en el cálculo
  for (j in 1:length(ma)){

```

```

        faux[j] <- (1/ (length(ma)*sel5(y)) ) *kex( (y-ma[j])/sel5(y) );
    }
    sum( faux );
}
fk2v <- Vectorize(fk2); fk3v <- Vectorize(fk3); fk4v <- Vectorize(fk4); fk5v <- Vectorize(fk5)
;
#Representación gráfica
plot(x1,fepv(x1),type='l',main='Método del k-vecino',ylab='Función de densidad',xlab='Redshift_
(z)',ylim=c(0,0.7)); lines(x,fk2v(x),col=2); lines(x,fk3v(x),col=3); lines(x,fk4v(x),col=4)
; lines(x,fk5v(x),col=5);
legend(3.7,0.6,c('Muestra completa','k_5','k_20','k_50','k_75'), col=c(1,2,3,4,5), lty=
c(1,1,1,1,1), cex=0.8 );
### Comparativa método del k-vecino y estimación kernel con amplitud de ventana referida a la
normal
s <- min(sd(ma), (quantile(ma,0.75)-quantile(ma,0.25))/1.349 ); #Estimación de la desviación tí
pica
hn2 <- ((4/3)^(1/5))*(length(ma)^(-(1/5)))*s ; #Anchura referida a la distribución normal
fkn <- function(y) (1/(length(ma)*hn2))*sum( kexv( (y - ma)/hn2 ) ); fknv <- Vectorize(fkn);
kaux <- 2*hn2*length(ma)*max( fepv(x) ); k6 <- round(kaux); d6 <- c();
for (j in 1:length(x)) {
    daux1 <- abs( x[j] - ma ); #Construcción de las distancias
    daux2 <- c();
    for (m in 1:length(ma)) {
        daux2[m] <- sum( 1*(daux1 <= abs( x[j] - ma[m] ) ) );
    }
    d6[j] <- min( daux1[which( daux2 >= k6 )] );
}
sel6 <- function(y) d6[which( x == y )] ;
fk6 <- function(y) {
    faux <- c(); #Vector de apoyo en el cálculo
    for (j in 1:length(ma)){
        faux[j] <- (1/ (length(ma)*sel6(y)) ) *kex( (y-ma[j])/sel6(y) );
    }
    sum( faux );
}
fk6v <- Vectorize(fk6);
# Valor entero k asociado (variable)
kaux2 <- 2*hn2*length(ma)*fepv(x); k7 <- round(kaux2);
d7 <- c();
for (j in 1:length(x)) {
    daux1 <- abs( x[j] - ma ); #Construcción de las distancias
    daux2 <- c();
    for (m in 1:length(ma)) {
        daux2[m] <- sum( 1*(daux1 <= abs( x[j] - ma[m] ) ) );
    }
    d7[j] <- min( daux1[which( daux2 >= k7[j] )] );
}
}

```



```

sel7 <- function(y) d7[which( x == y )] ;
fk7 <- function(y) {
  faux <- c(); #Vector de apoyo en el cálculo
  for (j in 1:length(ma)){
    faux[j] <- (1/ (length(ma)*sel7(y)) ) *kex( (y-ma[j])/sel7(y) );
  }
  sum( faux );
}
fk7v <- Vectorize(fk7);
# Representación gráfica
plot(x1,fepv(x1),type='l',main='Método del k-vecino',ylab='Función de densidad',xlab='Redshift (z)',ylim=c(0,0.7)); lines(x,fk7v(x), col=2); lines(x,fk6v(x), col=3); lines(x,fk7v(x), col=4);
legend(3.7,0.6,c('Muestra completa',paste("h_=",round(hn2,3)),paste("k_=",round(k6)), 'k_ variable'), col=c(1,2,3,4), lty=c(1,1,1,1), cex=0.8 );

```

A.5. Código del estudio de la Subsección 4.2.2

```

load("../SDSS_cuasares/qso.Rdata"); #Lectura de los datos
mc <- qso$z; #Selección del redshift (parámetro a estudiar)
ma <- sample(mc, 1000, replace = TRUE); #Construcción de una m.a.s. de tamaño 1000
### DATOS GLOBALES ###
haux <- min(sd(mc), (quantile(mc,0.75)-quantile(mc,0.25))/1.34 ); #Anchura referida a la
  distribución normal
hn <- ( (40*sqrt( pi ))^(1/5) )*(length(mc)^(-(1/5)))*haux;
fep <- function(x) (1/(length(mc)*hn))*sum( (3/4)*( 1 - ((mc-x)/hn)^2 )*( abs((mc-x)/hn) <= 1 )
  ); fepv <- Vectorize(fep);
### MÉTODO KERNEL ADAPTADO ###
#Construcción del estimador pivote
hpaux <- min(sd(ma), (quantile(ma,0.75)-quantile(ma,0.25))/1.34 );
hp <- (((2/3)/(35*((1/6)^2)))^(1/5))*3*hpaux*length(ma)^(-(1/5)); #Anchura por sobre suavizació
  n
fp <- function(x) (1/(length(ma)*hp))*sum( ( 1 - abs((ma-x)/hp) )*( abs((ma-x)/hp) <= 1 ) );
fpv <- Vectorize(fp); #Estimación pivote con función Triangular
evma <- fpv(ma); #Evaluación de la estimación pivote en los valores muestrales de ma
mug <- exp((1/length(ma))*(sum( log(evma) ))); #Cálculo de la media geométrica
alfa <- c(0, 0.25, 0.5, 0.75, 1); #Array de valores para el parámetro de sensibilidad
#Construcción de los factores de escala para la amplitud de ventana
lambda1 <- (evma/mug)^(-alfa[1]); lambda2 <- (evma/mug)^(-alfa[2]); lambda3 <- (evma/mug)^(-
  alfa[3]); lambda4 <- (evma/mug)^(-alfa[4]); lambda5 <- (evma/mug)^(-alfa[5]);
hest <- ((4/3)^(1/5))*hpaux*((length(ma))^(1/5)); #Referida a la distribución normal con
  kernel normal
kexp <- function(x) (1/sqrt(2*pi))*exp( - (x^2)/2 ); #Kernel exponencial
#Función de densidad para cada array de factores de escala

```

```

fest1 <- function(y) {
  aux1 <- c();
  for (k in 1:length(ma)) {
    aux1[k] <- (1/lambda1[k])*kexp( (y-ma[k])/(hest*lambda1[k]) );
  }
  (1/(length(ma)*hest))*sum(aux1)
}

fest2 <- function(y) {
  aux2 <- c();
  for (k in 1:length(ma)) {
    aux2[k] <- (1/lambda2[k])*kexp( (y-ma[k])/(hest*lambda2[k]) );
  }
  (1/(length(ma)*hest))*sum(aux2)
}

fest3 <- function(y) {
  aux3 <- c();
  for (k in 1:length(ma)) {
    aux3[k] <- (1/lambda3[k])*kexp( (y-ma[k])/(hest*lambda3[k]) );
  }
  (1/(length(ma)*hest))*sum(aux3)
}

fest4 <- function(y) {
  aux4 <- c();
  for (k in 1:length(ma)) {
    aux4[k] <- (1/lambda4[k])*kexp( (y-ma[k])/(hest*lambda4[k]) );
  }
  (1/(length(ma)*hest))*sum(aux4)
}

fest5 <- function(y) {
  aux5 <- c();
  for (k in 1:length(ma)) {
    aux5[k] <- (1/lambda5[k])*kexp( (y-ma[k])/(hest*lambda5[k]) );
  }
  (1/(length(ma)*hest))*sum(aux5)
}

#Vectorización de las funciones definidas
fev1 <- Vectorize(fest1); fev2 <- Vectorize(fest2); fev3 <- Vectorize(fest3); fev4 <- Vectorize
(fest4); fev5 <- Vectorize(fest5);

### REPRESENTACIÓN GRÁFICA ###
x <- seq(0,6, length=1e4); #Array para la representación gráfica
plot(x,fev(x),type='l',main='Método kernel adaptado',ylab='Función densidad', xlab='Redshift'
z',ylim=c(0,0.6)); lines(x,fev1(x),col=2); lines(x,fev2(x),col=3) ; lines(x,fev3(x),col=4)
; lines(x,fev4(x),col=5) ; lines(x,fev5(x),col=6)
legend(3.6,0.5,c('Muestra completa',expression(paste(alpha,"_=","_")),expression(paste(alpha,"_="
_")),expression(paste(alpha,"_=","_")),expression(paste(alpha,"_=","_")),expression
(paste(alpha,"_=","_")),col=c(1,2,3,4,5,6),lty=c(1,1,1,1,1,1),cex=0.8)

```

Bibliografía

- [1] M. B. Bell y D. McDiarmid. *Six Peaks Visible in the Redshift Distribution of 46,400 SDSS Quasars Agree with the Preferred Redshifts Predicted by the Decreasing Intrinsic Redshift Model*, The Astrophysical Journal, Vol. 648, pp. 140-147, 2006.
- [2] R. Cao, A. Cuevas y W. G. Manteiga. *A comparative study of several smoothing methods in density estimation*, Computational Statistics & Data Analysis, Vol. 17, pp. 153-176, 1994.
- [3] Y.-C. Chen. *Lecture 6: Density Estimation: Histogram and Kernel Density Estimator*, Introduction to Nonparametric Statistics, University of Washington, 2018.
- [4] V. A. Epanechnikov. *Non-Parametric Estimation of a Multivariate Probability Density*, Theory of Probability & Its Applications, Vol. 14, No. 1, pp. 153-158, 1969.
- [5] E. D. Feigelson y G. J. Babu. *Modern Statistical Methods for Astronomy*, Cambridge University Press, New York, UK, 2012.
- [6] P. Hall y J. S. Marron. *Estimation of integrated squared density derivatives*, Statistics & Probability Letters, Vol 6, pp. 109-115, 1987.
- [7] P. Hall, S. J. Sheather, M. C. Jones y J. S. Marron. *On Optimal Data-Based Selection in Kernel Density Estimation*, Biometrika, Vol. 78, No. 2, pp. 263-269, 1991.
- [8] W. Härdle, M. Müller, S. Sperlich y A. Werwatz. *Nonparametric and Semiparametric Models, An Introduction*, Springer, 2004.
- [9] M. C. Jones, J. S. Marron y B. U. Park. *A simple root n bandwidth selector*, The Annals of Statistics, Vol. 19, No. 4, pp. 1919-1932, 1991.
- [10] K. G. Karlsson. *On the Existence of Significant Peaks in the Quasar Redshift Distribution*, A&A, Vol. 58, pp. 237-240, 1977.
- [11] Y. P. Mack y M. Rosenblatt. *Multivariate k -Nearest Neighbor Density Estimates*, Journal of Multivariate Analysis, Vol. 9, pp. 1-15, 1979.
- [12] B. U. Park y J. S. Marron. *Comparison of Data-Driven Bandwidth Selectors*, Journal of the American Statistical Association, Vol. 85, No. 409, pp. 66-72, 1990.
- [13] E. Parzen. *On Estimation of a Probability Density Function and Mode*, The Annals of Mathematical Statistics, Vol. 33, No. 3, pp. 1065-1076, 1962.

- [14] M. Rosenblatt. *Remarks on Some Nonparametric Estimates of a Density Function*, Vol. 27, No. 3, pp. 832-837, 1956.
- [15] O. Y. Savchuk, J. D. Hart y S. J. Sheather. *Indirect Cross-Validation for Density Estimation*, Journal of the American Statistical Association, Vol. 105, pp. 415-423, 2010.
- [16] D. W. Scott. *On Optimal and Data-Based Histograms*, Biometrika, Vol. 66, No. 3, pp. 605-610, 1979.
- [17] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability, Londres, UK, 1986.
- [18] G. R. Terrell. *The Maximal Smoothing Principle in Density Estimation*, Journal of the American Statistical Association, Vol. 85, No. 410, pp. 470-477, 1990.
- [19] B. A. Turlach. *Bandwidth selection in kernel density estimation: A review*, CORE and Institut de Statistique, 1993.
- [20] A. Z. Zambom y R. Dias. *A Review of Kernel Density Estimation with Applications to Econometrics*, ArXiv:1212.2812v1, 2012.
- [21] Y.-Y. Zhou, Z.-G. Deng y Z.-L. Zou. *Quasar redshift distribution and selection effects*, Chinese Astronomy and Astrophysics, Vol. 9, pp. 20-26, 1985.