Biblioteca Digital
F C E N - U B A

BIBLIOTECA CENTRAL LUIS F LELOIR
BIBLIOTECA CENTRAL LELOIR
FACULTAD DE CIENCIAS EXACTAS Y NATURALES UBA

Tesis de Posgrado

Estimación de la densidad relativa

Iturmendi, Nilda Susana

Tesis presentada para obtener el grado de Magister de la Universidad de Buenos Aires en el área de Estadística Matemática de la Universidad de Buenos Aires

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:
Iturmendi, Nilda Susana. (). Estimación de la densidad relativa. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_3669_Iturmendi.pdf

Cita tipo Chicago:
Iturmendi, Nilda Susana. "Estimación de la densidad relativa". Tesis de Magister. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. .
http://digital.bl.fcen.uba.ar/Download/Tesis/Tesis_3669_Iturmendi.pdf

EXACTAS UBA
Facultad de Ciencias Exactas y Naturales

UBA
Universidad de Buenos Aires

Dirección: Biblioteca Central Dr. Luis F. Leloir, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. Intendente Güiraldes 2160 - C1428EGA - Tel. (++54 +11) 4789-9293

Contacto: digital@bl.fcen.uba.ar

Universidad de Buenos Aires
Facultad de Ciencias Exactas y Naturales

Estimación de la densidad relativa

Autor: Nilda Susana Iturmendi
Lugar de Trabajo: Universidad Nacional del Sur

Director: Dra. Graciela L. Boente Boente

**Tesis para optar al título de
Magister en Estadística Matemática**

13669

Junio de 2003

Agradecimientos

Quisiera expresar mi agradecimiento a todas las personas que estuvieron a mi lado, apoyándome, ayudándome en mis tareas, expresando su cariño y siempre alentándome.

En primer lugar a mi directora de tesis, Dra. Graciela Boente, que me recibió, me acompañó y me guió, con mucho afecto y profesionalismo.
A mis compañeros de maestría, por la amistad que me brindaron.
A mis profesores, por su paciencia y consideración.
A mis compañeros en Bahía Blanca, por su aliento constante.
Al Consejo Departamental y al Director del Dto. de Matemática de la Univ. Nac. del Sur, por haberme otorgado ayuda económica para realizar los viajes a Bs. As.
A Gabriela Cicileo, por su hospitalidad.
A mi hijos Claudina y Sebastián, por su presencia, cariño y apoyo.

Indice General

Pag.

Resumen 3

Abstract..... 4

Introducción..... 5

Capítulo 1. Estimadores de la densidad y criterios de bondad.

1-Estimadores de la densidad..... 9

2-Algunos criterios para medir el error al estimar la densidad. 11

2.1-Error cuadrático medio..... 12

2.2-Métricas en el espacio de las funciones de densidad... 13

2.3-Criterios globales..... 13

2.4-Algunas observaciones..... 14

Capítulo 2. Métodos de estimación no-paramétrica de la densidad.

1-Histograma..... 15

1.1-Error cuadrático medio integrado (MECI) del estimador histograma..... 17

1.2-Validación cruzada sesgada para estimar la ventana óptima asintótica..... 22

1.3-Validación cruzada insesgada para estimar la ventana óptima asintótica..... 22

2-Polígono de frecuencia..... 24

3-Histograma corrido promediado (HCP)..... 29

3.1-El HCP límite como estimador de núcleo..... 33

4-Estimador basado en núcleos..... 34

5-Estimador de vecinos más cercanos..... 41

6-Estimador de núcleo variable..... 43

7-Estimador de máxima verosimilitud penalizada..... 44

Capítulo 3. Elección de la abertura de ventana en estimadores de núcleos.

1-Minimizando la expresión del MECI..... 45

2-Validación cruzada de mínimos cuadrados (VCMC)..... 47

3-Validación cruzada de máxima verosimilitud penalizada(VCMVP). 49

Capítulo 4. Densidad relativa

1-Planteo del problema..... 51

2-Estimación de la densidad relativa..... 55

2.1-Métodos basados en núcleos..... 55
2.2-Métodos de vecinos más cercanos con núcleos..... 56
3-Estudio de Monte Carlo..... 60
3.1-Estimador de núcleos..... 61
3.2-Estimador de vecinos más cercanos..... 64
4-Resultados de las simulaciones..... 66
4.1-Caso 1..... 66
4.2-Caso 2..... 73
4.3-Caso 3..... 79
4.4-Conclusiones..... 85
4.5-Ejemplos de aplicación..... 88

Bibliografía..... 94

Apéndice 1. Programas de computación..... 98

Apéndice 2. Datos de biomasa aérea de gramíneas.....134

Resumen.

Dadas dos variables aleatorias independientes, X_0 y X, que llamaremos de *referencia* y de *comparación* respectivamente, se define la *variable relativa* como $R = F_0(X)$, siendo F_0 la función de distribución acumulada de la variable X_0. Existe abundante bibliografía sobre métodos para estimar la densidad de una variable a partir de observaciones independientes.

En esta tesis consideramos el problema de estimar la densidad de la variable R, llamada densidad relativa, a partir de muestras independientes, $X_{0,1},\ldots X_{0,n}$ y $X_1,\ldots X_m$, de ambas poblaciones. Para ello construímos predictores de la variable relativa obtenidos a partir de un estimador de la función F_0 calculada en observaciones de la variable X.

El conocimiento de la densidad de R, nos permite representar las similitudes o diferencias entre las distribuciones de X y X_0. Para estimar la densidad relativa hemos combinado dos estimadores de la función de distribución, la función de distribución empírica y un estimador suavizado de la misma, con estimadores de la densidad basados en *núcleos* y basados en *vecinos más cercanos con núcleos.* Respecto de la elección del parámetro de suavizado hemos considerado la ventana óptima y distintos procedimientos de validación cruzada.

Mediante un estudio de simulación, se estudió el comportamiento para muestras pequeñas de los estimadores de la densidad relativa propuestos. Como medida del error de estimación se consideró una aproximación del *error cuadrático medio integrado.* El estimador basado en núcleos utilizando una ventana de validación cruzada sesgada y el que utiliza la ventana óptima, son los procedimientos con el mejor comportamiento.

Abstract

Let X_0 and X be given independent random variables, called *reference variable* and *comparison variable*, respectively. The *relative variable* is defined by $R = F_0(X)$, where F_0 is the accumulated distribution function of the variable X_0. There is an extense bibliography dedicated to the estimation of the density of a given variable from the knowledge of independent observations.

In this thesis, we will deal with the question of estimating the density of the variable R, called *relative density,* from the knowledge of independent samples, $X_{01},\ldots,X_{0n}$, and $X_1,\ldots,X_n$, of the two populations under consideration. To achieve this, we build predictors of the relative variable, obtained from an estimator of the function F_0, calculated at observations of the variable X.

Knowledge of the density of R provides a way to represent similitudes and differences between the distributions of X_0 and X. To estimate the relative density, we have used two estimators of the distribution function, the empiric distribution function, and a smoothed estimator of it, with estimators based on *kernels,* and on *nearest neighbours with kernels.* Regarding the choice of the soothing parameter, we have considered the *optimal window* and several *cross-validation proceedures.*

By performing a simulation study, we have studied the behaviour for small samples of the proposed estimators of relative density. We have measured the error of estimation through an approximation of the mean integrated squared error.

The best perfomance of the error corresponds to the kernel method with biased cross-validation window, and the optimal window procedure comes in the second place.

Introducción

La función de densidad es un concepto fundamental en estadística que nos permite calcular probabilidades y valores característicos de una variable aleatoria.

Cuando es desconocida, existen numerosos procedimientos para estimarla. Estos se han agrupado en dos grandes ramas, la paramétrica y la no-paramétrica. En la primera se debe especificar la familia a la cual pertenece la densidad que se quiere investigar, mientras que en la segunda no se requiere que verifique un modelo dado y puede presentar cualquier forma.

Es sabido que los métodos paramétricos alcanzan una eficiencia mayor que los no-paramétricos cuando se cumple el modelo predeterminado, aunque son poco eficientes cuando éste no se verifica.

En este trabajo consideramos estimadores no-paramétricos de la densidad.

El estimador de la densidad f, que notamos f_n, se calcula a partir de una muestra de n observaciones independientes de la variable aleatoria considerada.

En el Capítulo 1 nos referimos a los conceptos mencionados y también a presentar diferentes medidas de error para evaluar la bondad del estimador.

La primer medida de error que presentamos es el error cuadrático medio (ECM), que es una medida de discrepancia puntual entre la densidad y el estimador, calculado en un punto dado, y se obtiene promediando sobre todas las muestras de tamaño n.

Como medidas de discrepancia global, para los valores donde se estima la densidad, se presentan aquellas basadas en las métricas L_∞, L_1 y L_2, las cuales dan como resultado variables aleatorias función de la muestra.

Al promediar sobre todas las muestras, se obtienen medidas de error globales. Por ejemplo, asociado al error de la métrica L_2, tenemos el error cuadrático medio integrado (MECI) que coincide con la integral del error cuadrático medio (IECM). Otra medida global es la propuesta por Devroye y Györfi (1985) que considera la métrica L_1 en lugar de la L_2.

Por último, al final de ese capítulo, mencionamos algunos resultados sobre el comportamiento de estos errores.

Seguidamente, en el Capítulo 2, se presentan los métodos de estimación no-paramétrica de la densidad, comenzando por el histograma. Discutimos el problema de selección de la ventana o parámetro de suavizado, que figurará también en los próximos estimadores, el cual controla la forma suave o irregular del estimador. Transcribimos los resultados de consistencia asintótica en media cuadrática y la expresión asintótica del MECI (AMECI) con el valor de ventana óptimo que lo minimiza. Este valor de ventana tiene su expresión en función de la derivada de la densidad desconocida, por lo cual nos referimos a dos métodos de aproximación de este valor óptimo. Ellos son, el método de validación cruzada sesgada y validación cruzada insesgada.
Considerando las discontinuidades que contiene el histograma se presenta posteriormente el polígono de frecuencias que elude ese inconveniente. La expresión asintótica del MECI posibilita encontrar una ventana óptima que la minimiza, la cual depende de la derivada segunda de la densidad. El método de validación cruzada sesgada provee una aproximación a esa ventana.
Teniendo en cuenta que el polígono se ve afectado por el valor inicial de la grilla de intervalos, presentamos un tercer estimador no paramétrico, llamado histograma corrido promediado, que no tiene ese inconveniente. Igual que en los casos anteriores, se muestra el valor del MECI, una expresión asintótica de él y la ventana óptima que minimiza dicha expresión. En el caso de una densidad normal, el método no-paramétrico más eficiente entre los considerados, resulta ser el histograma corrido promediado, siguiéndolo el polígono y muy lejos el histograma.
Se introduce el estimador de núcleos, el cual puede verse como límite del histograma corrido promediado. Este estimador fue estudiado por numerosos autores, comenzando por Rosenblatt (1956), Parzen (1962) y Cencov (1962) y siguiendo hasta la actualidad con autores como Hall (1983), Devroye y Györfy (1985), Marron (1985), entre otros.
Este estimador se basa en una suma promediada de los valores de una función K centrada en cada valor de la muestra, dependiendo de un parámetro de suavizado. A esta función K se la denomina núcleo y el estimador hereda sus propiedades de suavidad.
Se obtiene la expresión del MECI asintótico y el valor de ventana que lo minimiza.

El parámetro de suavizado no depende del punto en que se calcula el estimador, a diferencia del estimador de vecinos más cercanos que se introduce posteriormente.
El estimador de vecinos más cercanos, tiene una expresión similar al estimador de núcleos, excepto en que la ventana, o parámetro de escala del núcleo K, varía de un punto a otro. Esta ventana se calcula como la distancia del punto en que se estima la densidad, al k-ésimo vecino más cercano entre los valores observados de la variable, siendo k un número natural menor o igual al tamaño de muestra.
Por último mencionamos el estimador de núcleo variable y el de máxima verosimilitud penalizada.
Teniendo en cuenta que en este trabajo hemos elegido, como estimadores no-paramétricos de la densidad, el de núcleos y el de vecinos más cercanos, en el Capítulo 3 hacemos una reseña de algunos métodos para encontrar la abertura de ventana h, que luego usaremos en las estimaciones. El primero consiste en hallar el valor de h, la abertura de ventana, que minimice una aproximación del MECI o de su aproximación asintótica; esta abertura se denomina óptima, y la indicaremos h_{opt}. En el Capítulo 4, donde se presentan los resultados de simulación Monte Carlo, esta ventana se calcula explícitamente para distribuciones consideradas.
Otro método automático para encontrar la ventana es el de validación cruzada de mínimos cuadrados (VCMC), sugerido por Rudemo (1982) y Bowman (1984). A partir de un desarrollo para su cálculo se deriva una modificación debida a Scott y Terrell (1987) llamada de mínimos cuadrados sesgada.
El tercer método considerado para encontrar la ventana es el de validación cruzada de pseudo-máxima verosimilitud, en donde h maximiza la función de pseudo-verosimilitud.
En el Capítulo 4 presentamos la variable relativa, que se define como la función de distribución acumulada de una variable de referencia, calculada sobre una variable de comparación. La función de densidad de esta variable relativa, llamada densidad relativa, permite representar las posibles diferencias entre las funciones de densidad de dos variables, además de las de locación y escala.
Los resultados de un estudio de Monte Carlo se muestran en el Capítulo 4, donde estimamos la densidad relativa utilizando método

de núcleos y de vecinos más cercanos, usando las ventanas de validación cruzada y óptimas. Simulamos tres pares de variables de referencia y comparación. Como medida de error de estimación se usó una aproximación del error cuadrático medio integrado. Aplicamos el método que mostró mejor perfomance del error, a dos ejemplos de datos reales.

Finalmente, en el Apéndice se presentan los programas de computación realizados.

Capítulo 1

Estimadores de la densidad y criterios de bondad.

1-Estimadores de la densidad.

La función de densidad de probabilidad f, de una variable aleatoria univariada continua X, es un concepto fundamental en estadística. Está caracterizada por verificar las siguientes dos condiciones,

$$a)\ f(x) \geq 0 \quad \forall\ x \in R,$$

$$b)\ \int f(x)\ dx = 1.$$

Su conocimiento nos permite, entre otras aplicaciones, acceder al valor de los parámetros de la población y al cálculo de probabilidades de la forma $P(a<X<b)$ a través de la igualdad

$$P(a<X<b) = \int_{(a,b)} f(x)\ dx,\ \forall\ a<b.$$

En muchas situaciones, la distribución de las observaciones es desconocida y el objetivo es estimarla a partir de una muestra aleatoria, $X_1,\ldots,X_n$, con densidad común f.
El camino para lograrlo se ha dividido en dos, según la metodología que se aplique para ello: la estimación paramétrica y la estimación no paramétrica, siendo ambas totalmente diferentes en su enfoque.
Como se menciona en Scott ((1992), pag.33), R.A. Fisher y Karl Pearson han estado embarcados en un interesante debate sobre aspectos de la estimación paramétrica, con Pearson argumentando a favor de la contrapuesta.
En el enfoque paramétrico, se supone de antemano que la densidad desconocida sigue un modelo, el cual está especificado por el hecho de pertenecer a una familia de densidades $\mathfrak{F}$, caracterizada por un parámetro vectorial θ, o sea,

$$\mathfrak{F} = \{\ f(\bullet,\theta)\ /\ f \text{ función de densidad, } \theta \in \Theta \subseteq R^k\ \}.$$

Por ejemplo, la familia normal, corresponde a

$$f(x,\theta) = f(x,(\mu,\sigma^2)) = \exp(-(x-\mu)^2/2\sigma^2)/(2\Pi\sigma^2)^{1/2},$$

siendo $\theta = (\mu,\sigma^2) \in R x R^+ \subseteq R^2$, mientras que la familia exponencial está dada por

$$f(x,\theta) = \theta \exp(-(\theta x)) \quad \text{para } x>0,\ \theta \in R^+ \subseteq R.$$

La idea es estimar el parámetro θ por un valor θ_n, función de la muestra, y reemplazar esta estimación en la expresión de f, obteniendo así la aproximación buscada $f_n(x) = f(x,\theta_n)$.
Fisher (1922, 1932) reconoció dos pasos en la estimación paramétrica: el primero es la especificación del modelo y el segundo la estimación de la densidad, una vez determinado el mismo. En esta última etapa el objetivo es encontrar estimaciones óptimas, lo cual trae como consecuencia el desarrollo de distintos métodos para estimar el parámetro θ. Con respecto al primer paso, donde corresponde determinar la familia $\mathfrak{F}$, advirtió que era un problema de la estadística práctica basado en la experiencia y que una especificación equivocada del modelo podría ser detectada con un test a posteriori.
Sin embargo, una incorrecta especificación de la familia $\mathfrak{F}$ puede producir un sesgo considerable en la estimación, el cual no puede ser corregido con el hecho usual de aumentar el tamaño de muestra. Por otra parte, un test de bondad de ajuste generalmente rechaza modelos razonables cuando la muestra es grande.
Hay un consenso generalizado en que métodos paramétricos son eficientes cuando la elección de $\mathfrak{F}$ es acertada pero, los estimadores pueden resultar ineficientes, aún en el caso en que existan sólo pequeñas desviaciones respecto al modelo real.
La idea sería sacrificar un pequeño porcentaje de optimalidad a fin de obtener una mayor insensibilidad a incorrectas especificaciones. Los procedimientos robustos dan una solución a este enfoque.
Desde otro punto de vista, podemos considerar métodos de estimación no paramétricos, que eliminan la necesidad de una completa especificación del modelo, equilibrando su menor

eficiencia con el mayor riesgo de cometer una incorrecta elección en el primer paso de la estimación paramétrica.
Aunque se supondrá que la variable X tiene una función de densidad f, se permitirá que los datos hablen por sí mismos, sin restringirlos de antemano a pertenecer a una familia de densidades paramétricas. De esta forma, la función f será estimada en su totalidad por f_n y no a través de los parámetros que figuran en su expresión.
Podemos resaltar que la creciente importancia del uso de computadoras en investigación estadística, la disponibilidad de software calificado y la conciencia general de las ventajas de gráficos de alto nivel han hecho que la estimación no paramétrica se haya vuelto popular. Observamos su efectividad al explorar sesgo, multimodalidad o comportamiento en las colas de la distribución, ofreciendo una visión más flexible que la dada por los métodos paramétricos tradicionales obligados a cumplir un modelo determinado. Por otra parte, peculiaridades estadísticas pueden ser observadas a través de gráficos de estimadores no paramétricos.
En este trabajo, a fin de comparar dos muestras, hemos elegido una aproximación basada en métodos de estimación no paramétricos que nos permiten un amplio espectro de densidades a estimar.

2-Algunos criterios para medir el error al estimar la densidad.

A partir de las observaciones, distintos estimadores pueden ser propuestos para aproximar el parámetro o función de interés.
El objetivo es elegir el mejor entre todos ellos, o sea aquel que está más cerca del valor que se quiere estimar.
El concepto de cercanía está ligado al concepto de error en la estimación: cuanto más cercana sea la aproximación, menor es el error que se comete en la estimación.
Existen varios criterios para medir la discrepancia entre el estimador de la densidad, que indicaremos por f_n, y la verdadera densidad f. La preferencia por uno de estos criterios estará basada en las características que surgen de su desarrollo teórico y en el propósito para el cual se busca el estimador. A su vez, cada uno de ellos puede ser optimizado eligiendo el estimador

apropiado, lo cual se logra cuando la medida del error es lo menor posible.
Vamos a definir algunas medidas de error, utilizadas especialmente en estimación no paramétrica.

2.1-Error cuadrático medio.

Cuando se estima un parámetro θ, por un estadístico insesgado θ_n, la medida de error usual es la varianza del estadístico,

$$\mathrm{var}(\theta_n) = E[\theta_n - \theta]^2.$$

Si el estadístico es sesgado, entonces la expresión $E[\theta_n - \theta]^2$ se denomina error cuadrático medio (ECM) y puede escribirse como

$$ECM(\theta_n) = E[\theta_n - \theta]^2 = \mathrm{var}(\theta_n) + \{E[\theta_n] - \theta\}^2 = \mathrm{var}(\theta_n) + \mathrm{sesgo}^2(\theta_n).$$

De la misma forma, podemos calcular el error cuadrático medio, cuando estimamos $f(x)$ por medio del estadístico $f_n(x)$,

$$ECM[f_n(x)] = E[f_n(x)-f(x)]^2.$$

Esta es una medida de discrepancia puntual, o sea sólo para el punto x donde se calcula la función f_n, que se obtiene promediando sobre todas las muestras.
En este caso, el ECM puede expresarse también en función de la varianza de $f_n(x)$ y del sesgo,

$$ECM[f_n(x)] = \mathrm{var}(f_n(x)) + \mathrm{sesgo}^2(f_n(x)).$$

Si $ECM[f_n(x)] \to 0$ cuando $n \to \infty$ para cada $x \in R$ se dice que f_n es un estimador consistente, en media cuadrática, de f.
Es posible medir el error de estimación para todos los valores x posibles. Estas medidas se llaman medidas de error global y existen varios criterios para determinarlas. Mencionaremos los criterios L_∞, L_1 y L_2. Estos criterios nos permitirán obtener una medida de error que no variará con el punto x, pero dependerá de la muestra obtenida.

2.2-Métricas en el espacio de las funciones de densidad.

A fin de definir un criterio global del error, podemos considerar la norma L_∞, con una medida de discrepancia dada por

$$\sup_{x \varepsilon R} |f_n(x)-f(x)|.$$

En el otro extremo del espectro de las normas, está la norma L_1, generando una medida de error global

$$\int_R |f_n(x)-f(x)|dx.$$

Ninguno de estos dos últimos criterios es tan fácil de tratar matemáticamente como el correspondiente a la norma L_2, llamado error cuadrático integrado (ECI),

$$ECI = \int_R [f_n(x)-f(x)]^2dx.$$

Observemos que todos ellos dependen del estimador elegido f_n, de la densidad desconocida f, del tamaño de muestra n, y también de la muestra obtenida, ya que no se han calculado esperanzas.

2.3- Criterios globales.

Las métricas anteriores definen variables aleatorias, por lo tanto podemos quitar la dependencia sobre la muestra, promediando sobre todas las muestras, lo cual da lugar a distintas medidas de error globales. Entre ellas, es de destacar el error cuadrático medio integrado (MECI), que es la media del ECI,

$$MECI = E(ECI) = E\left(\int_R [f_n(x)-f(x)]^2dx\right).$$

Aplicando el Teorema de Fubini, podemos intercambiar el operador integral con la esperanza, obteniendo

$$MECI = E(ECI) = E\left(\int_R [f_n(x)-f(x)]^2dx\right)$$

$$= \int_R E\left([f_n(x)-f(x)]^2\right)dx = \int_R ECM[f_n(x)]dx = IECM,$$

donde IECM indica la integral del error cuadrático medio. Luego, el MECI tiene dos interpretaciones: es una medida del error global promedio y del error cuadrático medio puntual integrado.
Devroye y Györfi (1985) consideran un enfoque análogo basado en la distancia L_1,

$$E\left(\int_R |f_n(x)-f(x)|dx\right).$$

Cuando se considera el resultado asintótico del MECI, al tender n a infinito, estamos calculando una medida de error asintótico que notaremos AMECI.

2.4-Algunas observaciones.

Farrell (1972) probó que la mejor convergencia asintótica del MECI es del orden de $n^{-4/5}$ y Boyd y Steele (1978) probaron que no existe ningún estimador f_n que genere un MECI con una convergencia que mejore al orden n^{-1}, aún cuando f fuera una densidad normal.
Se han comparado también los errores generados por las normas L_1 y L_2. Una desventaja con la medida L_2, es que le da poca importancia al comportamiento en las colas, lo que no ocurre con la medida L_1.
Devroye y Györfi (1985) han desarrollado un importante trabajo teórico basado en L_1 pero la simplicidad del ECI y su buen comportamiento en trabajos de aplicación hace que la elección se incline hacia él.
Algunos resultados asintóticos (Hall y Wand (1988), Scott y Wand (1991)) demuestran que las diferencias entre los criterios L_1 y L_2 son razonablemente escasas excepto en situaciones extremas. Cabe agregar que los estimadores que optimizan ambos criterios son similares.
En este trabajo usaremos como medida de error del estimador, a una aproximación del MECI.

Capítulo 2

Métodos de estimación no-paramétrica de la densidad

En este capítulo, presentaremos las familias más conocidas de estimadores no paramétricos de la densidad. Comenzaremos con el histograma, aumentando la complejidad hasta llegar a los estimadores de vecinos más cercanos.
La estructura del histograma es útil para entender los pasos generales de la teoría y práctica no-paramétrica. Desde el cálculo del error cuadrático hasta los algoritmos de validación cruzada, estos tópicos se ven más simples en el contexto de este primer estimador no paramétrico y será más sencillo comprenderlos cuando recordemos la teoría de los demás estimadores.
Por otra parte, la construcción de estimadores cada vez más complejos surge naturalmente y paso a paso, como veremos.

1-Histograma.

La primera vez que se menciona un histograma fue en el año 1662, en el trabajo de J. Graunt, "Natural and Political Observations Made upon the Bills of Mortality", y hasta la publicación de una nota de Herbert Sturges (1926), no se presentaron trabajos para diseñarlos.
Allí aparece lo que hoy conocemos como regla de Sturges, para determinar el número de intervalos del histograma, que está dada por la expresión: $k = 1+\log_2 n$, bajo la suposición de n datos normales. Si la distribución es asimétrica o muy concentrada, se aconseja aumentar el número de intervalos.
Dada una muestra aleatoria $X_1,\ldots,X_n$, y un conjunto de intervalos disjuntos que cubren el soporte de la función f, el histograma $f_n(x)$ puede definirse como el porcentaje de observaciones que pertenecen al mismo intervalo de clase que x, dividido por la longitud de dicho intervalo. Más precisamente, si los intervalos están definidos por una red de puntos $\{t_k \ / \ k \in Z\}$ y si $B_k = [t_k, t_{k+1})$, siendo n_k la frecuencia de la muestra en dicho intervalo, obtenemos $f_n(x) = n_k/(n(t_{k+1} - t_k))$ si $x \in B_k$.
A fin de que las frecuencias n_k sean comparables, o bien los intervalos deberían tener todos la misma longitud, o bien se debe

definir el histograma como lo hemos hecho, tomando la frecuencia relativa dividida por la longitud de cada intervalo considerado. Si el soporte de f fuera infinito, se toma el de la muestra.
En el caso de intervalos de igual longitud, el histograma estaría completamente determinado por dos parámetros: el valor inicial t_0 y la amplitud del intervalo h, siendo $B_k = [t_0+kh,\ t_0+(k+1)h)$, $k \in Z^+$, $t_0 \leq X_{(1)}$, donde $X_{(1)}$ denota el primer estadístico de orden en la muestra aleatoria.
El histograma se define, en este caso, como

$$f_n(x) = (n.h)^{-1} \sum_{k=0}^{\infty} I_{B_k}(x) n_k, \qquad (1)$$

donde I_A es la función característica del conjunto A, definida por

$$I_A(t) = 0 \quad \text{si } t \notin A,$$

$$I_A(t) = 1 \quad \text{si } t \in A.$$

El parámetro h se conoce como parámetro de "suavizado" ya que controla la forma suave (h grande) o irregular (h pequeño) del gráfico del estimador.
Los gráficos de la Fig.1 indican el aumento de suavizado del estimador al aumentar el valor de h; se simularon 100 datos de una población normal de media 2 y desvío 1, el mínimo de la muestra fue el valor -0.4925 y el máximo 4.225.

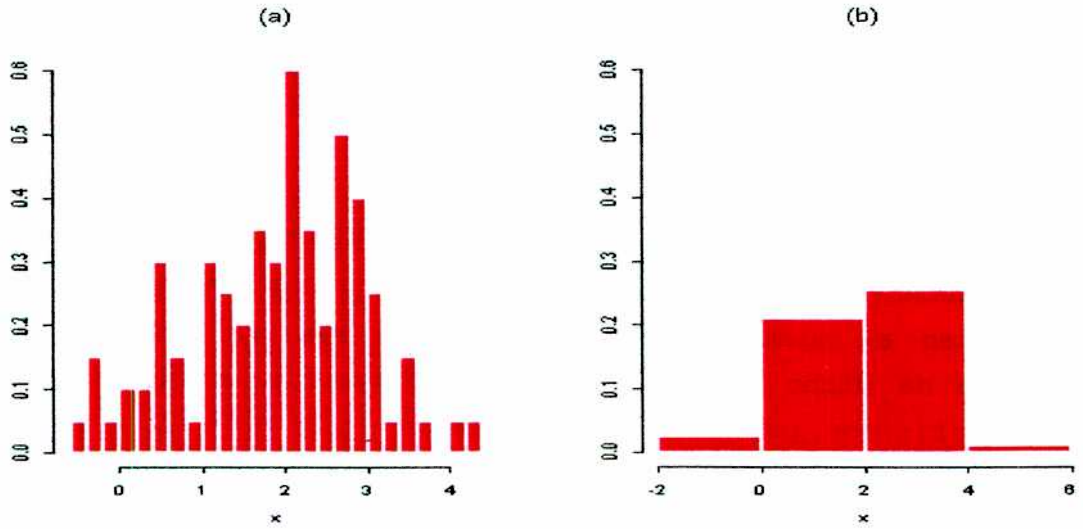

Fig. 1. Histograma de 100 observaciones de X~N(2,1) (a) h=0.2,(b) h=2.

Observemos que $f_n(x)$ es constante en cada intervalo. Es fácil probar que el histograma es el estimador de máxima verosimilitud de f, entre las posibles funciones escaleras, constante en cada uno de los intervalos (Härdle, (1991), pag.9).

1.1-Error cuadrático medio integrado del estimador histograma.

El análisis de la variable aleatoria $f_n(x)$, $x \in B_k$, es muy simple una vez que se ha comprendido que n_k es una variable aleatoria binomial,

$$n_k \sim B(n,p_k) \quad \text{siendo} \quad p_k = \int_{B_k} f(t)dt. \tag{2}$$

Luego,

$$E(f_n(x)) = E(n_k/(r,h)) = p_k/h ,$$

y

$$Var(f_n(x)) = Var(n_k)/(nh)^2 = p_k(1-p_k)/(nh^2).$$

Estudiemos el ECM cuando x pertenece al intervalo B_k. Por el Teorema del Valor Medio,

$$p_k = \int_{B_k} f(t)dt = h\ f(\xi_k), \qquad \xi_k \in B_k.$$

Luego,

$$Var(f_n(x)) \leq p_k/(nh^2) = f(\xi_k)/(nh).$$

Si f(x) es continua Lipschitz de constante γ_k en B_k, entonces,

$$|sesgo(f_n(x))| = |E(f_n(x)) - f(x)| = |p_k/h - f(x)|$$

$$= |f(\xi_k) - f(x)| \leq \gamma_k|\xi_k - x| \leq \gamma_k h .$$

Luego,

$$sesgo^2(f_n(x)) \leq \gamma_k^2 h^2$$

Acotando superiormente el ECM concluímos,

$$ECM(f_n(x)) \leq f(\xi_k)/(nh) + \gamma_k^2 h^2. \tag{3}$$

La desigualdad anterior muestra el comportamiento inverso del sesgo y la varianza con respecto a la elección de h: aumentando h disminuimos la varianza y agrandamos el sesgo del estimador, y a la inversa; lo cual implica un compromiso entre sesgo y varianza. Esto puede ser controlado eligiendo una amplitud de intervalo que decrezca cuando aumenta el tamaño de muestra de tal manera que cuando $n \to +\infty$, entonces $h \to 0$ y $nh \to +\infty$.
Recordando la definición de estimador consistente en media cuadrática y considerando la desigualdad (3), se obtiene el siguiente resultado.

Proposición 1. *Sean $X_1,\ldots,X_n$, independientes, idénticamente distribuídas (i.i.d). Sea $x \in B_k$ y supongamos que f es continua Lipschitz en B_k con constante de Lipschitz γ_k, entonces el estimador dado por (1) es consistente en media cuadrática cuando $n \to +\infty$, si se verifica que $h \to 0$ y $nh \to +\infty$.*

En ese caso, el valor h que minimiza la acotación (3) es igual a $h^* = [f(\xi_k)/2n\gamma_k^2]^{1/3}$ y la cota es de orden $n^{-2/3}$.
El decrecimiento de h^* es de orden $n^{-1/3}$, mucho mayor que el correspondiente a la regla de Sturges que es de orden $[\log_2(n)]^{-1}$, por lo que daría un estimador más suave.
Este resultado de consistencia acotando superiormente el ECM no es útil en la práctica, ya que, algunas veces el valor superior puede estar muy lejos del verdadero.
Aproximaciones más precisas pueden ser obtenidas suponiendo la existencia de derivadas primeras de f.
Vamos a analizar ahora el error global MECI, a partir de la norma L_2, para encontrar el mejor estimador de f utilizando un histograma.
Puesto que la expresión de MECI depende del parámetro de suavizado h, esto se reduciría a encontrar el h óptimo que lo minimice.
Si recordamos la expresión de MECI (pag.13-14), en su cálculo interviene la densidad f desconocida,

$$\text{MECI} = \text{IECM} = \int \text{ECM}[f_n(x)]dx = \int E([f_n(x)-f(x)]^2)dx$$

$$= \int [\text{var}(f_n(x)) + \text{sesgo}^2(f_n(x))]\, dx.$$

Si desconocemos f, para soslayar este problema, primero encontraremos una expresión asintótica del error global expuesto y, por último, una estimación de los valores desconocidos que figuran en esa expresión, utilizando cálculos de validación cruzada (VC) que consisten en utilizar las observaciones en forma secuencial, usando una parte de los datos para averiguar sobre los datos restantes (Wand y Jones, (1995), pag.63).
Recordando que, para $x \in B_k$,

$$\mathrm{Var}(f_n(x)) = p_k(1-p_k)/(nh^2) = \left(\int_{B_k} f(t)dt\right)\left(1-\int_{B_k} f(t)dt\right)/(nh^2).$$

Al desarrollar por Taylor la función f(t) alrededor de x,

$$f(t) = f(x) + (t-x)f'(x) + (t-x)^2 f''(x)/2 + \ldots,$$

e integrando f(t) sobre B_k, se llega a que,

$$\mathrm{Var}(f_n(x)) = \left(\int_{B_k} f(t)dt\right)\left(1 - \int_{B_k} f(t)dt\right)/(nh^2)$$

$$= (nh)^{-1}\left(h^{-1}\int_{B_k} f(t)dt\right)\left(1 - \int_{B_k} f(t)dt\right)$$

$$= (nh)^{-1}\left(h^{-1}\int_{B_k} f(t)dt\right)\left(1 - O(h)\right) = (nh)^{-1}\left(f(x) + o(1)\right)$$

$$= (nh)^{-1}f(x) + o((nh)^{-1}) \text{ cuando } nh \to +\infty \text{ y } x \in B_k.$$

Usando el mismo desarrollo de Taylor, se llega a que, para $x \in B_k$ y $h \to 0$, el

$$\mathrm{sesgo}(f_n(x)) = E(f_n(x)) - f(x) = p_k/h - f(x) = h^{-1}\int_{B_k} f(t)dt - f(x)$$

$$= h^{-1}\int_{B_k}(f(x) + (t-x)f'(x) + (t-x)^2 f''(x)/2 + \ldots)dt - f(x)$$

$$= (t_0 + (k + 0.5)h - x)f'(x) + O(h^2).$$

Luego,

$$ECM(f_n(x)) = (nh)^{-1}f(x) + o((nh)^{-1})$$

$$+(t_0 + (k + 0.5)h - x)^2 f'^2(x) + O(h^2) \text{ para } x \in B_k.$$

Por otra parte, recordando que $\int f(x)dx = 1$, podemos ver que,

$$\sum_k \int_{B_k} (nh)^{-1} f(x)dx = (nh)^{-1}.$$

Observemos además que, la integral sobre B_k, del tercer término de la expresión del $ECM(f_n(x))$ puede expresarse como,

$$\int_{B_k} (t_0 + (k + 0.5)h - x)^2 f'^2(x) = f'^2(\eta_k)\int_{B_k} (t_0 + (k + 0.5)h - x)^2 dx$$

$$= f'^2(\eta_k)h^3/12, \text{ para algún } \eta_k \in B_k.$$

Si sumamos las integrales sobre los B_k y usamos la aproximación de la integral (Scott (1992), pag.54),

$$\sum_k [\int_{B_k} (t_0 + (k + 0.5)h - x)^2 f'^2(x)] = (h^2/12)\sum_k f'^2(\eta_k)h$$

$$= (h^2/12) \int f'^2(x)dx + o(h^2) = (h^2/12)R(f') + o(h^2).$$

Finalmente, integrando el $ECM(f_n(x))$ sobre cada B_k y sumando, obtenemos

$$IECM(f_n) = 1/(nh) + (h^2/12)R(f') + o(h^2) + o(1/nh) = MECI(f_n),$$

donde $R(f') = \int f'^2(t)dt$.

Asintóticamente, el termino $o(h^2) + o(1/nh)$ puede ser ignorado cuando $nh \to +\infty$ y $h \to 0$.
Esto lleva a una expresión asintótica para el MECI que indicaremos por, $AMECI(h) = 1/(nh) + (h^2/12)R(f')$, el cual alcanza su mínimo cuando $h = [6/(R(f')n)]^{1/3}$.
El siguiente resultado sintetiza los pasos anteriores.

Proposición 2. *Suponiendo que f tiene derivada absolutamente continua y de cuadrado integrable, se obtiene que una expresión*

asintótica para el MECI, cuando $n \to +\infty$, $h \to 0$ *y* $nh \to +\infty$, *está dada por* $AMECI(f_n) = 1/(nh)+(h^2/12)R(f')$, *el cual alcanza su mínimo en* $h = h^*$ *donde*

$$h^* = [6/(R(f')n)]^{1/3} , \qquad (4)$$

y el AMECI mínimo está dado por

$$AMECI^* = [3/(4n)]^{2/3} \; [R(f')]^{1/3} .$$

El AMECI óptimo decrece en la misma proporción $n^{-2/3}$ que la cota mínima del ECM. El h^* depende de f sólo a través de la medida de irregularidad $R(f')$ de la primera derivada.
Es ilustrativo calcular h^* y $AMECI^*$ para el caso en que f es la densidad normal. Siendo $R(f') = 1/(4\pi^{1/2}\sigma^3)$, entonces $h^* = (24\pi^{1/2}\sigma^3/n)^{1/3}$ y $AMECI^* = (3/4n)^{2/3}(4\pi^{1/2}\sigma^3)^{-1/3} = 0.4297n^{-2/3}\sigma^{-1}$.
Es oportuno comparar el valor del MECI asintótico cuando se usa la estimación paramétrica y $f(x,\theta) = \varphi((x-\mu)/\sigma)$ donde φ es la densidad de una normal estandar y $\theta = (\mu,\sigma^2)$.
En ese caso $f_n(x) = f(x,\theta_n) = \varphi((x-\bar{x})/s)$ y el error estará dado por,

$$MECI = E[\int (\varphi((x-\bar{x})/s) - \varphi((x-\mu)/\sigma))^2 \, dx \,] .$$

Scott ((1992), pag.41) probó, que cuando $(\mu,\sigma^2) = (0,1)$, el error global asintótico tiene un valor igual a $7/(16n\pi^{1/2})$.
Los resultados encontrados indican, por ejemplo, que para un valor de $n = 100$, $AMECI = 0.002468$; mientras que, para el histograma se necesita una muestra de $n = 2297$ para lograr el mismo valor de AMECI.
Esto habla en forma contundente a favor de los métodos paramétricos cuando se conoce la familia a la cual pertenece la densidad que queremos estimar. Sin embargo, cuando la densidad se aparta levemente de la normal, el MECI puede no tender a cero (Scott, (1992), pag.58).
Presentaremos ahora dos maneras de aproximar la abertura de ventana óptima asintótica h^* de la Proposición 2, usando algoritmos de validación cruzada.

1.2.Validación cruzada sesgada para estimar la ventana óptima asintótica.

Observemos, que el h óptimo asintótico dado por la expresión (4), tiene como única cantidad desconocida el valor $R(f')$, el cual puede ser estimado usando los datos obtenidos de la muestra.
Una aproximación de $f'(t_k)$ (Scott, (1992), pag.75), puede basarse en una diferencia del valor del histograma, en los puntos medios de los intervalos B_k y B_{k+1} o sea, $f_n'(t_k) = [n_{k+1}/(nh)-n_k/(nh)]/h$. Un estimador posible de $R(f')$ sería,

$$R_1 = \sum_k [f_n'(t_k)]^2 h = \sum_k (n_{k+1}-n_k)^2/n^2h^3 .$$

Este método de selección del parámetro de suavizado se conoce como método plug-in y fue introducido por Woodroofe (1970).
Es fácil comprobar, usando desarrollo de Taylor y la aproximación a la integral de Riemann, que

$$E(R_1) = R(f') + 2/(nh^3) + O(h).$$

Con el valor de h^*, dado por la Proposición 2, se puede comprobar que $2/(nh^3)$ converge a $R(f')/3$. Luego, R_1 es un estimador sesgado de $R(f')$, siendo $(3/4)R_1$ asintóticamente insesgado, para $R(f')$.
Si consideramos $R_2 = R_1 - 2/(nh^3)$, como estimador de $R(f')$ en la expresión de AMECI de la Proposición 2, obtendremos un estimador de validación cruzada sesgada de MECI (Scott, (1992), pag.76), el cual será indicado por

$$VCS(h) = 5/(6nh) + \sum (n_{k+1}-n_k)^2/12n^2h.$$

El h que minimiza VCS(h), lo notamos h_{VCS} y converge a h^* con un error relativo de orden $n^{-1/6}$ (Scott y Terrell (1987)); la convergencia es lenta pero sin embargo es un algoritmo útil, especialmente para muestras grandes.

1.3.Validación cruzada insesgada para estimar la ventana óptima asintótica.

Un método para encontrar una aproximación a la abertura de ventana que minimiza el MECI, en base a las observaciones obtenidas, es el

llamado de validación cruzada insesgada (VCI). Este fue presentado por Rudemo (1982) y tiene por objetivo minimizar el ECI cuando tenemos una muestra dada, eligiendo la amplitud de intervalo o abertura de ventana h adecuada.
Recordemos que,

$$ECI(h) = \int [f_n(x)-f(x)]^2dx = R(f_n) - 2\int f_n(x)f(x)dx + R(f).$$

Minimizar ECI(h) es equivalente a minimizar $R(f_n) - 2\int f_n(x)f(x)dx$. La segunda integral depende de la función de densidad desconocida f que puede pensarse como la $E_X(f_n(X))$, donde la esperanza es con respecto a la variable aleatoria X, con densidad f e independiente de la muestra.
Esta esperanza puede estimarse con el algoritmo de validación cruzada; para cada observación X_i, se calcula el histograma en ese punto, basado en las restantes n-1 observaciones, que notamos $f_{n,-i}(X_i)$. Estos n valores de histograma se promedian para estimar la $E_X(f_n(X))$.
O sea, un estimador de $E_X(f_n(X))$ es

$$\sum_{i=1}^{n} f_{n,-i}(X_i)/n.$$

En efecto, considerando que $E_X(f_n(X))$ es una variable aleatoria función de la muestra $X_1,\ldots,X_n$, es fácil comprobar que $E(E_X(f_n(X)) = E(f_{n,-i}(X_i))$, $i=1,\ldots,n$, donde E es la esperanza respecto a $(X_1,\ldots,X_n)$, de donde se obtiene que un predictor de $E_X(f_n(X))$ es un promedio de $f_{n,-i}(X_i)$, $i=1,\ldots,n$.
Luego,

$$MECI(h)=E[\int [f_n(x)-f(x)]^2dx]$$

$$= E[R(f_n)] - 2E[\int f_n(x)f(x)dx] + E[R(f)]$$

$$= E[R(f_n)] - 2E[E_X(f_n(X)] + R(f)$$

$$= E[R(f_n)] - 2\sum_{i=1}^{n} E[f_{n,-i}(X_i)]/n + R(f)$$

$$= E[R(f_n) - 2\sum_{i=1}^{n} f_{n,-i}(X_i)/n\ ] + R(f).$$

Por lo tanto la expresión,

$$R(f_n) - 2 \sum_{i=1}^{n} f_{n,-i}(X_i)/n + R(f),$$

es un estimador insesgado de MECI(h).
Luego, la amplitud h que minimiza a la expresión,

$$VCI(h) = R(f_n) - 2 \sum_{i=1}^{n} f_{n,-i}(X_i)/n$$

$$= 2/(n-1)h - (n+1)/(n^2(n-1)h) \sum_{k=1}^{n} n_k^2,$$

(Scott (1992), pag.77) será un valor cercano al que minimiza el MECI(h) y no necesitamos conocer la densidad f para encontrarlo. Ese valor lo notaremos h_{VCI}.
Scott ((1992), pag.77) compara, para distintas circunstancias, los dos métodos presentados para encontrar la estimación de la abertura de ventana óptima asintótica, aconsejando un gráfico comparativo, con muestras simuladas, de VCI(h) y VCS(h) versus h.
Con lo dicho tenemos una idea somera del comportamiento del histograma, respecto a la medida de error dada por la media del error cuadrático integrado (MECI) o su valor asintótico. Existen numerosos trabajos que tratan más profundamente el tema y otros relacionados.
En el caso del histograma hemos hecho un desarrollo de algunas metodologías para obtener una apertura de ventana óptima y como encontrarla, para los demás estimadores de la densidad que describiremos sólo mencionaremos las técnicas de uso estandar.

2-Polígono de frecuencia.

Las discontinuidades del histograma, en los bordes de los intervalos, ha limitado su uso. El polígono de frecuencias viene a subsanar este problema, ya que es un estimador continuo de la densidad, basado en el histograma, y se obtiene interpolando linealmente su valor en los puntos medios de los intervalos igualmente espaciados. El polígono conecta el valor del histograma en dos intervalos adyacentes. Scott (1985a) presenta varios argumentos a favor del polígono.

Consideremos dos intervalos vecinos de amplitud h,

$$I_k=[t_0+(k-1)h, t_0+kh) \quad y \quad I_{k+1}=[t_0+kh, t_0+(k+1)h), \quad k \in Z .$$

El estimador por el polígono de frecuencia está definido, para $x \in J_k = [t_0 + (k - 0.5)h, t_0 + (k + 0.5)h)$, como el segmento que une los puntos de coordenadas $(t_0 + (k - 0.5)h; n_k/nh)$ y $(t_0 + (k + 0.5)h; n_{k+1}/nh)$, donde n_i es la frecuencia de observaciones en el intervalo I_i.
Para simplificar la notación consideraremos $t_0 = 0$ y $k = 0$. Los intervalos serán entonces $I_0 = [-h,0)$ y $I_1 = [0,h)$. Explícitamente, el estimador de la densidad será,

$$f_n(x)= (1/2 - x/h)(n_0/nh)+(1/2 + x/h)(n_1/nh), \quad \text{para } -h/2 \le x < h/2. \tag{5}$$

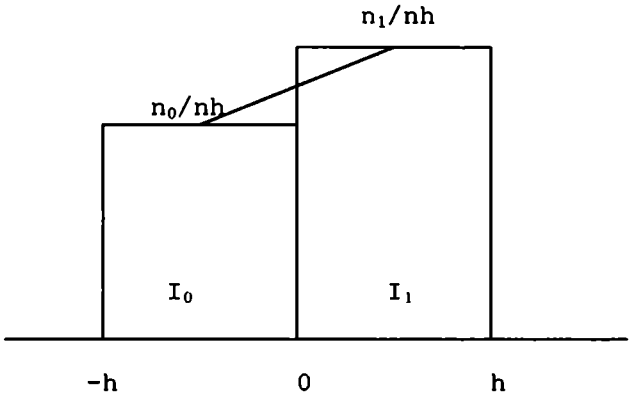

Fig. 2. Gráfico de polígono

A fin de encontrar la expresión de MECI, consideremos las siguientes aproximaciones, para $p_i = P(X \in I_i)$, $i = 0,1$;

$$p_0 = \int_{[-h,0)} f(t)dt \cong h\, f(0) - h^2\, f'(0)/2 + h^3\, f''(0)/6, \tag{6}$$

$$p_1 = \int_{[0,h)} f(t)dt \cong h\, f(0) + h^2\, f'(0)/2 + h^3\, f''(0)/6,$$

obtenidas usando el desarrollo de Taylor alrededor del punto cero, de la función f,

$$f(x) = f(0) + x\, f'(0) + x^2\, f''(0)/2 + x^3\, f'''(0)/6 + \ldots$$

Luego, como $n_k = Bi(n,p_k)$, por (5) y (6) obtenemos,

$$E[f_n(x)] = (1/2 - x/h)(p_0/h)+(1/2 + x/h)(p_1/h)$$

$$\cong f(0) + xf'(0) + h^2f''(0)/6, \qquad \text{para } x \in J_0 = [-h/2,h/2).$$

Por lo tanto, como los primeros sumandos son los primeros términos del desarrollo de Taylor de f(x) alrededor del punto 0,

$$f(x) = f(0) + xf'(0) + x^2f''(0)/2 + \ldots,$$

una aproximación para el sesgo está dada por

$$\text{Sesgo}(f_n(x)) = E[f_n(x)]-f(x) \cong (h^2-3x^2)f''(0)/6.$$

Con lo cual, se llega a

$$\int_{[-h/2,h/2)} \text{sesgo}^2(f_n(x))dx \cong \int_{[-h/2,h/2)} [(h^2-3x^2)f''(0)/6]^2\, dx$$

$$= [49\ h^4\ f''(0)^2/2880]h.$$

Para otros intervalos $J_k = [(k-0.5)h,(k+0.5)h)$, el resultado estará dado por, $[49\ h^4\ f''(kh)^2/2880]h$, ya que el desarrollo de Taylor se hará alrededor del punto kh.
Sumando sobre todos los intervalos y usando el resultado del Lema 3.1 de Scott (1985a), se obtiene

$$\int \text{sesgo}^2(f_n(x))dx = \sum_k \int_{I_k} \text{sesgo}^2(f_n(x))dx \cong \sum_k [49\ h^4\ f''(kh)^2/2880]h$$

$$=(49/2880)h^4 \sum_k f''(kh)^2 h = (49/2880)h^4(R(f'')+O(h\|2f''f'''\|_1))$$

$$= 49h^4R(f'')/2880 + O(h^5\|f''\|_2\|f'''\|_2), \qquad \text{donde } \|g\|_p = [\int |g(x)|^p]^{1/p}.$$

Es decir,

$$\int \text{sesgo}^2(f_n(x))dx \cong 49h^4R(f'')/2880 + O(h^5\|f''\|_2\|f'''\|_2)). \qquad (7)$$

Con respecto a la varianza, para $-h/2 \le x < h/2$,

$$\mathrm{Var}(f_n(x)) = (1/2 - x/h)^2 \mathrm{var}[(n_0/nh)] + (1/2 + x/h)^2 \mathrm{var}[(n_1/nh)]$$

$$+ 2[(1/2)^2 - (x/h)^2]\mathrm{cov}[(n_0/nh),(n_1/nh)]. \qquad (8)$$

Puesto que,

$$\mathrm{var}[(n_i/nh)] = np_i(1-p_i)/(nh)^2 \cong [f(0)(1 - hf(0)]/(nh), \quad i = 0, 1,$$

$$\mathrm{cov}[(n_0/nh),(n_1/nh)] = -np_0p_1/(nh)^2 \cong -f(0)^2/n;$$

sustituyendo estas aproximaciones en (8), se obtiene

$$\mathrm{var}(f_n(x)) \cong [2x^2/(nh^3) + 1/(2nh)]f(0) - f(0)^2/n .$$

Integrando sobre $(-h/2,h/2)$,

$$\int_{[-h/2,h/2)} \mathrm{var}(f_n(x))\, dx \cong [2f(0)/(3nh) - f(0)^2/n]h .$$

Similarmente, sobre $J_k = [(k-0.5)h,(k+0.5)h)$, se obtiene la aproximación,

$$\int_{J_k} \mathrm{var}(f_n(x))\, dx \cong [2f(kh)/(3nh) - f(kh)^2/n]h .$$

Aplicando el Lema 3.1 de Scott (1985a), y recordando que $\int f(x)dx = 1$, tenemos

$$\int \mathrm{var}(f_n(x))\, dx = \sum_k \int_{J_k} \mathrm{var}(f_n(x))\, dx \cong \sum_k [2f(kh)/(3nh) - f(kh)^2/n]h$$

$$= 2/(3nh) - R(f)/n + O(\|f\|_2\|f'\|_2 h/n)). \qquad (9)$$

Apelando a las aproximaciones (7), (9) y a la expresión de MECI se llega al siguiente resultado.

Proposición 3. *Supongamos que $f \in \mathcal{C}^3$, f'' es absolutamente continua y que se cumple $R(f^{(k)}) < +\infty$, $k=0,1,2,3$. Entonces se obtiene que una expresión asintótica del MECI, cuando $n \to +\infty$, $h \to 0$ y $nh \to +\infty$, es*

$$AMECI(h) = 2/(3nh) + h^4 R(f'')49/2880,$$

con lo cual, la amplitud h que minimiza la expresión de AMECI está dado por,

$$h^{*} = 2[15/(49n\ R(f''))]^{1/5}, \qquad (10)$$

siendo,

$$AMECI(h^{*})=AMECI^{*}=(5/12)[49n^{-4}R(f'')/15]^{1/5}. \qquad (11)$$

A fin de comparar el histograma con el polígono de frecuencias, al tomar una muestra de 800 observaciones de una población normal, la amplitud óptima h^{*} correspondiente al polígono resulta ser el 50% mayor que el correspondiente al histograma. Aparentemente, a causa de las discontinuidades del histograma, éste debe ser más irregular para aproximarse a los cambios de la densidad. La continuidad del polígono permite aproximar mejor la densidad continua por ajustes lineales sobre intervalos más amplios. Por otra parte, se ha encontrado (Scott, (1992), pag.99) que un valor de AMECI igual a 0.002468, en una densidad normal estandar, que requería una muestra de 2297 observaciones en el histograma, se logra con sólo 546 en el polígono, aunque siempre en desventaja con el ajuste paramétrico frente a la especificación correcta del modelo.

Para la densidad normal, $R(f'')=3/(8\pi^{1/2}\ \sigma^{5})$, entonces, por (10) y (11), $h^{*} = 2.15\sigma\ n^{-1/5}$ y $AMECI^{*} = 0.387\ \sigma^{-1}\ n^{-4/5}$, siendo este último valor bastante menor que el calculado para el histograma.

Una situación en la cual el polígono está en desventaja es cuando la densidad presenta discontinuidades, respondiendo mejor el histograma si se sitúan los bordes de los intervalos en los puntos de discontinuidad.

Como antes, la amplitud óptima h^{*} depende de la densidad desconocida. Como en el caso del histograma, aproximaciones a h^{*} pueden obtenerse mediante validación cruzada. El siguiente estimador de $R(f'')$ fue propuesto por Scott y Terrell (1987),

$$R_1(f'') = (1/n^2h^5)\sum_k (n_{k+1} - 2n_k + n_{k-1})^2 - 6/(nh^5),$$

y reemplazando en (11), obtenemos la expresión asintótica a minimizar,

$$VCS(h) = 271/(480nh) + 49/(2880n^2h)\sum_k (n_{k+1} - 2n_k + n_{k-1})^2.$$

Hemos visto que tanto el histograma como el polígono, tienen totalmente determinados sus intervalos, especificando el origen t_0 y el parámetro de suavizado h. El valor t_0 no cumple rol alguno en los resultados asintóticos vistos, sin embargo un gráfico de histograma puede cambiar sustancialmente al variar solamente el origen, por ejemplo de unimodal a multimodal, como se aprecia en los histogramas (Härdle, (1991), pag.28) de la Fig. 3.

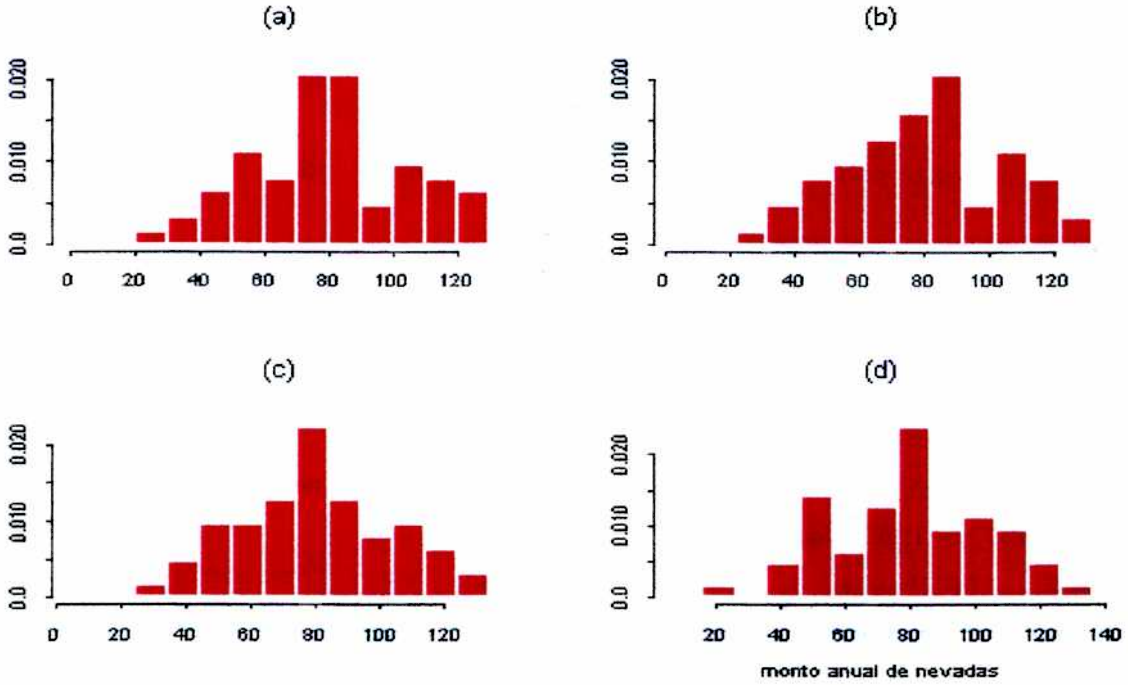

Fig. 3. Histograma para el monto anual de nieve caída en Buffalo, durante los años 1910-1972 (Scott, (1992), pag.279). (a) origen t_0=0, (b) t_0=2, (c) t_0=4, (d) t_0=6. La amplitud de intervalo es la misma en los cuatro histogramas.

Para el polígono existen más elecciones de t_0, ya que el ancho óptimo de los intervalos es mayor. Se aconseja elegir un t_0 que permita el gráfico más suave. Como esto es muy subjetivo, el próximo estimador que veremos elimina el efecto de este parámetro.

3. Histograma corrido promediado.

Este histograma elimina el problema de la ubicación de los intervalos, conservando las ventajas computacionales de los estimadores basados en las frecuencias relativas. Scott (1985b)

propone promediar polígonos con distintos orígenes pero con el mismo parámetro de suavizado. Un algoritmo más simple de describir, analizar y comprender, es el que promedia histogramas corridos y tendrá también la apariencia de un histograma, o sea con valor constante sobre intervalos disjuntos.

Veamos cómo se define el histograma corrido promediado (HCP).

Consideremos un conjunto de m histogramas $H_1,...H_m$ con el mismo parámetro de suavizado h y con orígenes t_0 dados por los valores 0, h/m,....,(m-1)h/m respectivamente, o sea H_j tiene el comienzo en (j-1)h/m para j=1,2,...m.

Notemos a los m histogramas H_j, estimadores de la densidad f, calculados a partir de n observaciones, como $H_j = f_{j,n}$ para j=1,2,...,m.

Definimos entonces f_n , el estimador del HCP, como

$$f_n(x) = \sum_{j=1}^{m} f_{j,n}(x)/m.$$

En un sentido, el parámetro t_0 ha sido reemplazado por el parámetro m que debe ser especificado.

Los intervalos donde f_n es constante tienen ahora la forma $B_k^* = [k\delta,(k+1)\delta)$ donde $\delta = h/m$ y k=0,1,2,...

Estos intervalos B_k^* tienen amplitud h/m, frecuencia n_k^* y son generados dividiendo los intervalos de cada histograma $f_{j,n}$ en m subintervalos, estando contenidos en, por lo menos, uno de ellos.

Notemos con $B_{l,j}$ el l-ésimo intervalo del histograma $f_{j,n}$, entonces $B_{l,j} = [(j-1)h/m + (l-1)h,(j-1)h/m + lh)$.

El subintervalo B_k^* estará contenido en ese intervalo si $l-1 \leq (k-j+1)/m < l$ con j=1,2,..m, l=1,2,...

Para encontrar el valor del HCP en B_k^*, debemos promediar las frecuencias relativas $n_{l,j}/nh$ de los intervalos $B_{l,j}$ que contengan a B_k^*.

Por otra parte, la frecuencia de un $B_{l,j}$ puede encontrarse sumando las frecuencias de los subintervalos B_k^* contenidos en él y que serán adyacentes.

Luego, para definir el HCP en un intervalo B_k^*, tengamos en cuenta que es el promedio del valor de los histogramas corridos, cada uno calculado en el intervalo que contiene a B_k^*, que a su vez puede determinarse como la suma de la frecuencia de los B_k^*, que están contenidos en él.

Luego, el valor en los histogramas corridos en función de los n_j^* será,

$$(n_{k-(m-1)}^*+\ldots+n_{k-2}^*+n_{k-1}^*+n_k^*)/nh,$$

$$(n_{k-(m-2)}^*+\ldots+n_{k-1}^*+n_k^*+n_{k+1}^*)/nh,\ldots,$$

$$(n_k^*+n_{k+1}^*+\ldots+n_{k+m-1}^*)/nh.$$

Promediando esos m valores, obtenemos finalmente,

$$f_n(x) = [\sum_{i=1-m}^{m-1} (m-|i|)n_{k+i}^*/nh]/m = [\sum_{i=1-m}^{m-1} (1-|i|/m)n_{k+i}^*]/nh$$

$$= [\sum_{i=1-m}^{m-1} w_m(i)n_{k+i}^*]/nh \qquad \text{para } x \in B_k^*.$$

Las frecuencias de los intervalos B_{k+i}^* más cercanos a B_k^* están pesadas por la cantidad $w_m(i)=1-|i|/m$, $|i|\leq m-1$, que se hace más pequeña al distanciarse del intervalo central B_k^*.
La representación gráfica de $w(i)$ corresponde a un triángulo isósceles sobre el intervalo $(-1,1)$; para $i = 0$ alcanza el valor mayor que corresponde a la frecuencia n_k^*.
A consecuencia del contorno que asumen los pesos, se ha generalizado el HCP aplicando otras curvas como multiplicadores de las frecuencias, teniendo en cuenta que la integral del estimador sea igual a uno. Para ello debe cumplirse que los pesos sumen m.
Una manera simple de definir pesos está dada por,

$$w_m(i) = mK(i/m)/\sum_{j=1-m}^{m-1} K(j/m) \qquad 1-m \leq i \leq m-1,$$

donde $K(.)$ es una función continua definida sobre $(-1,1)$. A menudo K es una función de densidad.
Con respecto al valor del MECI asintótico y su relación con h, vamos a enunciar el siguiente resultado debido a Scott (1985b).

Proposición 4 *Suponiendo que la densidad f verifica que f″ es absolutamente continua y que* $R(f''') < +\infty$, *entonces el MECI del estimador* f_n *del HCP, con parámetro de suavizado h y parámetro de corrimiento m está dado por,*

$$MECI = (2/3nh)(1+1/2m^2) - R(f)/n + h^2R(f')/12m^2$$

$$+ h^4 R(f'')(1-2/m^2+3/5m^4)/144 + O(h/n + h^5).$$

Ignorando términos de orden n^{-1} *y haciendo tender m a infinito, el h que minimiza el* $AMECI = (2/3nh)(1+1/2m^2) + h^2R(f')/12m^2 + h^4 R(f'')(1-2/m^2+3/5m^4)/144$, *está dado por*

$$h^* = [24/(nR(f''))]^{1/5}$$

Observemos que cuando m = 1 los dos primeros términos del AMECI coinciden con el AMECI calculado para el histograma (Proposición 2), y cuando $m \to +\infty$ es similar al del polígono (Proposición 3).

Ahora podemos comparar la eficiencia de los tres estimadores no paramétricos y un estimador paramétrico, cuando f es la densidad normal. La siguiente tabla da los tamaños de muestra necesarios para que el AMECI sea aproximadamente 0.0025.

Estimador	$N(\bar{x}, s^2)$	HCP	Poligono	Histograma
Tamaño de Muestra	100	436	546	2297

AMECI = 0.0025

La conveniencia del HCP respecto al polígono no es tan acentuada como respecto al histograma.

En algunas situaciones adversas, tal como muestras pequeñas de una población con gran irregularidad en la densidad, el histograma puede competir con el HCP. Bajo condiciones más restrictivas, asintóticamente la eficiencia del histograma será cero respecto del HCP o polígono de frecuencia debido a las distintas

velocidades de convergencia del MECI (ver Proposición 2 y Proposición 3).

3.1 El HCP límite como estimador por núcleo.

La determinación del parámetro m es menos importante que la del parámetro t_0, ya que, eligiendo adecuadamente el h y cuidando que $m \geq 2$, vamos a conseguir una estimación conveniente.
El análisis del AMECI cuando $m \to +\infty$ es importante porque el límite del estimador da como resultado, lo cual explicaremos inmediatamente, el estimador por núcleo.
Con h y n fijo, y m creciendo, es fácil aislar el efecto de una observación X_j sobre $f_n(x)$ para algún x.
Si $x \in B_k^*$ y $X_j \in B_{k+i}^*$ con $|k+i|<m$, la influencia de X_j sobre x es proporcional al peso asignado a n_{k+i}^*,

$$1-|i|/m = 1-|i|\delta/\delta m = 1-|i|\delta/h = 1-|x-X_j|/h + |c|\delta/h$$

$$= 1-|x-X_j|/h + O(\delta/h). \qquad |c|<1$$

Si $|k+i| \geq m$ ó, equivalentemente, $X_j \notin (x-h,x+h)$, entonces la influencia de X_j es nula.

Finalmente,

$$\lim_{m \to +\infty} f_n(x) = (1/nh)\sum_{j=1}^{n}[1-|x-X_j|/h]\ I_{[-1,1]}((x-X_j)/h). \qquad (12)$$

Si llamamos $K(t) = (1-|t|)\ I_{[-1,1]}(t)$, entonces la igualdad (12) quedará expresada como,

$$\lim_{m \to +\infty} f_n(x) = (1/nh)\sum_{j=1}^{n} K((x-X_j)/h). \qquad (13)$$

El segundo miembro de (13) también representa el estimador de núcleo de la densidad f, donde K se denomina función núcleo siendo en este caso particular, el núcleo triangular con soporte en (-1,1).

Veremos que otros núcleos pueden ser aplicados y algunos con soporte infinito, como en el caso del núcleo normal.

4-Estimador basado en núcleos.

Fix y Hodges (1951), introdujeron un algoritmo básico para obtener un estimador no paramétrico de la densidad, focalizados a resolver el problema de discriminación estadística, cuando se desconoce la densidad. En las décadas siguientes, varios algoritmos y formas teóricas de análisis alternativos fueron introducidas por Rosenblatt (1956), Parzen (1962) y Cencov (1962). Posteriormente, importantes desarrollos teóricos fueron llevados a cabo por Watson y Leadbetter (1963), Loftsgaarden y Quesenberry (1965), Schwartz (1967), Epanechnikov (1969), Tarter y Kronmal (1970) y Wahba (1971). Finalmente en los años 70 aparecen los primeros trabajos sobre la aplicación de estos métodos a partir de los trabajos de Scott y colaboradores (1978) y Silverman (1978).
El estimador por núcleo f_n se define como,

$$f_n(x) = (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)h^{-1}) = n^{-1} \sum_{i=1}^{n} K_h(x-X_i), \qquad (14)$$

donde $K:R\to R$ se denomina función núcleo y verifica $\int K(w)dw=1$ $K\geq 0$.
$K_h(t) = K(t/h)/h$, siendo h el parámetro de suavizado.
Aunque la motivación natural de este estimador pareciera ser el HCP, siguiendo otros orígenes, como el análisis numérico, las series de tiempo y los procesos de señales, podemos alcanzar un entendimiento profundo de los métodos por núcleos.
Rosenblatt (1956), lo presentó como una aproximación numérica a la derivada de la función de distribución acumulada. En efecto, como

$$f(x) = \lim_{h\to 0} (F(x+h) - F(x-h))/2h,$$

podemos estimar f(x) por

$$f_n(x) = (F_n(x+h)-F_n(x-h))/(2h) = (nh)^{-1} \sum_{i=1}^{n} (1/2) I_{(-1,1)}((x-X_i)h^{-1})$$

$$= (nh)^{-1} \sum_{i=1}^{n} K((x-X_i)h^{-1}),$$

donde el núcleo K es la densidad uniforme en (-1,1) y F_n es la acumulada empírica.
El análisis estadístico de estimadores por núcleos es más simple que para los histogramas, ya que es un promedio de n variables aleatorias independientes e idénticamente distribuídas $K_h(x-X_i)$, para i = 1, 2,..., n.
Como función de x el estimador $f_n(x)$ hereda las propiedades de continuidad y diferenciabilidad de K.
Para llegar a la expresión de MECI comencemos a calcular la varianza y la esperanza del estimador (14).

$$E(f_n(x)) = E(n^{-1} \sum K_h(x-X_i)) = E(K_h(x-X))$$

$$= (1/h)\int K((x-t)/h)f(t)dt = \int K(w)f(x-hw)dw. \qquad (15)$$

Desarrollando por Taylor f(x-hw) en la última expresión, obtenemos

$$E(f_n(x)) = f(x)\int K(w)dw - hf'(x)\int wK(w)dw + (1/2)\ h^2\ f''(x)\int w^2K(w)dw$$

$$- (1/3!)\ h^3\int f^{(3)}(\xi)w^3\ K(w)dw, \qquad (16)$$

donde $\xi = \xi(w)$ es un punto entre x y x-hw.

Por otra parte,

$$Var(f_n(x)) = Var(n^{-1} \sum K_h(x-X_i)) = Var(K_h(x-X))/n, \qquad (17)$$

donde,

$$Var(K_h(x-X)) = E(K_h(x-X)^2) - [E(K_h(x-X))]^2$$

$$= (1/h^2)\int K((x-t)/h)^2f(t)dt - [f(x)\int K(w)dw$$

$$- hf'(x)\int wK(w)dw + (1/2)\ h^2\int f''(\xi)w^2K(w)dw]^2$$

$$= (1/h)\int K(w)^2\ f(x-hw)dw - [f(x)\int K(w)dw$$

$$- hf'(x)\int wK(w)dw + (1/2)\ h^2\int f''(\xi)w^2K(w)dw]^2$$

$$= f(x)R(K)/h - [f(x)\int K(w)dw - hf'(x)\int wK(w)dw$$

$$+ (1/2)\ h^2\int f''(\xi)w^2K(w)dw]^2. \qquad (18)$$

Si el núcleo K satisface

$$\int K(w)dw = 1, \qquad \int w\ K(w)dw = 0, \qquad \int w^2\ K(w)dw = \sigma_K^2 > 0, \quad (19)$$

las ecuaciones (16) y (19) implican,

$$\text{sesgo}(f_n(x)) = \sigma_K^2 h^2 f''(x)/2 + o(h^2), \qquad h\to 0. \qquad (20)$$

Luego,

$$\int \text{sesgo}^2(f_n(x))dx = \sigma_K^4 h^4 R(f'')/4 + o(h^4),\ h\to 0.$$

Por las ecuaciónes (17), (18) y (19) obtenemos,

$$\text{Var}(f_n(x)) = f(x)R(K)/(nh) - f^2(x)/n + O(h/n),$$

$$= f(x)R(K)/(nh) + o((nh)^{-1}), \quad nh\to +\infty. \qquad (21)$$

y

$$\int \text{Var}(f_n(x))dx = R(K)/(nh) + o((nh)^{-1}), \quad nh\to +\infty.$$

De las expresiones (20) de sesgo y (21) de varianza del estimador f_n podemos deducir que el $ECM(f_n)$ se puede expresar como,
$ECM(f_n) = \text{var}(f_n(x)) + \text{sesgo}^2(f_n(x))$

$$= f(x)R(K)/(nh) + h^4\ (f''(x)\sigma_K^2)^2/4 + o((nh)^{-1}) + o(h^4)$$

$$nh\to +\infty,\ h\to 0.$$

Entonces, la integral del $ECM(f_n)$ puede expresarse como

$$IECM(f_n) = MECI(f_n) = R(K)/(nh) + h^4\ R(f'')\sigma_K^4\ /4 + o((nh)^{-1}) + o(h^4)$$

$$nh\to +\infty,\ h\to 0,$$

llegando finalmente al siguiente resultado.

Proposición 5. *Para un estimador por núcleo, con núcleo $K \geq 0$ que cumple (19), se verifica que una expresión asintótica del MECI cuando $n \to +\infty$, $h \to 0$ y $nh \to +\infty$, está dada por*

$$AMECI(h) = R(K)/(nh) + \sigma_K^4 h^4 R(f'')/4.$$

Luego, el h óptimo que minimiza el AMECI(h) está dado por,

$$h_{opt} = [R(K)/(n\sigma_K^4 R(f''))]^{1/5},$$

siendo el valor óptimo

$$AMECI(h_{opt}) = AMECI^* = (5/4)[\sigma_K R(K)/n]^{4/5} R(f'')^{1/5}.$$

Las condiciones bajo las cuales se cumple la Proposición 5 han sido estudiadas por muchos autores. Un conjunto posible de hipótesis es que el núcleo sea una función de densidad, de cuadrado integrable, media cero y varianza finita; y que la densidad f cumpla que f'' sea absolutamente continua y f''' sea de cuadrado integrable.

Observemos que $AMECI^* = R(K)/(nh^*) + \sigma_K^4 h^{*4} R(f'')/4$, correspondiendo el primer sumando a la varianza integrada (VI) y el segundo al sesgo al cuadrado integrado (SCI), estando en una relación 4:1, por lo tanto el SCI representa el 20% del error global asintótico óptimo, mientras que la VI el 80%.

Comparemos la Proposición 5, con núcleo K triangular, con la Proposición 3 del estimador por polígono. Siendo $R(K) = 2/3$ y $\sigma_K^4 = 1/36$, los dos resultados son similares.

Ocurre lo mismo al comparar el resultado de la Proposición 4, con $m \to +\infty$ del HCP, cuando consideramos K triangular, en el estimador por núcleo.

Como referencia para valores posibles de la ventana óptima, según la Proposición 5, el valor de $h^* = (4/3)^{1/5}\sigma\, n^{-1/5}$, cuando $f(x) = \varphi(x/\sigma)$ y K es el núcleo normal.

En cuanto al núcleo K, vamos a presentar las expresiones de los más populares en la tabla siguiente; todos verifican (19).

Nucleo	K(u)
Uniforme	$I_{[-1,1]}(u)/2$
Triángular	$(1-\|u\|)\ I_{[-1,1]}(u)$
Epanechnikov	$3(1-u^2)\ I_{[-1,1]}(u)/4$
Cuartico	$15(1-u^2)^2 I_{[-1,1]}(u)/16$
Cúbico	$35(1-u^2)^3 I_{[-1,1]}(u)/32$
Gaussiano	$\exp(-u^2/2)/\sqrt{2\pi}$
Coseno	$\cos(u\pi/2)\, I_{[-1,1]}(u)\,(\pi/4)$

Aunque la calidad del estimador depende principalmente de h, y en menor grado del núcleo, es posible mejorar la elección del núcleo a través de un conocimiento más detallado de sus propiedades.
Si la función K verifica,

$$\int K(w)dw = 1 \qquad \int w^j\ K(w)dw = 0 \quad \text{para } j=1,..p-1, \qquad \int w^p\ K(w)dw \neq 0,$$

diremos que K es de orden p.
En ese caso, continuando el desarrollo (16) llegamos a

$$\text{sesgo}(f_n(x)) = \mu_p h^p f^{(p)}(x)/p! + \ldots, \quad \text{donde } \mu_p = \int w^p\ K(w)dw.$$

Quedando la expresión (21) sin cambio, llegamos al siguiente resultado.

Proposición 6. *Suponiendo que f es suficientemente diferenciable y que el núcleo K es de orden p entonces,*

$$AMECI(h) = R(K)/(nh) + \mu_p^2\ h^{2p}\ R(f^{(p)})/(p!)^2,$$

siendo $h^* = [(p!)^2 R(K)/(2p\ \mu_p^2\ R(f^{(p)}))]^{1/(2p+1)}\ n^{-1/(2p+1)}$,

y $AMECI^* = [\mu_p^2 R(K)^{2p} R(f^{(p)})]^{1/(2p+1)} n^{-2p/(1+2p)}$.

Este resultado se acerca al orden n^{-1} del AMECI paramétrico, sin embargo el valor h^* aumenta con p, anulando en parte ese beneficio (Scott (1992), pag.133).
Una recomendación conservadora sería elegir una función núcleo que sea suave, unimodal y simétrica alrededor del origen.
Otra manera de evaluar los posibles núcleos es midiendo la eficiencia (Scott (1992),pag.139) con respecto al núcleo de Epanechnikov $K(t)= (3/4)(1-t^2)I_{[-1,1]}(t)$ que minimiza la expresión $[\sigma_K R(K)]^{4/5}$, la cual está contenida en la expresión de $AMECI^*$ de la Proposición 5. La eficiencia relativa de los núcleos listados anteriormente respecto al de Epanechnikov, presentan valores comprendidos entre 1.0061 y 1.0758, lo cual daría una diferencia poco significativa.
Otros criterios de evaluación y perfomance de funciones núcleo se mencionan en Silverman ((1992),pag. 68) y Scott ((1992), pag.138).
Veamos una ejemplificación gráfica que nos permitirá observar la diferencia del estimador $f_n(x)$ al considerar dos valores distintos para la abertura de ventana h. Se representó en ambos gráficos las funciones núcleo $K_h(x-X_i)$, i=1,..n, que promediadas dan como resultado $f_n(x)$.
Elegimos el núcleo K normal y fijamos cinco valores observados: $X_1 = -1$, $X_2 = -0.4$, $X_3 = 0$, $X_4 = 0.4$ y $X_5 = 1$. En la Fig.4(a) se eligió h = 0.2 y en la Fig.4(b), h = 1.
En ambos gráficos, centrada en cada X_i está dibujada la curva $(nh)^{-1} K((x-X_i)h^{-1})$ del núcleo normal, i=1,...,n, siendo n = 5; la curva más alta corresponde al promedio de estas cinco curvas dando lugar al estimador $f_n(x)$.
La Fig.4(b) presenta un estimador más suave que el primero, ya que el h es mayor. En cada uno de ellos el estimador $f_n(x)$ es mayor para los x cercanos a los X_i, i=1,...,5. Este ejemplo sirve sólo para el cometido mencionado, ya que n = 5 es una muestra demasiado pequeña para encontrar un estimador.

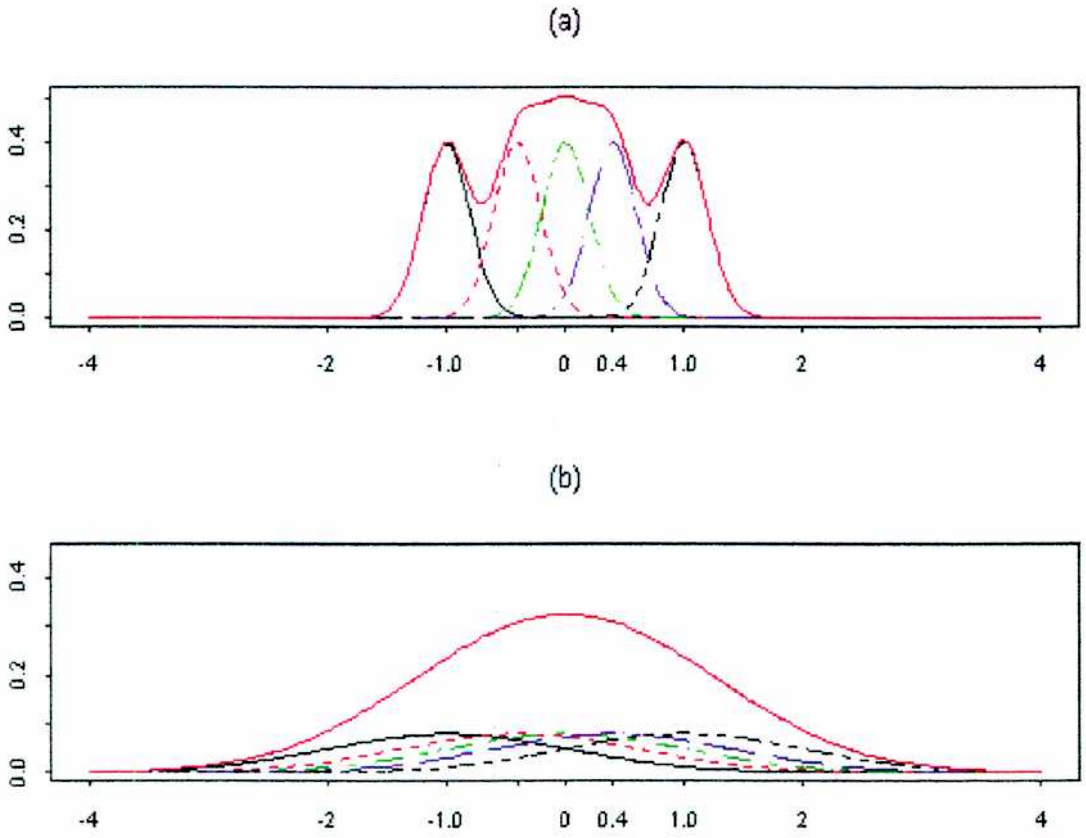

Fig.4. Estimador de la densidad, por método de núcleos, basado en el núcleo normal para las observaciones $X_1=-1$, $X_2=-0.4$, $X_3=0$, $X_4=0.4$, $X_5=1$. En (a) h=0.2, (b) h=1.

Debemos tener en cuenta que si los datos provienen de distribuciones de colas pesadas, existe una tendencia a producir mucho ruido en las colas de la estimación, ya que el h es fijo; si para solucionar este inconveniente se aumenta el parámetro de suavizado, los detalles en la parte principal de la distribución se enmascaran (Silverman (1992), pag.18). La Fig.5 muestra este hecho, con una muestra de tamaño 100 de una población distribuída según una densidad chi-cuadrado con tres grados de libertad. Las aberturas elegidas para este caso particular fueron h = 0.2 y h = 1.

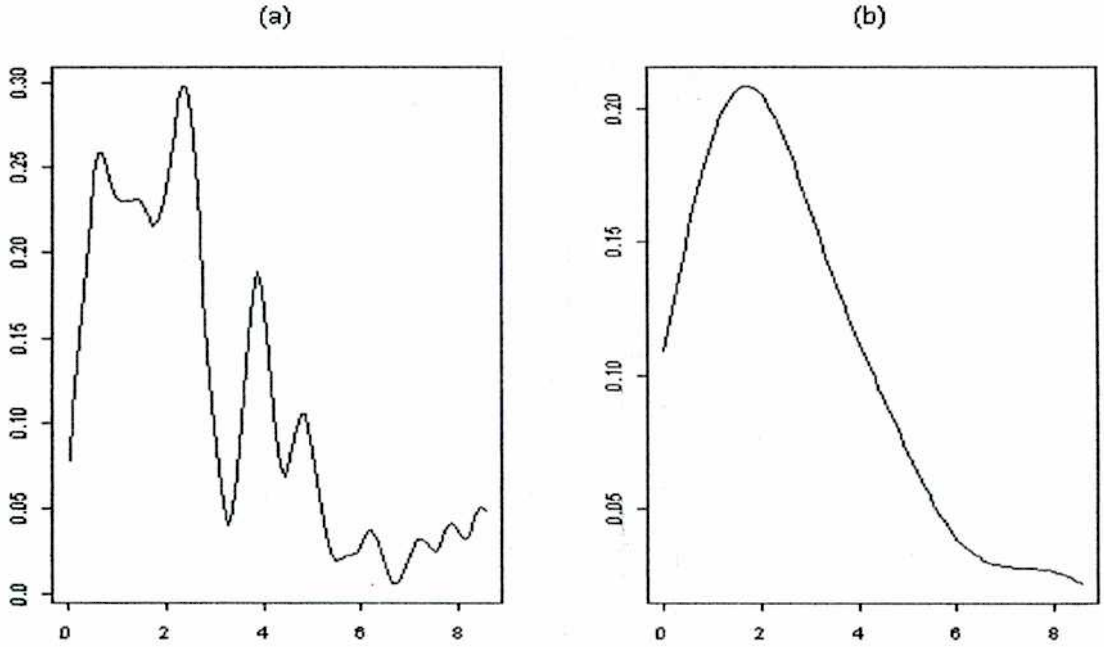

Fig.5. Estimador de la densidad basado en el núcleo normal, para 100 observaciones de una variable chi-cuadrado con 3 grados de libertad. En (a) h=0.2, (b) h=1.

Una descripción de los métodos para seleccionar la abertura de ventana en base a los datos, se da en el Capítulo 3.

5-Estimador de vecinos más cercanos.

Dada una muestra $X_1,\ldots,X_n$ y un punto t de la recta, indiquemos las distancias de t a los valores observados, en forma ordenada ascendente,

$$d_1(t) \leq d_2(t) \leq d_3(t) \leq \ldots \leq d_n(t),$$

donde $d_i(t)$ representa el i-ésimo estadístico de orden de las distancias de t a los puntos de la muestra $X_1,\ldots,X_n$. Llamaremos $d_i(t)$ la distancia de t a la i-ésima observación más cercana.
Si f(t) es la densidad en el punto t, uno esperaría que en el intervalo [t-r, t+r], de longitud 2r, se encontrarían, aproximadamente, 2rnf(t) observaciones.
Además, por definición de $d_k(t)$, hay exactamente k datos en el intervalo $[t-d_k(t);\ t+d_k(t)]$, si no hay empates.
Por lo tanto, un estimador de f(t) puede definirse como aquél que verifique $k = 2n\, d_k(t) f_n(t)$ o equivalentemente,

$$f_n(t) = k/(2n\ d_k(t)). \qquad (22)$$

Esta función es inversamente proporcional a la distancia de t al k-ésimo vecino más cercano. En las colas de la distribución esta distancia puede ser mayor que en la parte principal, lo que evitaría, como advertimos en el estimador de núcleo, la necesidad de suavizar la estimación en las colas.

Este estimador es continuo, pero su derivada presenta discontinuidades en los mismos puntos que d_k y su integral no es igual a la unidad.

Podemos generalizarlo utilizando una función núcleo K,

$$f_n(x) = \sum_{i=1}^{n} K((x-X_i)/d_k(x))/(nd_k(x)), \qquad (23)$$

la cual contiene al anterior (22) para el caso en que el núcleo es uniforme.

Cuán suave sea en su totalidad el estimador depende del número de vecinos k; mientras que el ancho de ventana $H_{k,x} = d_k(x)$ depende de la densidad de las observaciones cercanas a x. La diferencia respecto del estimador de núcleos, está en que el parámetro de suavizado $H_{k,x}$ variará en cada punto estimado.

El aporte del núcleo K al estimador f_n en el punto x, dependerá directamente de cuan próximos estén los datos a dicho punto.

En la Fig.6, se ha representado el estimador de vecinos cercanos $f_n(x)$ de la densidad f(x) donde los datos son los mismos del ejemplo anterior y K es el núcleo gaussiano. La primera curva corresponde al valor k = 2, y para cada x el valor del estimador en ese punto es igual al promedio

$$(1/5) \sum_{i=1}^{5} K((x-X_i)/H_{k,x})/H_{k,x},$$

de cinco densidades normales centradas en X_i, i = 1,...,5, con desvío $H_{k,x}$, calculadas en el punto x. En el segundo gráfico el valor dado a k es 3.

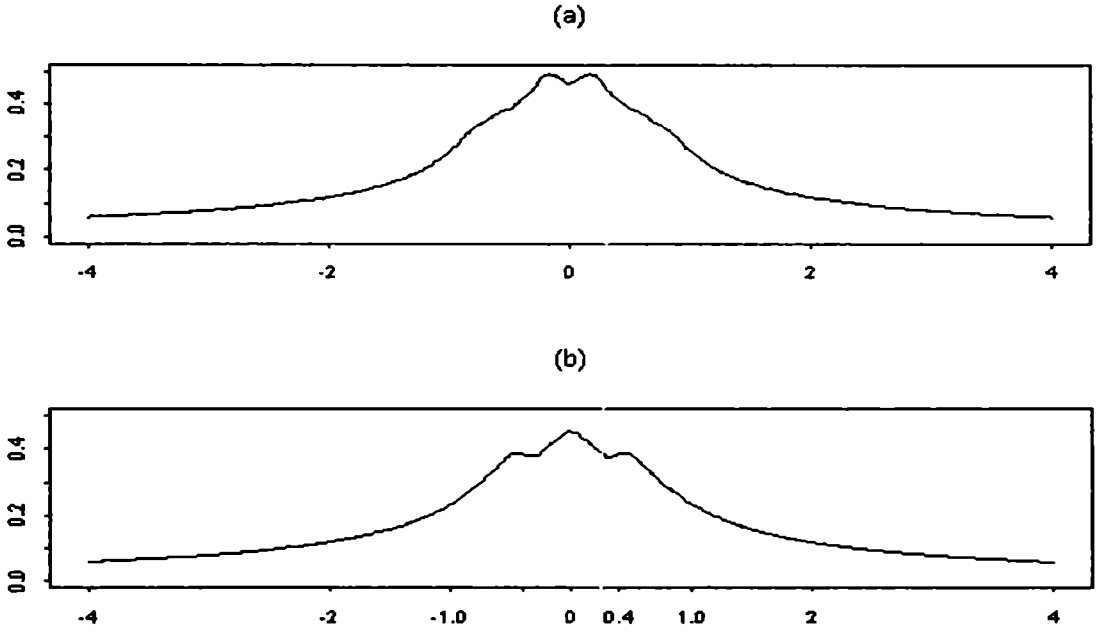

Fig.6. Estimador de la densidad, por método de vecinos cercanos, basado en el núcleo normal para las observaciones $X_1=-1$, $X_2=-0.4$, $X_3=0$, $X_4=0.4$, $X_5=1$. En (a) k=2, (b) k=3.

6-Estimador de núcleo variable.

Este método, como el anterior, adapta la cantidad de suavizado a la densidad local de los datos. El estimador se construye de manera similar al de núcleo, pero el parámetro de escala, de la función núcleo K, variará de una observación a otra.

Si k es un entero positivo menor que n, llamamos $H_{k,j}$ a la distancia de la observación X_j al k-ésimo vecino más cercano entre las restantes n-1 observaciones. Por lo tanto $H_{k,j}$ dependerá de k y variará con X_j , siendo independiente del punto donde se calcula el estimador.

El estimador de f será,

$$f_n(x) = n^{-1} \sum_{j=1}^{n} H_{k,j}^{-1} K((x-X_j)H_{k,j}^{-1}) .$$

El aporte del núcleo K al estimador f_n en el punto x, dependerá directamente de cuan cercanos estén los datos entre sí, en las inmediaciones de x.

7-Estimador de máxima verosimilitud penalizada.

La función de verosimilitud de la densidad f estaría dada por la expresión,

$$L(f|X_1,\ldots,X_n) = \prod_{i=1}^{n} f(X_i),$$

siendo $X_1,\ldots,X_n$ observaciones independientes e identicamente distribuídas con densidad f.
La función de verosimilitud no alcanza el máximo sobre la familia de funciones de densidad sin imponer algunas restricciones.
Si suponemos que $X_i \in (a,b)$ con $a = t_1 < t_2 <\ldots< t_{m+1} = b$, el histograma (1), puede verse como un estadístico de máxima verosimilitud correspondiente al subespacio $L_1(a,b)$ definido por

$$S_0(t_1, t_2,\ldots, t_m) = \{\sum_{i=1}^{m} y_i I_{B_i} / B_i = [t_i,t_{i+1}),\ y_i \in R\},$$

(ver Thompson y Tapia (1992), pag.95).
A fin de imponer regularidad al estimador obtenido, se considera entonces lo que se denomina el logaritmo de la función de verosimilitud penalizada,

$$\mathrm{Log}(L_P\ (f|X_1,\ldots,X_n)) = \sum_{i=1}^{n} \log(f(X_i)) - \alpha \int [f''(x)]^2\, dx$$

donde α es un parámetro de suavidad positivo y el factor que lo acompaña es una medida de la irregularidad de f, y la indicamos por $R(f'')$.
La función f_n que maximiza la función de verosimilitud, o equivalentemente su logaritmo, es el estimador buscado de f.
El parámetro α controla la suavidad del estimador y el primer término de $\mathrm{Log}(L_P)$, la fidelidad a los datos.

Capítulo 3

Elección de la abertura de ventana en estimadores de núcleo

En este capítulo, nos concentraremos en los estimadores de la densidad f basados en núcleos. Estos estimadores que hemos definido en el Capítulo 2 a través de la expresión (14), dependen de un parámetro de suavizado o abertura de ventana que indicaremos por h. Es de suma importancia una elección apropiada de h ya que dicho valor influye en el suavizado de f_n y también en la medida del error de estimación.

Si sólo estamos explorando, a fin de sugerir modelos de densidades, será suficiente elegirlo subjetivamente y dibujar varios estimadores para distintos valores de h, con el fin de elegir el más apropiado al problema particular planteado. Sin embargo, muchas aplicaciones requieren una elección automática del parámetro de suavizado, pudiendo usarse como punto de partida para hacer ajustes subjetivos posteriores.

Generalmente, se quiere hacer referencia al método aplicado para encontrar h, a fin de comparar o informar resultados, lo cual nos lleva a utilizar métodos automáticos, como los que describiremos.

1-Minimizando la expresión del MECI.

Sea $K:R\to R$ un núcleo, o sea, una función tal que $\int K(w)dw=1$ $K\geq 0$. Queremos hallar el valor de h que minimice el MECI como medida de error de estimación.

Si conociéramos f(x), a partir de los valores de la muestra $X_1,\ldots,X_n$, podríamos elegir h de modo tal que minimice la siguiente aproximación del MECI,

$$\text{MECI} \cong \sum_{i=1}^{m} [f_n(t_i)-f(t_i)]^2/m,$$

donde t_i ,i=1,2,...m, son puntos del rango de f, y f_n se calcula en los puntos t_i como el estimador de núcleos dado por (14), es decir,

$$f_n(t_i) = (nh)^{-1} \sum_{j=1}^{n} K((t_i-X_j)h^{-1}) \qquad \text{para } i=1,2,\ldots m.$$

Otra manera de encontrar h es calcular la aproximación asintótica del MECI y estimar, con los algoritmos de validación cruzada o por métodos plug-in, las incógnitas referidas a f o a sus derivadas. Como ya habíamos visto en la Proposición 5,

$$\text{AMECI} = (nh)^{-1} \int K^2(w)\, dw + (1/4)\, h^4\, \sigma_K^4 \int f''^2(x)dx\ ,$$

y el h óptimo que minimiza asintóticamente el MECI está dado por:

$$h_{opt} = \{\ R(K)/\ (n\ \sigma_K^4\ R(f''))\ \}^{1/5},\ \text{donde}\ R(g) = \int g^2(u)\ du\ .$$

Luego, el valor óptimo de MECI asintótico será igual a

$$\text{AMECI}^* = (5/4)\ (\sigma_K\ R(K))^{4/5}\ R(f'')^{4/5}\ n^{-4/5}.$$

Ambas fórmulas dependen de la densidad desconocida, sin embargo pueden sacarse de ellas conclusiones provechosas.
Primero, la abertura de ventana óptima converge a cero cuando aumenta el tamaño de muestra. Segundo, puesto que el término $R(f'')$ mide en algún sentido la rapidez de fluctuación de f, uno podría esperar que valores pequeños de h serían apropiados para densidades muy irregulares.
Los dos sumandos que aparecen en la expresión de AMECI están en una proporción de 4 a 1 cuando $h = h_{opt}$, lo que indica que la integral del sesgo al cuadrado aporta el 20% del AMECI y la integral de la varianza el 80% restante.
Cuando f es aproximadamente normal, una estimación de la ventana, que minimiza la medida asintótica de error AMECI, podría ser $h^* = (4/3)^{1/5}\, s\, n^{-1/5}$, donde s es el desvío estándar de la muestra.
Si f es desconocida, el método plug-in sugiere estimar f'' por medio de un estimador de núcleos con una ventana inicial y posteriormente calcular $R(f'')$, en base al estimador obtenido; una descripción detallada de este procedimiento puede verse en Wand y Jones (1995).
Veremos ahora dos variaciones del algoritmo de validación cruzada, para encontrar el parámetro de suavizado: el de mínimos cuadrados y el de pseudo-verosimilitud. Cao, Cuevas y Gonzalez-Manteiga (1994) presentan comentarios y bibliografía sobre estas y otras aberturas de ventanas.

2-Validación cruzada de mínimos cuadrados (VCMC).

El método VCMC para encontrar la abertura de ventana h, fue sugerido por Rudemo(1982) y Bowman(1984) y está basado en un concepto muy simple que es minimizar la expresión del MECI a partir del ECI, el cual se calcula con una sola muestra,

$$ECI= \int [f_n(x)-f(x)]^2dx = \int f_n^2(x)dx -2 \int f_n(x)f(x)dx + \int f^2(x)dx.$$

Como el último término no depende de f_n, y por lo tanto tampoco de h, sería equivalente a minimizar la expresión,

$$H_h(f_n) = \int f_n^2(x)dx -2 \int f_n(x)f(x)dx.$$

La última integral depende de la densidad desconocida, entonces la idea es construir una estimación de $H_h(f_n)$ a partir de los datos que no dependa de f, y encontrar su mínimo sobre h.
Para ello sea

$$f_{n,-i}(x)= (n-1)^{-1} h^{-1} \sum_{j\neq i}^{n} K((x-X_j)h^{-1}),$$

el estimador de núcleos de f(x) obtenido considerando toda la muestra menos la i-ésima observación.
La expresión $\int f_n(x)f(x)dx$ puede considerarse como la esperanza de $f_n(X)$ respecto a la variable X con igual distribución que las X_i e independiente de ellas. Es decir, $E_X(f_n(X)) = \int f_n(x)f(x)dx$, resultará una variable aleatoria función de la muestra $X_1,\ldots,X_n$.
Por otra parte, $E[K((x-X_j)h^{-1}] = E[K((x-X)h^{-1})]$, para todo j=1,...n y para cada x fijo.
Luego,

$$E[n^{-1} \sum_{i=1}^{n} f_{n,-i}(x)] = E[f_{n,-i}(x)] = E[f_n(x)],$$

para cada x fijo, con la esperanza calculada respecto a las variables de la muestra.
Por otra parte la $E(f_{n,-i}(X_i))$, calculada respecto a todas las variables de la muestra, la podemos descomponer en la esperanza condicional respecto a las variables de la muestra excluída la variable X_i, de la esperanza respecto a la variable X_i, o sea

$$E(f_{n,-i}(X_i)) = E_{X_j \neq X_i}\,[E_{X_i}\,(f_{n,-i}(X_i))] = E_{X_j \neq X_i}[\int f_{n,-i}(x) f(x)\,dx\,]$$

$$= \int E_{X_j \neq X_i}[f_{n,-i}(x)]\; f(x)\,dx = \int E[f_n(x)] f(x)\,dx = E(f_n(x)),$$

donde la última esperanza se toma respecto de todas las variables de la muestra. Rudemo (1982) propuso reemplazar la $\int f_n(x) f(x)\,dx$ por el estimador insesgado

$$n^{-1} \sum_{i=1}^{n} f_{n,-i}(X_i).$$

Luego una aproximación de $H_n(f_n)$ será

$$H_h(f_n) \cong \int f_n^2(x)\,dx - 2\, n^{-1} \sum_{i=1}^{n} f_{n,-i}(X_i) = M(h), \qquad (24)$$

y la última expresión no depende de la densidad f.

El método de validación cruzada de mínimos cuadrados (VCMC) selecciona la abertura de ventana h como la que minimiza la expresión M(h) y la notamos h_{VCMC}. Hall (1983) y Stone (1984) probaron que es un estimador consistente del ancho de ventana óptimo y la distribución asintótica de $M(h_{VCMC})$ y de h_{VCMC} fue obtenida por Hall y Marron (1987b), y Scott y Terrell (1987).

Observemos que $M(h) + \int f^2(x)\,dx$, para todo h, es un estimador insesgado de MECI. Por lo tanto, minimizar E(M(h)) es equivalente a minimizar el MECI sobre h. Por otra parte, uno espera que la ventana que minimiza M(h) sea similar a la que minimiza E(M(h)).

Estas reflexiones nos inducen a pensar que la ventana h buscada por este método es una buena elección del parámetro de suavizado de f_n.

Scott y Terrell (1987) han propuesto una versión modificada del algoritmo VCMC, que llamaron de validación cruzada de mínimos cuadrados sesgada (VCMCS) y se basa en estimar R(f") en las expresiones de la Proposición 5.

Un estimador natural de R(f") sería $R(f_n'')$. Sin embargo, Scott y Terrell (1987) demuestran que $E[R(f_n'')] = R(f'') + R(K'')/(nh^5) + O(h^2)$. Luego $R(f_n'') - R(K'')/(nh^5)$ es un estimador asintóticamente insesgado de R(f") que puede ser utilizado.

El término $R(K'')/(nh^5)$, es asintóticamente una constante representando el ruido que existe en el estimador de núcleos.

Sustituyendo en la Proposición 5 el valor de R(f") por la expresión

$R(f_n'') - R(K'')/(nh^5)$ la estimación del MECI sería,

$$VCS(h) = R(K)/(nh) + [\sigma_k^4/(2n^2h)] \sum_{i<j} \psi(\Delta_{i,j}), \qquad (25)$$

donde

$$\psi(\Delta) = \int K''(w)\ K''(w+\Delta)\ dw$$

y

$$\Delta_{i,j} = (X_i - X_j)/h.$$

En el caso en que la función K sea el núcleo normal la expresión para VCS(h) es la siguiente

$$VCS(h) = 1/(2nh\pi^{1/2}) +$$

$$+(1/(64n^2h\pi^{1/2})) \sum_{i<j}(\Delta_{i,j}^4 - 12\Delta_{i,j}^2 + 12)\exp(-\Delta_{i,j}^2/4). \qquad (26)$$

Puesto que $\lim_{h \to \infty} VCS(h) = 0$, el h se elige como el mayor valor de un mínimo local.

3-Validación cruzada de máxima verosimilitud penalizada (VCMVP).

El método de validación cruzada de máxima verosimilitud (VCMV) está basado en la idea de considerar la abertura de ventana h, como un parámetro que puede ser estimado por el procedimiento de máxima verosimilitud.
Considerando que f es desconocida, la función de verosimilitud tendría la forma,

$$L(h) = \prod_{i=1}^{n} f_n(X_i, h)$$

donde $f_n(X_i, h)$ es un estimador de f, con parámetro desconocido h.
Este problema lleva a elegir como ventana h que maximiza la verosimilitud, a h = 0 ya que si $h \to 0$ la contribución a la verosimilitud en $x = X_i$ se hace infinita. Es por ello que se modificó el criterio imponiendo que el estimador de núcleos de f en cada observación X_i, se obtenga a partir de las n-1 observaciones restantes,

$$f_{n,-i}(X_i,h) = (n-1)^{-1} h^{-1} \sum^{n} K((X_i-X_j)h^{-1}).$$

De esta forma, la función a maximizar, con respecto a h, será finalmente,

$$L^{*}(h) = \prod_{i=1}^{n} f_{n,-i}(X_i,h), \qquad (27)$$

también llamada de validación cruzada de pseudo-máxima verosimilitud o de verosimilitud penalizada. El valor de h que la maximiza lo indicamos h_{MVP}.

La idea básica de este método fue sugerida por Duin(1976) y Habbema, Hermans y van den Broek (1974). Propiedades de consistencia y comportamiento asintótico fueron presentadas por diversos autores, entre ellos Chow, Geman y Wu (1983), Devroye y Györfi (1985) y Marron (1985).

R. Cao, A. Cuevas y W. Gonzalez-Manteiga (1994) presentan comentarios y bibliografía sobre estas dos aberturas de ventanas, (ver también Izenman (1991)).

Capítulo 4

Densidad relativa

1-Planteo del problema.

Sea F_0, la función de distribución acumulada (fda) de una variable aleatoria de referencia X_0, y F la fda de una variable aleatoria (v.a.) de comparación X, ambas continuas.
En el caso de estar interesados en estudiar las diferencias entre las dos distribuciones, un camino sería definir una v.a. R que mida la relación entre X_0 y X: $R = F_0(X)$. R es una v.a. continua con rango en el intervalo [0,1]. Puesto que, R mide la distribución relativa de X respecto de X_0, nos referiremos a ella como la variable relativa. Su función de distribución acumulada G está dada por $G(r) = P(R \leq r) = P(F_0(X) \leq r) = P(X \leq F_0^{-1}(r)) = F(F_0^{-1}(r))$ para $0<r<1$, donde $F_0^{-1}(r) = \inf\{ y : F_0(y) \geq r\}$.
Sean f y f_0 las densidades de X y X_0 respectivamente. En este caso, la v.a. R tiene una densidad, llamada densidad relativa, que está dada por la expresión

$$g(r) = f(F_0^{-1}(r)) / f_0(F_0^{-1}(r)) \qquad \text{para } 0 < r < 1,$$

Si $F_0 = F$, entonces g es la densidad de una v.a. uniforme en (0,1).
A partir de la expresión de la densidad g, es posible interpretar la relación entre las distribuciones de X y X_0, ya que es el cociente entre ambas densidades, calculadas en el punto $F_0^{-1}(r)$.
Para los valores de r, donde $g(r) > 1$, se cumple que $f(F_0^{-1}(r)) > f_0(F_0^{-1}(r))$, o sea f se mantiene mayor que f_0 en los puntos $x = F_0^{-1}(r)$. Por otra parte, si g(r) toma valores cercanos a 1, la f se mantendrá cercana a f_0 en los puntos mencionados. Observamos por lo tanto, que cuando g(r) se mantiene menor que uno, la función f_0 se mantiene mayor que f en los puntos $F_0^{-1}(r)$, mientras que cuando g(r) es mayor que uno, la densidad f_0 es menor que f en los puntos $F_0^{-1}(r)$. El conocimiento de la densidad relativa nos permitirá entonces comparar f con f_0.
Las figuras 7, 8 y 9 dan los gráficos de la densidad relativa para distintas situaciones de f y f_0, correspondientes a las variables

(i) $X = N(0,1)$, $X_0 = N(1,2^2)$,

(ii) $X = N(0,1)$, X_0 = t-Student con k grados de libertad, k=1,2,

(iii) $X = N(0,1)$, X_0 = Chi-cuadrado con 3 grados de libertad.

También se representan las densidades f y f_0 que dan lugar a g. Las distintas densidades f_0, están marcadas con líneas cortadas y las correspondientes f con líneas llenas, las densidades relativas g correspondientes están en los gráficos de la derecha.

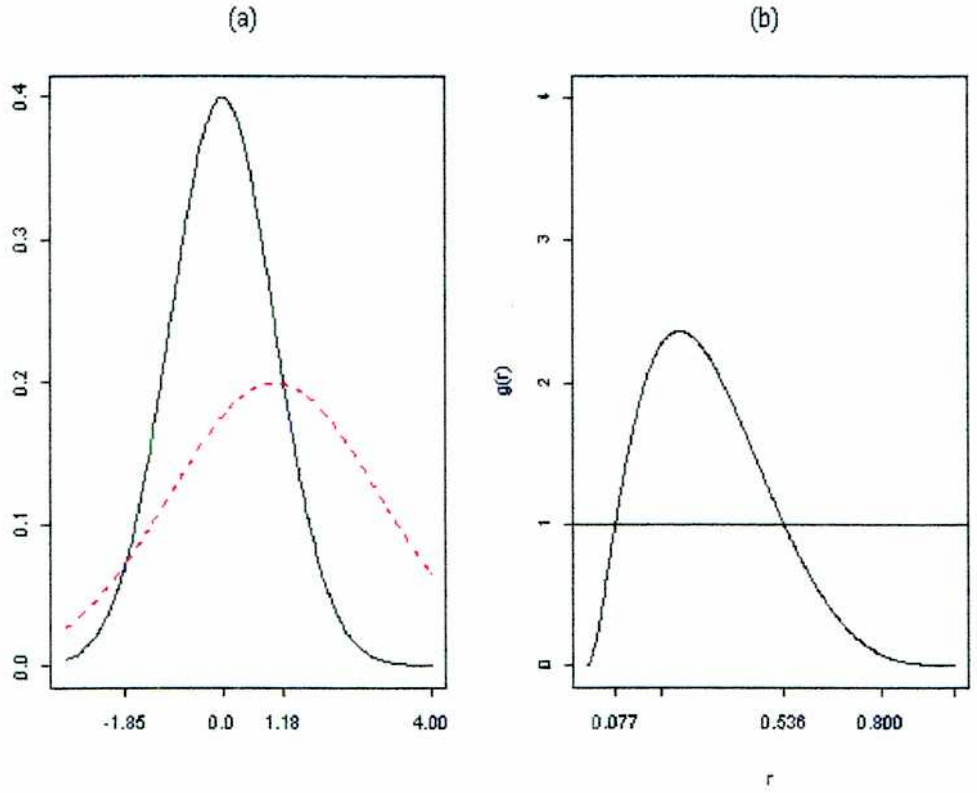

Fig.7. (a) Gráfico de densidades normales; con línea llena representamos la densidad normal estándar y con líneas cortadas la densidad normal $N(1,2^2)$. (b) Gráfico de la correspondiente densidad relativa. En el eje horizontal se marcaron los valores 0.077 y 0.536 donde la densidad es igual a 1.

En la Fig.7(a) se presenta el gráfico de la densidad f_0 de una variable X_0 con distribución $N(1,2^2)$ y de la densidad f de una normal estándar X.

La función g(r) correspondiente, representada en Fig.7(b), se mantiene menor que 1, para r entre 0 y 0.077. A partir del punto $F_0^{-1}(0.077) = -1.847$, f_0 comienza a ser menor que f. Por otra parte, para x entre -1.847 y 1.181, f(x) es mayor que $f_0(x)$, y por lo tanto, g(r) es mayor que 1 para r entre $F_0(-1.847) = 0.077$ y

$F_0(1.181) = 0.536$. A partir de $r = 0.536$, $g(r)$ se mantiene menor que 1, siendo $F_0^{-1}(0.536) = 1.181$ el punto en donde f comienza a ser menor que f_0.

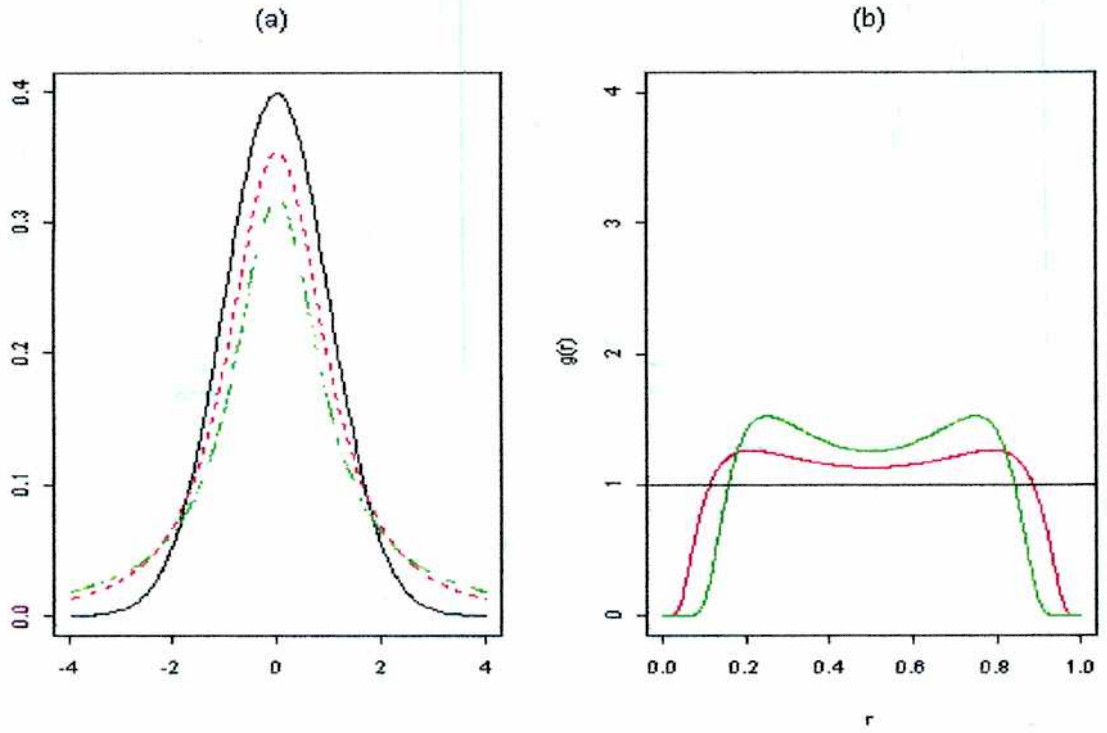

Fig.8. (a) Gráfico de densidades; con línea negra representamos una normal estándar, con roja una t-Student con dos grados de libertad y con verde una Cauchy. (b) Gráfico de la densidad relativa, cuando comparamos la normal con t_2, línea roja, y con Cauchy, línea verde.

En la Fig.8 estamos comparando una variable normal estándar con una t-Student con dos grados de libertad (t_2) y con una Cauchy (t_1). Se observa en el gráfico de las densidades relativas correspondientes, que existe una mayor diferencia de la variable normal respecto de la Cauchy que con respecto a la variable t_2, dada por el mayor alejamiento del valor 1. La densidad normal se muestra mayor que las restantes densidades en los valores centrales, mientras que en los extremos se encuentra por debajo de las densidades comparadas. Esto estaría explicado por el gráfico de ambas densidades relativas, respecto a su relación con el valor 1.

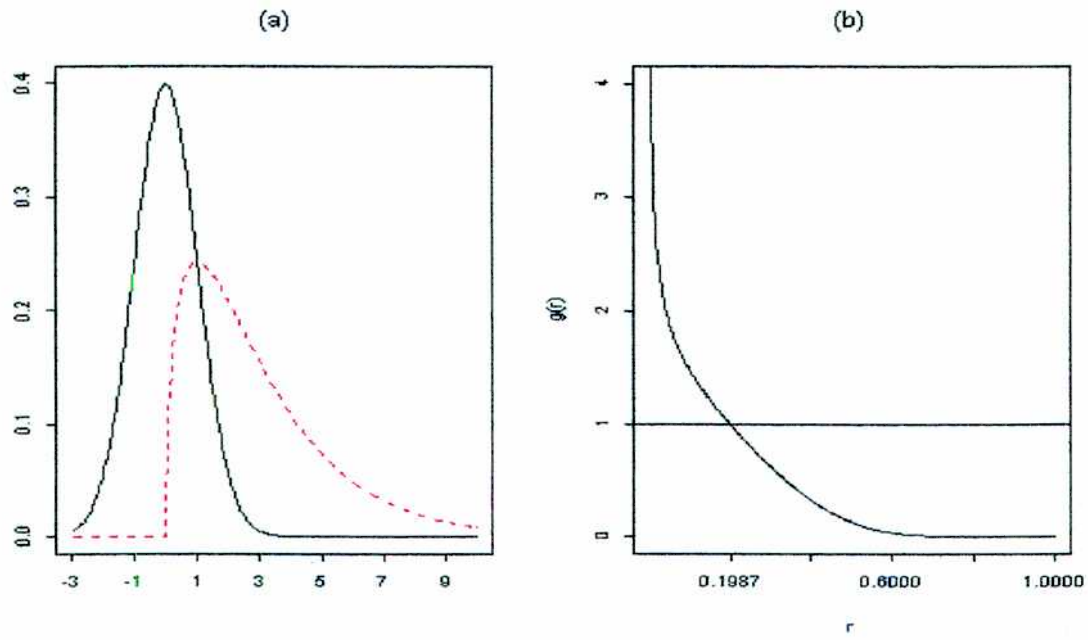

Fig.9. (a) Gráfico de densidades; con línea llena representamos una densidad normal estándar y con líneas cortadas la densidad de una variable chi-cuadrado con 3 grados de libertad. (b) Gráfico de la correspondiente densidad relativa. En el eje horizontal está marcado el valor 0.1987 en donde la densidad es igual a 1.

En la Fig.9(b), se observa que cerca del origen g(r) toma valores relativamente grandes, lo que indica que f es bastante mayor que f_0. La densidad relativa es mayor que 1 hasta r = 0.1987, por lo tanto, como $F_0^{-1}(0.1987) = 1$, a partir de 1 f es menor que f_0.

Si comparamos los tres gráficos de g, es evidente que el segundo indica una mayor semejanza entre f y f_0, por estar g(r) muy cercano a uno en la casi totalidad de su rango.

Respecto a la interpretación de la función de distribución acumulada relativa, podemos decir que $G(r) = F(F_0^{-1}(r))$ representa la proporción de la variable X que está debajo del percentil $F_0^{-1}(r)$ de la variable X_0.

Más precisamente, G(r) es la proporción de valores de la v.a. X que están por debajo del valor $x_r = F_0^{-1}(r)$ de la v.a. X_0, el cual deja una proporción r de valores de X_0 por debajo de él.

El gráfico de G(r) nos dará un "p-p plot" de F contra F_0 o sea el gráfico de los pares $(F_0(x),F(x))$, $x \in R$, coincidirá con el de los pares $(r,G(r))$, $0<r<1$.

La densidad relativa ha sido estudiada, entre otros, por Eubank, LaRiccia y Rosenstein(1987), Parzen (1992) y Handcock y Janssen (2000).

2-Estimación de la Densidad Relativa.

En la Sección 3, vamos a estimar, a partir de muestras aleatorias simuladas de X y X_0, la densidad relativa g aplicando como procedimientos posibles el método de núcleos y el método de vecinos más cercanos para estimar la densidad, combinados con estimadores de la función de distribución. Ello implicará la elección de algoritmos para estimar la ventana h o la cantidad k de vecinos más cercano.
El núcleo K elegido en este trabajo es la función de densidad normal estándar, que indicaremos por φ.
Si las muestras son de tamaño n y m, para cada una de las poblaciones, indicamos con $X_1,\ldots,X_m$, la muestra aleatoria de la variable de comparación X y con $X_{0,1},\ldots.X_{0,n}$, la de la variable aleatoria de referencia X_0.
Al estimador de F_0 lo notaremos $F_{0,n}$ o $F_{0,n,h}$, como se definirá más adelante, según si consideramos la función de distribución empírica o una versión suavizada de la misma con abertura de ventana h. Luego, los predictores naturales de R serán $R_{i,m} = F_{0,n}(X_i)$ o $R_{i,m,h} = F_{0,n,h}(X_i)$, para i=1,2,...m, y por lo tanto, no serán variables independientes.
Entonces, dadas muestras aleatorias de X y X_0, obtendremos valores $R_{i,m}$ o $R_{i,m,h}$, a partir de los cuales estimaremos g(r) por $g_{n,m}(r)$, donde $g_{n,m}(r)$ será un estimador basado en núcleos como el definido en (14) construído a partir de los valores $R_{i,m}$ o un estimador de vecinos más cercanos con núcleos como el definido en (23) obtenido a partir de los valores $R_{i,m,h}$.
Una vez obtenido el estimador de g(r), mediremos el error de estimación basándonos en un estimador del MECI = E(ECI)= $E\left(\int[g_{n,m}(r)-g(r)]^2dr\right)$, que notaremos $MECI_1$, siendo

$$MECI_1 = (1/s)\sum_{i=1}^{s}(g_{n,m}(r_i)-g(r_i))^2, \qquad (28)$$

donde $r_1,\ldots,r_s$, representa una grilla del intervalo (0,1).

2.1-Método basado en núcleos.

Handcock y Janssen (2000) proponen definir el estimador $g_{n,m}(r)$ mediante la expresión

$$g_{n,m}(r) = (mh)^{-1} \sum_{i=1}^{m} K((r-R_{i,m})h^{-1}) = m^{-1} \sum_{i=1}^{m} K_h(r-R_{i,m}), \quad (29)$$

donde $K_h(t)= K(t/h)/h$ y $h=h_m$ es la apertura de ventana.
Las variables $R_{i,m}$, se calcularon por medio de la relación $R_{i,m} = F_{0,n}(X_i)$, donde $F_{0,n}$ es la distribución empírica obtenida de la muestra $X_{0,1},\ldots,X_{0,n}$.
Ellos mostraron en particular, que el estimador de $G(r)$, dado por $G_{n,m}(r) = F_m(F_{0,n}(r))$, siendo F_m la distribución empírica calculada a partir de la muestra $X_1,\ldots,X_m$, tiene una distribución asintótica conjuntamente normal. La distribución asintótica marginal fue obtenido previamente por Gastwirth (1968).
El estimador $g_{n,m}(r)$ dado por (29) es asintóticamente insesgado. Más aún, en su Teorema 2, Handcock y Janssen (2000) establecen que bajo condiciones de regularidad en el núcleo K y las densidades f y f_0, de X y X_0 respectivamente, se tiene que,

$$(mh_m)^{1/2}(g_{n,m}(r)-g(r)) \rightarrow N(0,g(r)R(K)+v^2g^2(r)R(K)),$$

cuando $m/n\rightarrow v^2$, $h_m\rightarrow 0$, $mh_m^3\rightarrow\infty$, $mh_m^5\rightarrow 0$, donde $R(K)=\int K^2(z)dz$.
A partir de este resultado, los mencionados autores obtienen intervalos de confianza.
La abertura de ventana h puede elegirse mediante alguno de los siguientes tres procedimientos: el que minimiza el MECI o el AMECI, que notamos h_{opt}; el de validación cruzada de mínimos cuadrados sesgada que notamos h_{vcs} y el de máxima verosimilitud penalizada o pseudo máxima verosimilitud, que notamos h_{MVP}, los cuales han sido definidos en el Capítulo 3.

2.2-Método de vecinos más cercanos con núcleos.

Otra propuesta para $g_{n,m}(r)$ está basada en el estimador de la densidad de vecinos más cercanos con núcleos. En este caso, se define

$$g_{n,m}(r) = \sum_{i=1} K[(r-R_{i,m,k}))H_{k,r}^{-1}]/(mH_{k,r}). \quad (30)$$

Como en el caso anterior, K es el núcleo normal, pero en este caso, $H_{k,r}$ representa la distancia de r al k-ésimo vecino más

cercano entre los $R_{i,m,h}$, $i=1,\ldots,m$, predictores de la variable relativa R.

Los $R_{i,m,h}$ serán calculados para este método, usando un estimador suavizado de la función de distribución acumulada, que fue estudiado por Fernholz (1991). Mehra y Krishnaiah (2002) estudiaron estimadores suavizados de la función de distribución condicional.

Daremos a continuación, una breve descripción de estos estimadores. Sabiendo que,

$$F_0(x)=\int_{-\infty}^{x} f_0(t)\,dt,$$

un estimador de F_0 puede calcularse a partir de un estimador de núcleos de la densidad f_0 como,

$$F_{0,n,h}(x) = \int_{-\infty}^{x} f_{0,n}(t)\,dt = \int_{-\infty}^{x} (1/n)\sum_{i=1}^{n} K_h(t-X_{0,i})\,dt =$$

$$= (1/n)\sum_{i=1}^{n} \int_{-\infty}^{x} K[(t-X_{0,i})/h]\,dt$$

$$= (1/n)\sum_{i=1}^{n} \int_{-\infty}^{(x-X0,i)/h} K(u)\,du = (1/n)\sum_{i=1}^{n} K_1[(x-X_{0,i})/h],$$

donde, en nuestro caso, $K_1=\phi$ es la fda de la variable normal estándar.

La abertura de ventana h elegida en la expresión suavizada $F_{0,n,h}$ corresponde al mínimo de una aproximación del MECI, encontrado por método Monte Carlo, o al h óptimo (h_{opt}), como puntualizaremos más adelante.

De esta forma se define

$$R_{j,m,h} = F_{0,n,h}(X_j)= (1/n)\sum_{i=1}^{n} \phi[(X_j - X_{0,i})/h],\ j=1,\ldots,m. \quad (31)$$

El siguiente resultado muestra que el estimador definido por

$$g_{n,m}(r) = g_{n,m}(r;h,v) = \sum_{i=1}^{m} K[(r-R_{i,m,h})v_m^{-1}]/(mv_m),$$

siendo v_m una sucesión numérica positiva dependiendo de m, resulta ser consistente.

Proposición 1. *Sean $X_1,\ldots, X_m$, $X_{0,1},\ldots, X_{0,n}$, variables aleatorias independientes tales que $X_i \sim F$, $1\le i\le m$, $X_{0,j}\sim F_0$, $1\le j\le n$. Sea*

$$F_{0,n,h}(x) = (1/n)\sum_{i=1}^{n} K_1[(x-X_{0,i})/h].$$

Supongamos que F_0 es Lipschitz en **R**, *entonces,*

(i) $\sup_x |F_{0,n,h}(x)-F_0(x)| \to 0$, *casi seguramente (a.s.)*

si $h=h_n$, cumple que $h_n\to 0$ y $nh_n\to+\infty$.

(ii) Supongamos que f_0 es Lipschitz en **R**. *Sea $g_{n,m}(r;h,v)$ el estimador de $g(r)$, dado por,*

$$g_{n,m}(r;h,v) = \sum_{i=1}^{m} K[(r-R_{i,m,h}))v_m^{-1}]/(mv_m).$$

Si $h_n\to 0$, $nh_n\to+\infty$, $nh_n^2\to 0$, $v_m\to 0$, $mv_m^4\to+\infty$, $n/(n+m) \to \lambda$, $0<\lambda<1$, K derivable con derivada continua y acotada, entonces $g_{n,m}(r;h,v)\to g(r)$ en probabilidad.

Demostración.

(i) Se deduce del Teorema de Glivenko-Cantelli (Loeve(1962), pag.20), del Teorema 3 de Devroye y Penrod (1984a) y de la igualdad (Devroye y Penrod (1984b) pag.1252),

$$\sup_B \left|\int_B f_n(x)dx - \int_B f(x)dx\right| = (1/2)\int |f_n(x)-f(x)|dx,$$

(ii) El estimador $g_{n,m}(r)= g_{n,m}(r;h,v)$, lo podemos escribir como,

$$g_{n,m}(r;h,v) = \sum_{i=1}^{m} K[(r-F_{0,n,h}(X_i))v_m^{-1}]/(mv_m)$$

$$= \sum_{i=1}^{m} K[(r-F_0(X_i)+F_0(X_i)-F_{0,n,h}(X_i))v_m^{-1}]/(mv_m)$$

Haciendo un desarrollo de Taylor de orden 1, si η_i indican puntos intermedios obtenemos

$$g_{n,m}(r;h,v) = 1/(mv_m)\sum_{i=1}^{m} \{K[(r-F_0(X_i))v_m^{-1}] +$$

$$+K'(\eta_i/v_m)[F_0(X_i)-F_{0,n,h}(X_i)]/v_m\}$$

$$= 1/(mv_m)\sum_{i=1}^{m} K[(r-F_0(X_i))v_m^{-1}] + G_m.$$

Por el Teorema 3 de Devroye y Penrod (1984a), el primer sumando de la expresión anterior converge a $g(r)$ en probabilidad si $mv_m \to +\infty$. Faltaría ver que $G_m \to 0$ en probabilidad.

$$|G_m| = |(1/mv_m)\sum_{i=1}^{m} K'(\eta_i/v_m)[F_0(X_i)-F_{0,n,h}(X_i)]/v_m|$$

$$\leq (1/mv_m^2)\|K'\|_\infty \sum_{i=1}^{m} |F_0(X_i)-F_{0,n,h}(X_i)|$$

$$\leq (1/v_m^2)\ \|K'\|_\infty \sup_x |F_0(x)-F_{0,n,h}(x)|$$

$$= (1/v_m)^2\|K'\|_\infty\ \|F_0-F_{0,n,h}\|_\infty.$$

Como $nh_n \to +\infty$, $n^{1/2}h_n \to 0$, el Teorema 2.3 de Fernholz (1991), implica que $n^{1/2}\|F_{0,n}-F_{0,n,h}\|_\infty \to 0$, en casi todo punto. Por otra parte, $D_n = n^{1/2}\|F_0-F_{0,n}\|_\infty$ es acotado en probabilidad. Luego, $D_{n,h} = n^{1/2}\|F_0-F_{0,n,h}\|_\infty$ es acotado en probabilidad. Obtenemos, por lo tanto,

$$|G_m| \leq (1/v_m)^2\|K'\|_\infty\ \|F_0-F_{0,n,h}\|_\infty$$

$$= n^{-1/2}v_m^{-2}\ \|K'\|_\infty\ n^{1/2}\|F_0-F_{0,n,h}\|_\infty = n^{-1/2}v_m^{-2}\ \|K'\|_\infty\ D_{n,h}.$$

Como $\lambda_n = n/(n+m) \to \lambda$, $0<\lambda<1$, tenemos

$|G_m| \leq (m^{-1/2}v_m^{-2})(1-\lambda_n)^{1/2}\lambda_n^{-1/2}\ \|K'\|_\infty\ D_{n,h}$, con lo cual como $m^{1/2}v_m^2 \to +\infty$, se obtiene el resultado. ■

<u>Observación</u>. Usando argumentos análogos a los del Teorema 2.3 y Teorema 3.1 de Devroye (1987) se podría obtener que $G_m \to 0$, en casi todo punto, si para algún $\alpha<1/2$, $m^\alpha v_m^2$ está acotado y $mv_m/\log(m) \to \infty$, pero la demostración está más allá del objetivo de esta tesis.
El estimador que nosotros consideramos en (30) utiliza una ventana v_m aleatoria e igual a $H_{k,r}$. Teniendo en cuenta la convergencia uniforme de $F_{0,n,h}$ a F_0, es de esperar que el estimador dado en (30) se comporte como el estimador obtenido cuando la ventana v_m se toma igual a la distancia de r a su k-ésimo vecinos más cercano entre $F_0(X_j)$, $j=1,\ldots,m$; para los cuales vale un resultado análogo al de la Proposición 1(ii), ya que puede utilizar los resultados de Moore y Yackel (1977) si el número de vecinos cumple que $k_m/m \to 0$ y $k_m^4/m^3 \to +\infty$.

Respecto a la cantidad de vecinos k o a la distancia $H_{k,r}$ que aparece en la expresión (30) de $g_{n,m}(r)$, se puede determinar de tres maneras diferentes: a partir de la ventana óptima h_{opt} que minimiza el MECI o AMECI y considerando que $k/(2nH_{k,r}) \rightarrow f(x)$ en probabilidad (Moore y Yackel (1976)) se puede igualar $h_{opt}=H_{k,r}$ o bien por el procedimiento de validación cruzada de mínimos cuadrados o por el de pseudo máxima verosimilitud.

3-Estudio de Monte Carlo.

En todos los casos se hicieron 500 replicaciones de muestras aleatorias de tamaño n = m = 100 correspondientes a las distribuciones que se enumeran a continuación, con el fin de obtener los predictores $R_{i,m}$ o $R_{i,m,h}$, i=1,2,...100. El estimador que hemos notado como $g_{n,m}(r)$ en (29) o (30), lo indicaremos $g_n(r)$ ya que n=m.

<u>Caso 1</u>: $F = F_0 \sim N(0,1)$,
<u>Caso 2</u>: $F \sim N(0,1)$, $F_0 \sim N(1,2^2)$,
<u>Caso 3</u>: $F \sim N(0,1)$, $F_0 \sim t_2$ (t-Student con 2 grados de libertad).

Aplicamos los siguientes dos procedimientos para estimar la densidad relativa,

a) Estimador de densidad basado en núcleo combinado con distribución empírica dado en (29), que llamaremos *estimador de núcleo*

b) Estimador de densidad basado en vecinos más cercanos con núcleo, combinado por un estimador suavizado de la distribución empírica dado en (30), que llamaremos *estimador de vecinos más cercanos con núcleos.*

Para cada procedimiento calculamos tres tipos de ventanas,

a1) Ventana fija u óptima,

a2) Validación cruzada de mínimos cuadrados,

a3) Validación cruzada de pseudo máxima verosimilitud.

Presentaremos los valores descriptivos y gráficos del estimador del error cuadrático medio integrado (MECI), que notamos $MECI_1$ en

la expresión (28), para las distintas poblaciones simuladas, métodos de estimación de la densidad relativa y procedimientos de cálculo de ventanas.

3.1-Estimador de núcleos.

Se estimó F_0 por medio de la acumulada empírica, $F_{0,n}$, calculada a partir de las observaciones independientes $X_{0,1},\ldots X_{0,n}$. Una vez obtenidos los valores $R_{i,m} = F_{0,n}(X_i)$, calculamos $g_n(r)$ mediante la expresión (29), con n=m, para r variando entre 0.01 y 0.09 con diferencias de un centésimo. Para ello calculamos la abertura de ventana según los tres algoritmos citados anteriormente. Una vez obtenido el estimador, calculamos el error aproximado $MECI_1$, según la expresión (28).
Podemos mencionar previamente el resultado a que hace referencia Scott ((1992), pag.165), sobre una ventana que suaviza más, o sobresuaviza, el gráfico del estimador y que indicaremos como h_{os}, siendo mayor que la ventana h_{opt} de la Proposición 5, y se calcula a partir de la expresión de h_{opt}, reemplazando $R(f'')$ por una cota inferior. El valor de esa ventana, para el núcleo gaussiano, es $h_{os} = 1.144\sigma_R n^{-1/5} = 0.4554\sigma_R$. Para las distribuciones del Caso 1, R resulta ser una variable aleatoria uniforme en (0,1) con lo cual $\sigma_R = 1/\sqrt{12} = 0.28867$, y por lo tanto $h_{os} = 0.1314$.
Por otra parte, para el Caso 2, $\sigma_R = 0.1634$, obteniendose $h_{os} = 0.0744$, y para el Caso 3, $\sigma_R = 0.2460$, con lo cual $h_{os} = 0.1120$

Ventana óptima.
Para el Caso 1 se calculó h, minimizando el MECI, mientras que para el Caso 2 y 3 se minimizó el AMECI utilizando la Proposición 5 del Capítulo 2. Para estos dos últimos casos la expresión será, $h_{opt} = [R(\varphi)/(n\sigma_\varphi^4\ R(g'')]^{1/5}$, siendo φ el núcleo gaussiano.
Luego, como $R(\varphi) = \int \varphi^2(u)du = 1/(2\pi^{1/2})$ y $\sigma_\varphi^4 = 1$; queda por calcular $R(g'')$ para esos dos casos.
Recordemos que $g(r)= f(F_0^{-1}(r))/f_0(F_0^{-1}(r))$.

Para el Caso 2, tenemos

$$g(r) = \sigma\ \varphi(\sigma\ \phi^{-1}(r) + \mu)/\varphi(\phi^{-1}(r))$$

y

$$g'(r) = g(r)\,[\phi^{-1}(r)(1-\sigma^2) - \sigma\mu]/\varphi(\phi^{-1}(r)),\ \text{con } \sigma=\sigma_{X_0}=2\ .$$

Para calcular $g''(r)$ y luego integrar su cuadrado, hicimos el cambio de variable $r = \phi(s)$ lo que implica que $dr/ds = \varphi(s)$. Luego,

$$\int_0^1 [g''(r)]^2\, dr = \int_{-\infty}^{+\infty} [g''(r(s))]^2\, \varphi(s)\, ds$$

$$= \int_{-\infty}^{+\infty} [\sigma^2\, \varphi^2(\sigma s + \mu)\{s^2 a - sb + c\}^2/\varphi^5(s)]\, ds,$$

donde $a = (1 - \sigma^2)(2 - \sigma^2)$, $b = \mu\sigma(3 - 2\sigma^2)$ y $c = \mu^2\sigma^2 + (1 - \sigma^2)$, siendo en este caso $\sigma = 2$ y $\mu = \mu_{X0} = 1$.
El valor óptimo de ventana dio como resultado $h_{opt} = 0.04927824$.

Por otra parte para el Caso 3, tenemos $f(u) = (2\pi)^{-1/2}\exp(-u^2/2)$, $f_0(u) = (2\pi)^{-1/2}\,\Gamma(3/2)(1+u^2/2)^{-3/2}$. De donde,
$f(u)/f_0(u) = [\Gamma(3/2)]^{-1}(1+u^2/2)^{3/2}\exp(-u^2/2)$, con $u = u(r) = F_0^{-1}(r)$.
Utilizando la regla de la cadena obtenemos $g''(r) = H''(u)\,(u')^2 + H'(u)u''$, donde $u' = 1/f_0(u)$ y

$$\int_0^1 [g''(r)]^2\, dr = \int_{-\infty}^{+\infty} [g''(r(u))]^2\, u'\, du = 16009.34,$$

por lo tanto $h_{opt} = 0.0445866$.

Finalmente, se da a continuación la derivación de la ventana óptima para el Caso 1,

$$MECI = E(ECI) = E\left(\int_0^1 [g_n(r)-g(r)]^2 dr\right) = \int_0^1 E\left([g_n(r)-1]^2\right) dr)$$

$$= \int_0^1 \{var[g_n(r)] + sesgo^2(g_n(r))\}\, dr. \qquad (32)$$

Por la igualdad (15) del Capítulo 2, se obtiene,

$$E[g_n(r)] = (1/h)\int_0^1 \varphi((r-t)/h)\; g(t)\; dt = \int_{(r-1)/h}^{r/h} \varphi(w)\; dw$$

$$= \phi(r/h) - \phi((r-1)/h),$$

con lo cual

$$Sesgo^2(g_n(r)) = [\phi(r/h) - \phi((r-1)/h) - 1]^2 \qquad 0<r<1 \qquad (33)$$

Por otra parte, por (17),

$$Var(g_n(r)) = Var(K_h(r-R))/n = (E[K_h^2(r-R)] - \{E[K_h(r-R)]\}^2)/n$$

$$= (\int_0^1 K_h^2(r-w)g(w)dw - \{\phi(r/h) - \phi((r-1)/h)\}^2)/n$$

$$= ((1/h)\int_{(r-1)/h}^{r/h} K^2(t)g(r-th)dt - \{\phi(r/h) - \phi((r-1)/h)\}^2)/n$$

$$= ((1/h)\int_{(r-1)/h}^{r/h} \varphi^2(t)\ dt - \{\phi(r/h) - \phi((r-1)/h)\}^2)/n$$

$$= (2^{-1}h^{-1}\pi^{-1/2}[\phi(r2^{1/2}/h) - \phi((r-1)2^{1/2}/h)] - \{\phi(r/h) - \phi((r-1)/h)\}^2)/n \quad (34)$$

Las expresiones (33) y (34) se integran para encontrar el MECI tal como indica la expresión (32), dando a h sucesivamente valores entre 0.01 y 0.99, con diferencias de un centésimo.
El MECI mínimo se alcanza cuando $h = h_{opt} = 0.11$ y su valor es 0.04126.
El siguiente cuadro resume los resultados anteriores, con valores de ventanas.

Ventana	Caso 1	Caso 2	Caso 3
h_{os}	0.1314	0.0744	0.1120
h_{opt}	0.1100	0.0493	0.0446

Ventana de validación cruzada de mínimos cuadrados sesgada.
Como vimos en el Capítulo 3, es posible aplicar la versión modificada de Scott y Terrell (1987) para encontrar la ventana en la expresión de $g_n(r)$.
Para cada conjunto de valores $R_{i,m}$, i=1,...,100, encontramos la ventana h_{VCS} que minimiza la expresión VCS(h) (26), donde los posibles valores de h varían entre 0.01 y 0.09, o entre 0.01 y h_{os}. Con ese valor calculamos $g_n(r)$, según (29), y la aproximación del error dado por el $MECI_1$.
Cabe aclarar que aquí consideramos dos tratamientos para los posibles valores de ventanas que minimizan VCS(h), uno siguiendo la advertencia de Scott ((1992), pag.166) en cuanto a que el valor de ventana h_{VCS} debe ser menor que el valor de h_{os}, y otro sin imponer una acotación a los posibles valores de h. Cuando

consideramos la recomendación mencionada, al repetir el procedimiento acotando los posibles valores de h superiormente con h_{os}, las ventanas que minimizan la expresión de VCS(h) coinciden, con h_{os} en un 93.8% para el Caso 1, en un 80% para el Caso 2 y en un 92.2% para el Caso 3.

Ventana de validación cruzada de pseudo máxima verosimilitud.
A partir de la expresión (27), explicitamos L(h) para el caso que nos compete, es decir

$$L(h) = \prod_{i=1}^{n} g_{n,-i}(R_{i,m},h), \qquad (35)$$

donde, n=100 y

$$g_{n,-i}(R_{i,m},h) = (n-1)^{-1}\, h^{-1} \sum^{n} \varphi((R_{i,m}-R_{j,m})h^{-1}),$$

y calculamos la ventana h que maximiza la expresión (35).

3.2-Método de vecinos más cercanos.

Calculamos el estimador de g, según la expresión (30) para $g_n(r)$, con r variando entre 0.01 y 0.09, con diferencias de un centésimo. Los predictores $R_{i,m,h}$, i=1,...,100, los calculamos con la expresión suavizada (31) vista en este capítulo. Una vez obtenido el estimador, se calcula el error $MECI_1$.
Para el Caso 1, el valor de ventana h usado en la expresión suavizada fue el que se obtuvo al minimizar en el intervalo [0.01;1) con diferencia de un centésimo, la aproximación del MECI,

$$MECI_1(F_{0,n,h}) = (1/n) \sum_{i=1}^{n} (F_{0,n,h}(X_i)-F_0(X_i))^2.$$

Al comparar los resultados para distintas replicaciones, se decidio tomar h=0.36.
Para los otros dos casos se utilizó la ventana óptima h_{opt} que resultó ser igual a 0.0493 y 0.0446, para el Caso 2 y 3 respectivamente.
Observamos que este estimador $F_{0,n,h}$, de F_0, aún para ventanas alejadas de la elegida, da valores de error relativamente pequeños y similares, ya que el estimador no suavizado $F_{0,n}$ es un estimador consistente de F_0. Con los valores predichos, $R_{i,m,h}$, i=1,...,100, de la variable relativa R, pasamos a determinar el número de vecinos, según los algoritmos que detallamos a continuación.

Número de vecinos óptimo.

Calculamos el orden del vecino más cercano utilizando que $k_n/(2nH_{k,r})\rightarrow g(r)$ en probabilidad, donde $H_{k,r}$ es la distancia de r a su k-ésimo vecino más cercano. Luego si queremos que $H_{k,r}$ esté cerca de la ventana óptima h_{opt}, deberíamos elegir k de modo que $k=k_{opt}=[2nh_{opt}g(r)]$, donde [x] indica la parte entera de x. Llamamos a esta elección de número de vecinos *número de vecinos óptimo* porque está asociada a la abertura de ventana óptima. En todos los casos, tomamos g(r)=1 que corresponde al Caso 1 en que ambas poblaciones tienen la misma distribución.

La ventana $H_{k,r}$ en la expresión (30), será entonces la distancia del valor r a su vecino k_{opt} más cercano entre los $R_{i,m,h}$ calculados.

Si consideramos la observación de Scott ((1992), pag.165), acerca del sobresuavizado del estimador $g_n(r)$ y aplicamos el valor h_{os} en la aproximación de Moore y Jackel (1976), tendríamos un valor de vecino más cercano $k_{os} = 2nh_{os}$.

El siguiente cuadro resume los valores de ventana óptima y de número de vecinos asociado para cada situación.

Ventana	Caso 1	Caso 2	Caso 3
h_{opt}	0.1100	0.0493	0.0446
$k_{opt}=[2nh_{opt}]$	22	10	9
h_{os}	0.1314	0.0744	0.1120
$k_{os}=[2nh_{os}]$	26	15	22

Número de vecinos de validación cruzada de mínimos cuadrados.

Calculados los valores $R_{i,m,h}$, encontramos el número de vecinos y por lo tanto, la amplitud de ventana, que depende del orden del vecino más cercano, minimizando la expresión (24), que ahora tiene la forma,

$$VC(h,\ k) = \int_0^1 g_n^2(r)dr - 2*n^{-1}\sum_{i=1}^{n} g_{n,-i}(R_{i,m,h}), \tag{36}$$

donde,

$$g_n(r) = \sum_{i=1}^{n} K[(r-R_{i,m,h}))H_{k,r}^{-1}]/(nH_{k,r}),$$

y

$$g_{n,-i}(x) = (n-1)^{-1}\ H_{k,x}^{-1}\sum_{j\neq i}^{n}\varphi((x-R_{j,m,h})H_{k,x}^{-1}). \tag{37}$$

Esta expresión depende de la abertura de ventana para el estimador suavizado de la función de distribución y del número de vecinos. Como hemos mencionado anteriormente, la abertura de ventana h se eligió previamente como h=0.36, 0.0493 y 0.0446 para los Casos 1, 2 y 3 respectivamente. Por lo tanto, sólo resta buscar el número de vecinos k que minimiza (36).
El valor de la integral en (36) lo aproximamos por una suma sobre una grilla de valores de la función a integrar y buscamos el valor de $k = k_{VC}$, cuando k varía entre 4 y 96, que minimiza VC(h, k). De esta forma calculamos $g_n(r)$ utilizando como ventana $H_{k,i}$, la distancia de r al k_{vc}-ésimo vecino más cercanos entre los $R_{i,m,h}$.

Número de vecinos de validación cruzada de pseudo máxima verosimilitud.
A partir de la forma (27), consideramos la expresión

$$L(h,k) = \prod_{i=1}^{n} g_{n,-i}(R_{i,m,h}). \qquad (38)$$

Fijado h como hemos descripto, la forma de $g_{n,-i}(R_{i,m,h})$, dada por la expresión (37), depende del orden k del vecino más cercano, por lo tanto debemos encontrar el valor de $k = k_{MVP}$ que maximice L(h,k), dándole a k valores entre 4 y 96, para cada conjunto de valores $R_{i,m,h}$ calculado. Con ese valor k_{MVP} calculamos el estimador de vecino cercano, $g_n(r)$, donde $H_{k,i}$ es la distancia de r al k_{MVP} vecino más cercano entre los $R_{i,m,h}$.

4-Resultados de las simulaciones.

4.1-Caso 1.

El primer conjunto de datos que generamos corresponde a dos poblaciones independientes, ambas normal estándar, X_0 y X.
Según vimos al comienzo del Capítulo 4, la variable relativa R tendrá distribución uniforme en el intervalo (0,1).

Comportamiento de los valores de $MECI_1$.
El método de núcleos presentó mejor comportamiento que el de vecinos más cercanos cada vez que se usaba el mismo método de amplitud de ventana. Los percentiles del $MECI_1$ se presentan en la Tabla 1 para los distintos procedimientos descriptos. La Tabla 2

da medidas resumen para las 500 ventanas calculadas por cada procedimiento. En las Fig. 13 y 14 se grafica la densidad del $MECI_1$ para los estimadores basados en núcleos y en vecinos más cercanos, respectivamente.

En la Fig. 15 se grafican los valores de ventanas obtenidos y su respectivo $MECI_1$.

Observemos que el rango de valores de $MECI_1$ no difiere sustancialmente en Fig. 15 (a) y (b). Sin embargo un alto porcentaje de valores de $MECI_1$ se mantuvieron, al considerar ventanas menores o iguales que 0.13, menores que los valores de $MECI_1$ correspondientes a las mismas observaciones, con ventanas sin cota. Un gráfico comparativo se presenta en Fig. 11, mostrando que los valores del error, cuando consideramos ventanas acotadas, son en su mayoría menores que los obtenidos para ventanas sin acotar. Presentamos algunos gráficos de valores de h versus VCS(h), para observar que no existen mínimos locales, en la Fig. 10.

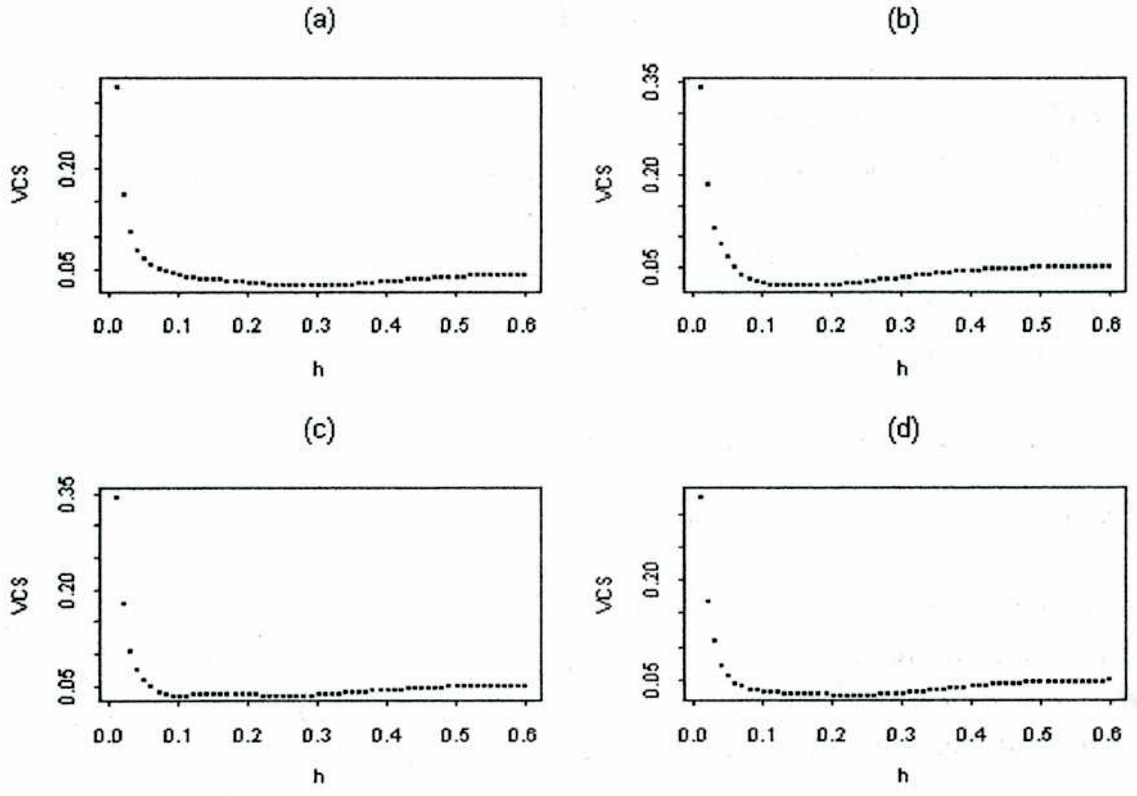

Fig. 10. $F=F_0 \sim N(0,1)$. Gráficos de h versus VCS(h) para cuatro replicaciones. El valor mínimo de VCS(h) se alcanzó en, (a) $h_{VCS}=0.28$, (b) $h_{VCS}=0.15$, (c) $h_{VCS}=0.25$, (d) $h_{VCS}=0.23$. Al acotar $h \leq h_{os}= 0.13$, el mínimo de VCS(h) se alcanzó en (a) h=0.13, (b) h=0.13, (c) h=0.10, (d)h=0.13.

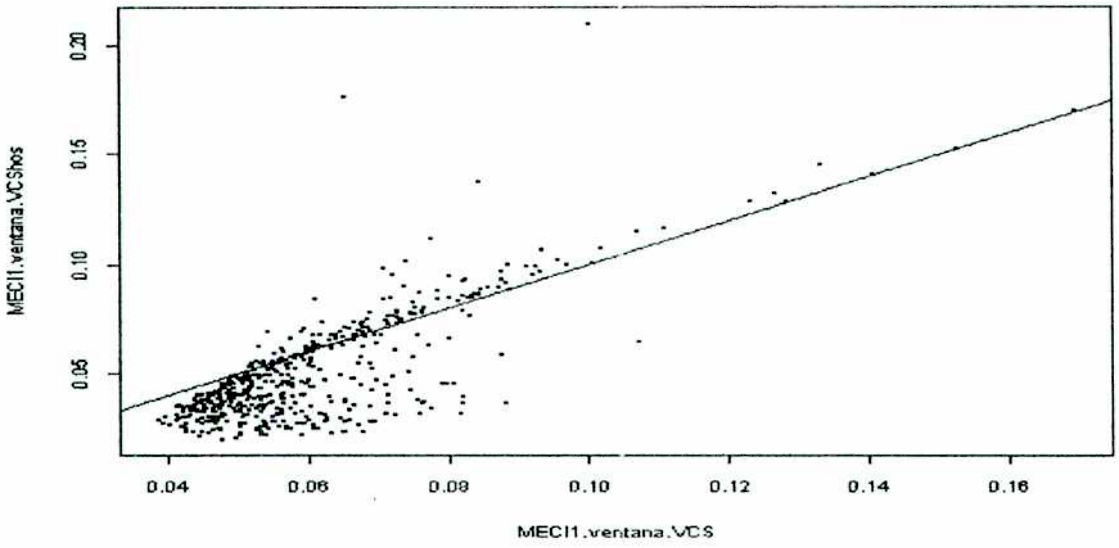

Fig. 11. Pares ($MECI_1(h_{VCS})$, $MECI_1(h_{VCS}<h_{os})$) de valores de $MECI_1$ con ventanas sin considerar la cota de h_{os} y considerándola. En un gran porcentaje, el valor de $MECI_1$ fue menor cuando se consideraron ventanas menor que h_{os}, según se deduce por la cantidad de puntos por debajo de la bisectriz.

Tabla 1- Valores descriptivos de $MECI_1$.

ESTIMADOR	VENTANA	Percentiles de $MECI_1$ Q_1	Q_2	Q_3	MAD
de Nucleo	Optima	0.03287	0.04527	0.06496	0.020947
	VCMCS $h \le h_{os}=0.13$	**0.03438**	**0.04433**	**0.06269**	**0.017815**
	VCMCS	0.04912	0.05593	0.06674	0.012271
	VCMVP	0.1839	0.2779	0.4358	0.1724155
de Vecino Cercano	Optima	0.04259	0.05886	0.08149	0.0279522
	VCMCS	0.05054	0.06632	0.09633	0.0309914
	VCMVP	*0.2374*	*0.3564*	*0.5025*	*0.190251*

En la Tabla 1, se marcaron en negrita los valores preferenciales del $MECI_1$ y en cursiva los menos favorables.

Para método de vecinos más cercanos, los valores de $MECI_1$ dieron muy similares cuando tomamos k_{opt}= 26 o k_{os}=22, al calcularlos en el 50% de las simulaciones.

Se obtuvieron mejores resultados del $MECI_1$, para método núcleo, cuando la ventana fue calculada por el procedimiento de mínimos cuadrados sesgado, cuando se verifica la acotación superior de h_{os} y por ventana fija u óptima. Lo podemos apreciar en la Fig.12 y de Fig.13. Para las mismas observaciones, se calculó el error de estimación $MECI_1$ cuando se aplicó ventana óptima o fija y la ventana que minimiza VCS(h) para valores menores que h_{os}, los pares correspondientes se representan en Fig. 12(a), donde se observa que las diferencias no son demasiado importantes. En la Fig. 12(b), representamos los pares de $MECI_1$ para la ventana de mínimos cuadrados sesgados (VCMC), con ventana acotada superiormente, y para la ventana de validación cruzada de máxima verosimilitud penalizada (VCMVP). En esta figura se puede apreciar la diferencia existente entre ambos valores de error, favorable al algoritmo de ventana de mínimos cuadrados.

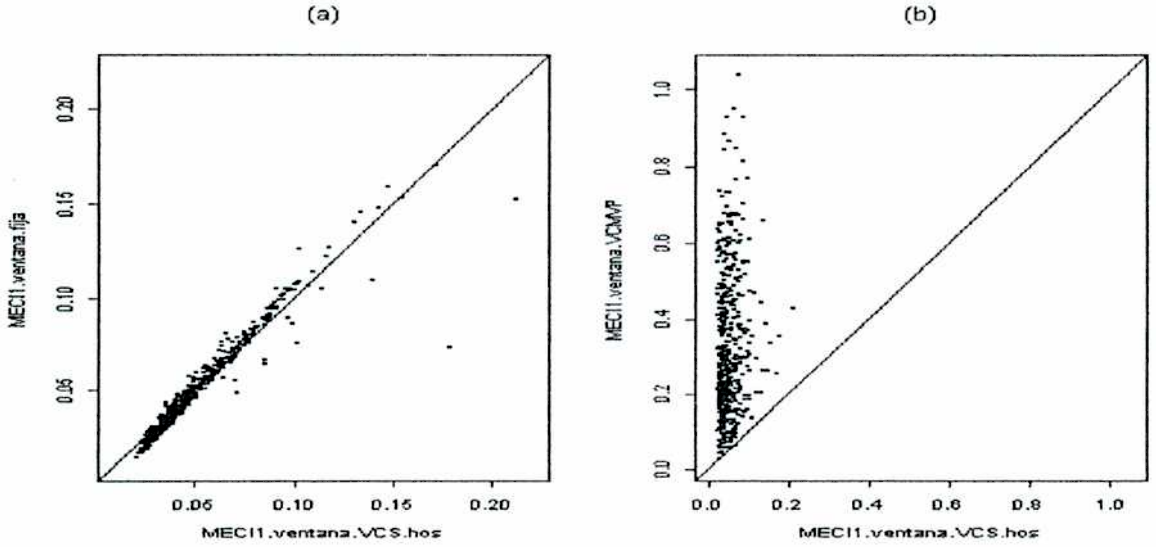

Fig.12. F_0=F~N(0,1).Método núcleos.(a) Pares de valores de $MECI_1$, la primera componente corresponde a ventana VCMCS, $h_{VCS} \leq h_{os}$=0.13, la segunda componente corresponde a ventana fija u óptima, h_{opt}=0.11 (b) pares de valores de $MECI_1$, la primera componente con ventana VCMCS, $h_{VCS} \leq h_{os}$=0.13, la segunda componente con ventana h_{MVP}. La recta que figura en ambos es la bisectriz.

La Fig. 13 muestra la distribución de los valores de $MECI_1$, para el métodos núcleos, para las tres ventanas elegidas. La ventana de máxima verosimilitud penalizada resulta ser la más desfavorable. Las dos ventanas restantes presentan distribuciones para el $MECI_1$ muy similares.

La Fig. 14, presenta gráficos de distribución del $MECI_1$, cuando se estimó la densidad relativa por método de vecinos más cercanos aplicando distintos algoritmos para encontrar la ventana $H_{k,r}$ de la expresión (30). El número de vecinos óptimo es el más conveniente para este método, siguiéndole el de mínimos cuadrados, con mayor dispersión. El número de vecinos de máxima verosimilitud penalizada tuvo, como en el método núcleo, valores de $MECI_1$ relativamente grandes. Al comparar con los gráficos de la Fig. 13 se observa la conveniencia del método núcleo para cada una de las ventanas.

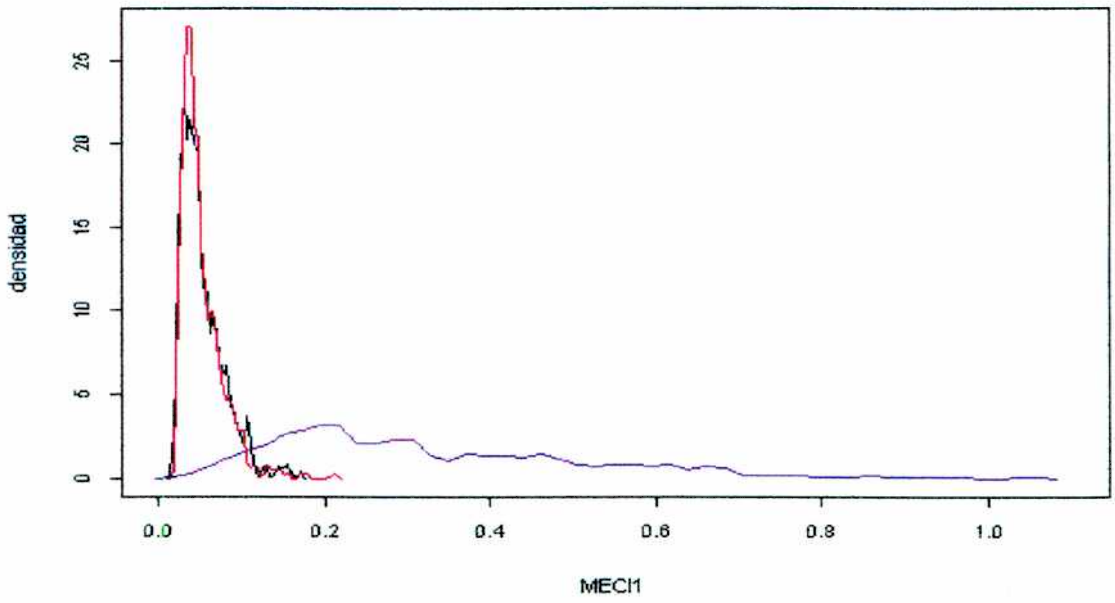

Fig. 13. Gráfico de densidades de $MECI_1$ para métodos núcleos, $F_0=F=N(0,1)$. La línea negra corresponde a estimador de núcleo con ventana fija, la línea roja con ventanas de validación cruzada de mínimos cuadrados sesgada (VCMCS, $h_{VCS}<h_{OS}$) y la azul con ventanas de validación cruzada de maxima verosimilitud penalizada (VCMVP).

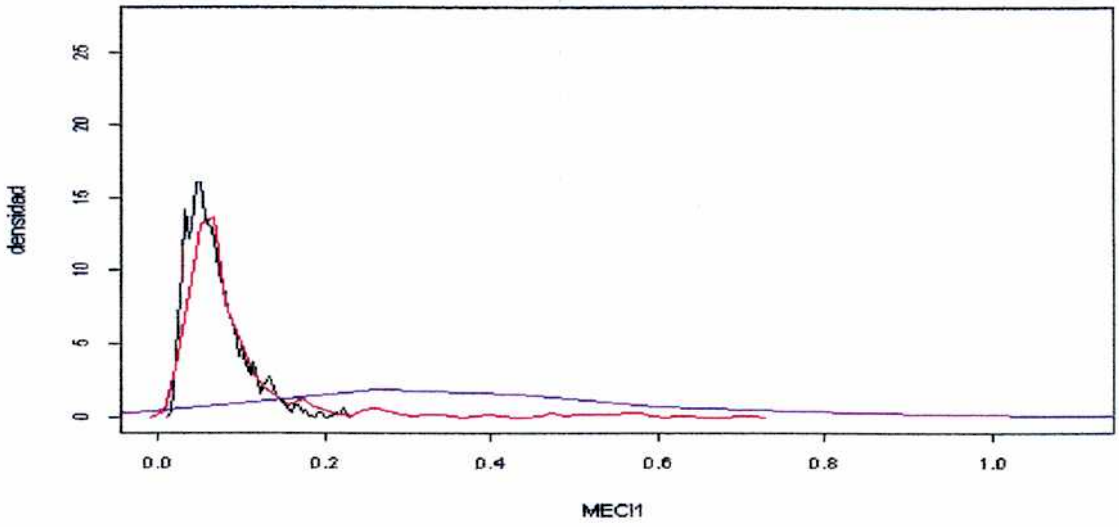

Fig. 14 Gráfico de densidades de $MECI_1$ para métodos de vecinos cercanos, $F_0=F\sim N(0,1)$. La línea negra corresponde a estimador con número de vecinos óptimo, la línea roja con número de vecinos de validación cruzada de mínimos cuadrados sesgada (VCMCS) y la azul con número de vecinos de validación cruzada de máxima verosimilitud penalizada (VCMVP).

Comportamiento de los valores de ventanas en el método de núcleos.

La siguiente tabla muestra como medidas resúmen de los valores de ventanas de método de núcleo, calculadas para cada una de las 500 simulaciones, según el procedimiento de ventana óptima, de VCMCS o de VCMVP, el cuartil inferior y superior Q_1 y Q_3, la mediana Q_2 y la MAD.

Tabla 2- Valores descriptivos de h_{VCS} y h_{MVP}.

Percentiles de ventanas

VENTANA	Q_1	Q_2	Q_3	MAD
Optima	0.11	0.11	0.11	0.
VCMCS $h_{VCS} \leq h_{os}$	0.13	0.13	0.13	0.
VCMCS	0.17	0.2	0.23	0.044478
VCMVP	0.01	0.02	0.02	0.014826

Los gráficos siguientes (Fig. 15) dan los pares formados por los valores de ventana h y el correspondiente $MECI_1(h)$, para los tres procedimientos que no son de ventana fija. Es posible observar que

los valores de ventanas de Fig.15(b) y Fig.15(c) tienen recorridos que no se superponen.

En el caso de VCMCS con valores de ventanas sin acotar superiormente (Fig. 15(b)) con h_{os}, se obtuvieron ventanas mayores o iguales que la ventana óptima o fija h_{opt} =0.11, mientras que para VCMVP las ventanas han dado valores mucho menores que el valor 0.11. Se puede observar también, el rango de valores de MEC_1 en los tres conjuntos dados. La ventana de validación cruzada sesgada presenta, en sus dos formas, valores menores de error. Con el método de máxima verosimilitud penalizada, las ventanas son muy pequeñas y los valores de error relativamente grandes. En este caso la ventana fija h_{opt}=0.11 y los valores de ventana $h_{VCS} \leq h_{os}$ son muy similares, ocurriendo lo mismo con los valores de error.

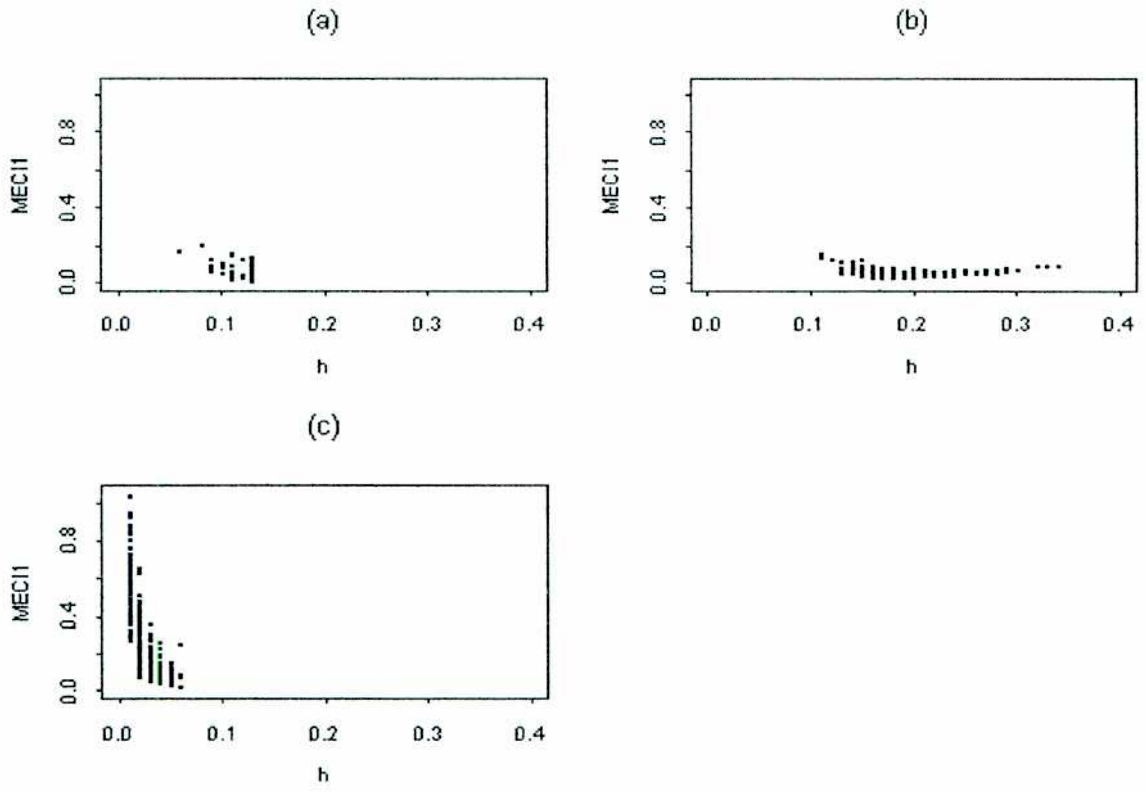

Fig. 15. Gráficos de pares (h, $MECI_1(h)$), método núcleo, $F=F_0 \sim N(0,1)$. (a) Ventanas $h = h_{VCS} \leq h_{os}=0.13$, calculadas con el algoritmo de validación cruzada de mínimos cuadrados sesgado, . (b) Ventanas $h = h_{VCS}$, sin cota, calculadas con el algoritmo de validación cruzada de mínimos cuadrados sesgado,(c) Ventanas $h=h_{MVP}$, calculadas con el algoritmo de validación cruzada de máxima verosimilitud penalizada.

Conclusión: Tanto para el método de núcleos como para el de vecinos más cercanos con núcleos, el algoritmo de ventana más

desfavorable fue el de máxima verosimilitud penalizada, alcanzando un valor máximo del $MECI_1$ de 1.044 y 7.729, respectivamente.
Entre los estimadores de núcleos considerados, el estimador con ventana fija tuvo un comportamiento muy similar, respecto a la medida de error considerada, al de ventana de VCMCS con cota, presentando ambos los menores valores de $MECI_1$.
El método núcleo se comportó mejor, para cada ventana que el método de vecinos cercanos asociado.

4.2-Caso 2.

Se consideró a la variable de referencia X_0, con distribución normal de media 1 y desvío estándar 2, mientras que la variable de comparación X, se eligió como normal estándar. El gráfico de g(r) se mostró en el Capítulo 4, Fig.7, siendo una densidad unimodal y asimétrica.

Comportamiento de los valores de $MECI_1$.
La tabla siguiente muestra el comportamiento de los valores de $MECI_1$.
Tabla 3-Valores descriptivos de $MECI_1$.

		Percentiles de $MECI_1$			MAD
ESTIMADOR	VENTANA	Q_1	Q_2	Q_3	
de Núcleo	Optima	0.08061	0.1321	0.206	0.0876174
	VCMCS	0.0526	0.09722	0.162	0.0741579
	VCMCS $h_{VCS} \leq h_{os}$	**0.04968**	**0.09071**	**0.1527**	**0.0699**
	VCMVP	0.1569	0.3147	0.5908	0.2852692
de Vecino Cercano	Optima	0.3561	0.6058	1.401	0.4616495
	V.C.M.C.	0.295	0.5161	0.8218	0.3842716
	VCMVP	*2.143*	*10.34*	*178*	*14.19743*

En la Tabla 3, los valores marcados en negrita fueron los más favorables y en itálica los menos favorables a la estimación.

La aplicación de método de núcleos dio resultados razonables. El mejor estimador dentro de esta clase corresponde a la selección de ventana de VCMCS, en primer lugar considerando $h \leq h_{os}$ en la expresión VCS(h) (26) y en segundo lugar sin acotar los posibles valores de ventanas. En la Fig.17 se puede visualizar que las diferencias son escasas, entre los valores de $MECI_1$, para estas dos formas de darle valores a h. Los puntos formados por pares de valores de $MECI_1$ calculados sobre las mismas observaciones, se sitúan muy cercanos a la bisectriz, lo que indica que en este caso son valores muy similares en sus componentes. La Fig.16 muestra los valores de ventanas h versus VCS(h), para cuatro conjuntos de de replicaciones, observándose que no existen mínimos locales.

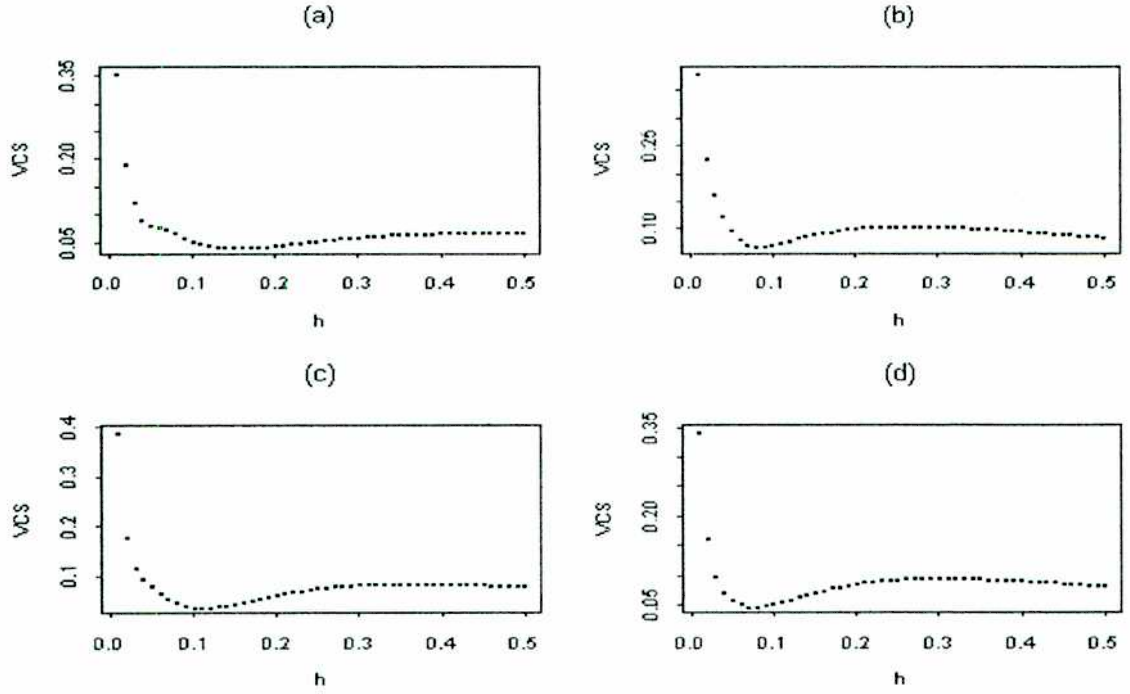

Fig. 16. $F \sim N(0,1), F_0 \sim N(1,2^2)$. Gráficos de h versus VCS(h) para cuatro replicaciones. El valor mínimo de VCS(h) se alcanzó en, (a) $h_{VCS}=0.15$, (b) $h_{VCS}=0.08$, (c) $h_{VCS}=0.11$, (d) $h_{VCS}=0.08$. Al acotar $h \leq 0.08 \cong h_{os}$, el valor mínimo se alcanzó, en los cuatro conjuntos de valores, en h=0.08.

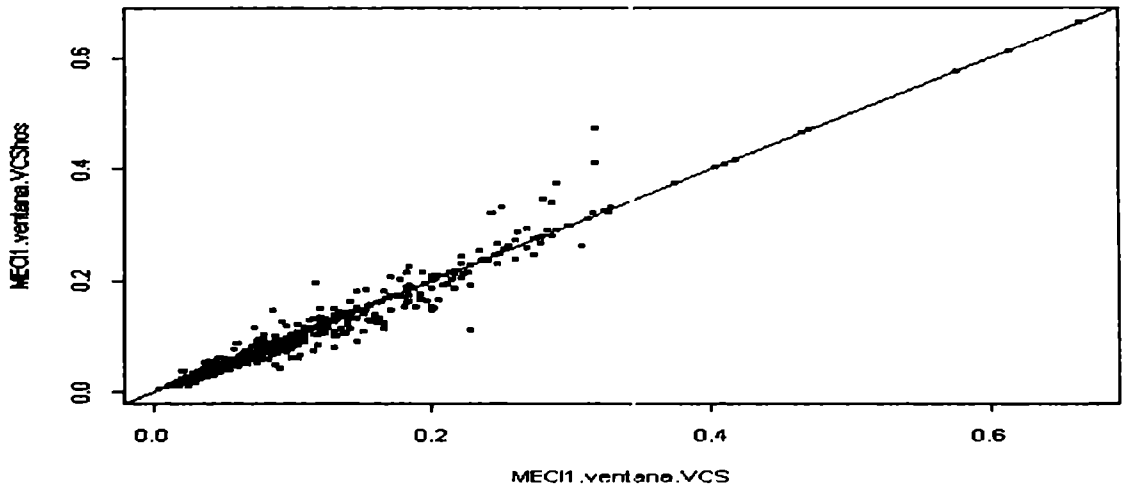

Fig.17. $F \sim N(0,1), F_0 \sim N(1,2^2)$. Pares $(MECI_1(h_{VCS}), MECI_1(h_{VCS}<h_{os}))$ de valores de $MECI_1$ con ventanas sin considerar la cota de $h_{os}=0.0744$ y considerándola, en ese orden. En su mayoría los puntos se encuentran muy cercanos a la bisectriz.

Un estimador de la densidad de los valores del $MECI_1$, está representada en la Fig. 19 para los tres tipos de ventanas, tomando como ventana de VCMCS la que corresponde al procedimiento de acotar superiormente los valores de ventanas. Se observa que, para ventanas de mínimos cuadrados sesgados, se presentan con mayor densidad los valores menores de $MECI_1$. El máximo y el mínimo de los valores de error $MECI_1$ para ventana fija, son casi coincidentes con los mismos valores, para ventanas de mínimos cuadrados. La ventana de máxima verosimilitud penalizada fue la que dio los valores mayores de error.

La Fig.18 muestra el comportamiento de pares de valores de $MECI_1$, calculados sobre el mismo conjunto de cantidades de $R_{i,m}$, a fin de comparar la estimación con ventana VCMCS, con las dos restantes ventanas, en método núcleo. Se observa en Fig.18(b) la mayor diferencia con ventana de VCMVP. En ambos gráficos se visualizan los valores máximos y mínimos de los tres conjuntos de valores de $MECI_1$.

El método de vecinos más cercanos presentó valores de $MECI_1$, muy desfavorables para las tres elecciones de número de vecinos propuestas. En el caso de número de vecinos fijo, los valores mayores que el tercer percentil aumentan hasta alcanzar un máximo igual a $2844*10^5$. En la Fig.20 se muestra la distribución de

aquellos menores que 10, los cuales conforman el 94% de los valores obtenidos. Un 80% conservó valores menores que 2, entre 2 y 10 un 14% y mayores que 10, creciendo hasta el máximo, un 6%.
Para número de vecinos de VCMCS sólo hubo un valor de $MECI_1$ sumamente desfavorable (=17680), este valor correspondió a una simulación que aún en método de núcleo dio valores de $MECI_1$ mayores que el tercer percentil, para todos los tipos de ventana. El resto de los valores se mantuvo menor que 3.78. En la Fig. 20 se grafica la densidad de estos valores, sin incluir el valor extremo mencionado. No se hubiera podido realizar un gráfico comparativo incluyendo la totalidad de los valores de $MECI_1$.
El número de vecinos de VCMVP presenta un valor muy grande para el tercer cuartil, dando un valor para la MAD superior a todos los calculados hasta este momento e imposibilitando un gráfico que permita comparar el comportamiento en la distribución del $MECI_1$ con las otras estimaciones.

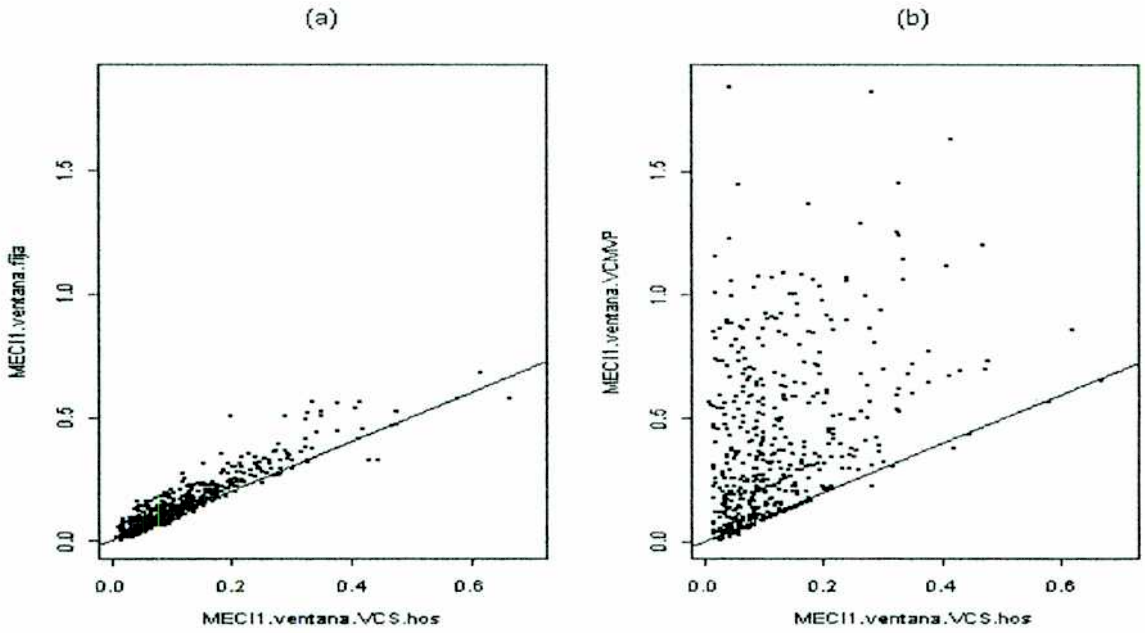

Fig. 18. Método núcleo. $F_0 \sim N(1,2^2)$, $F \sim N(0,1)$. (a)Pares de valores de $MECI_1$, la primera componente corresponde a ventanas de mínimos cuadrados sesgados $h_{VCS} \leq h_{os}$, la segunda componente corresponde a ventana fija u óptima $h_{opt}=0.049$, (b)pares de valores de $MECI_1$, la primera componente con ventana $h_{VCS} \leq h_{os}$, la segunda componente con ventana h_{MVP}. La recta que figura en ambos es la bisectriz.

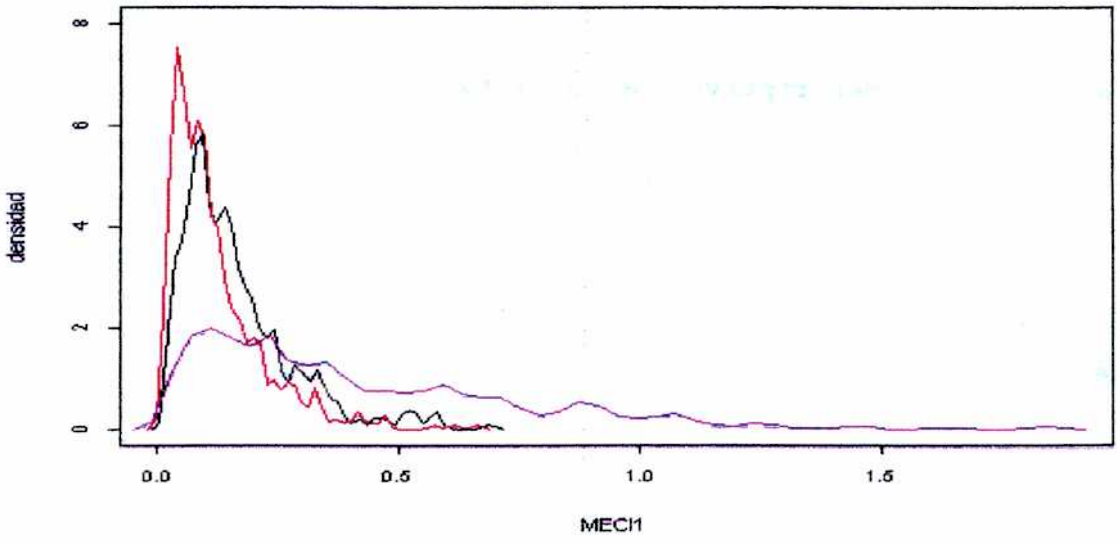

Fig. 19. Gráfico de densidades de $MECI_1$ para métodos núcleos, $F_0 \sim N(1,2^2)$, $F \sim N(0,1)$. La línea negra corresponde a estimador de núcleo con ventana fija, la roja con ventanas VCMCS y la azul con ventanas de VCMVP.

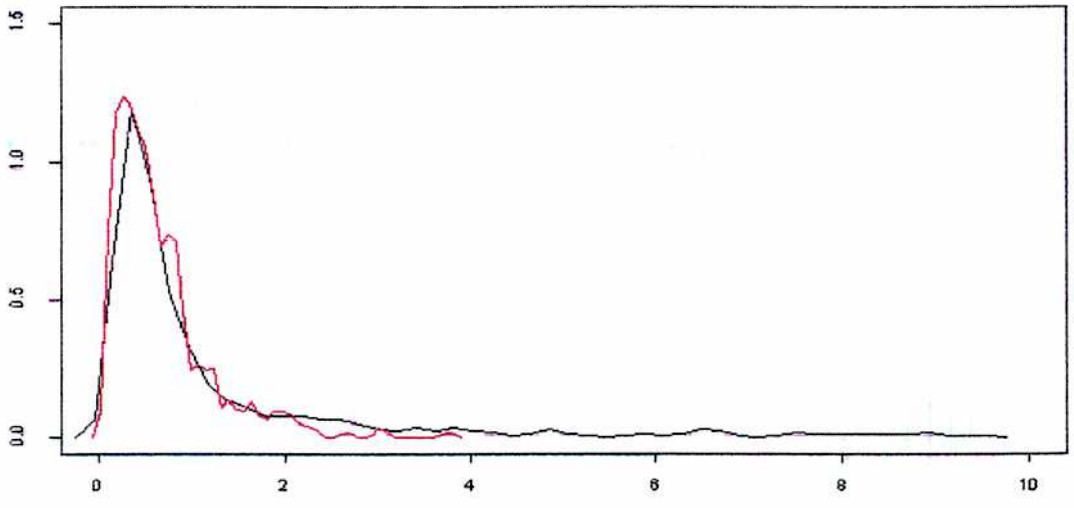

Fig. 20. Gráfico de densidades de $MECI_1$ para métodos de vecinos más cercanos, $F_0 \sim N(1,2^2)$, $F \sim N(0,1)$. La línea negra corresponde a estimador con número de vecinos óptimo, la línea roja con número de vecinos de VCMC. Para el número de vecinos óptimo se consideró solamente los valores de $MECI_1$ menores que 10, para vecinos de VCMC se omitió el valor 17680.

Comportamiento de los valores de ventanas en el método de núcleos.
La tabla siguiente muestra las características de las ventanas para los tres procedimientos de obtención automática así como la ventana óptima.

Tabla 4. Valores descriptivos de h_{VCS} y h_{MVP}.

Percentiles de ventanas

VENTANA	Q_1	Q_2	Q_3	MAD
Optima	0.04927	0.04927	0.04927	0.
VCMCS, $h_{VCS} \leq h_{os}$	0.08	0.08	0.08	0.
VCMCS	0.08	0.09	0.11	0.0148
VCMVP	0.02	0.03	0.04	0.0148

Los percentiles con valores más altos se presentan para el procedimiento de VCMCS que fue el método que obtuvo mejor comportamiento respecto de los valores de $MECI_1$.

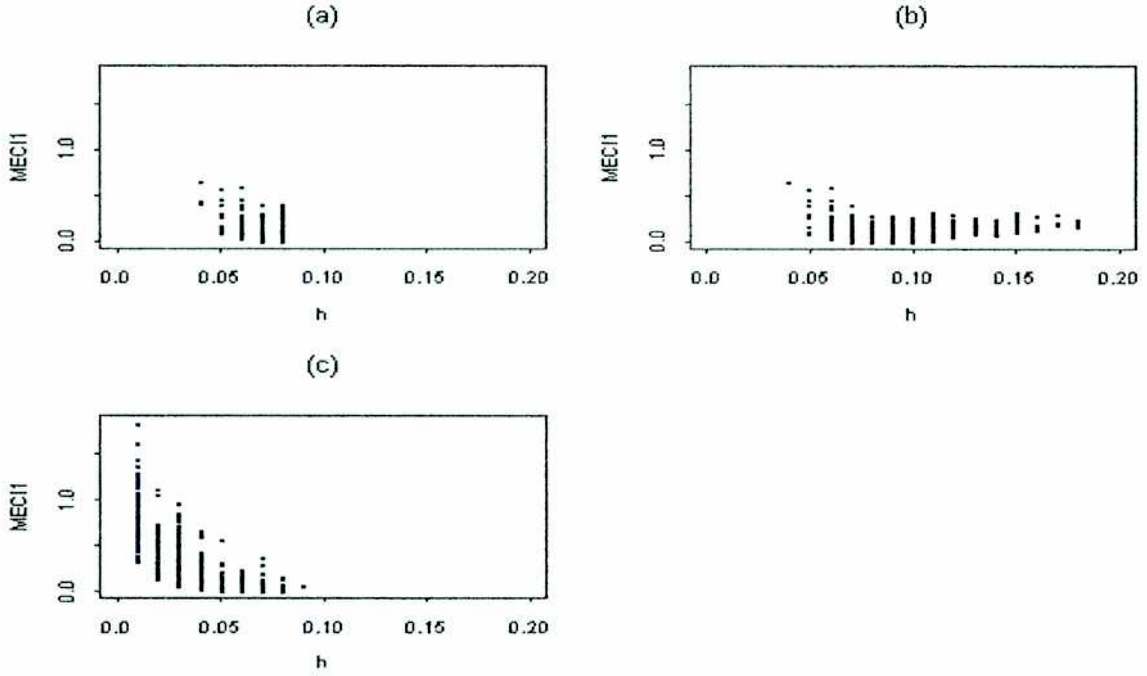

Fig. 21. Gráficos de pares (h, $MECI_1(h)$), $F \sim N(0,1), F_0 \sim N(1,2^2)$. (a) Ventanas $h = h_{VCS} \leq h_{os} = 0.0744$, calculadas con el algoritmo de validación cruzada de mínimos cuadrados sesgado. (b) Ventanas $h = h_{VCS}$, calculadas con el algoritmo de validación cruzada de mínimos cuadrados sesgado, sin restringir a $h \leq h_{os}$. (c) Ventanas $h = h_{MVP}$, calculadas con el algoritmo de validación cruzada de máxima verosimilitud penalizada.

La ventana fija h_{opt}=0.049 fue un valor posible en los conjuntos de ventanas de validación cruzada y aparentemente, en base a los resultados de la Tabla 3, resultó un valor demasiado pequeño. La Fig. 21 nos permiten comparar las ventanas y sus correspondientes valores de $MECI_1$.

Conclusión

El método de vecinos más cercanos presentó valores desfavorables de $MECI_1$, excepto en el caso de número de vecinos elegido por VCMCS que presentó un solo valor extremo. Los otros dos procedimientos de selección del número de vecinos han invalidado una posible estimación de g con este método.

El método núcleo con ventana de VCMCS, con cota h_{os} o sin cota en segundo lugar, sería la mejor elección en este caso, le sigue la ventana óptima o fija, con valores similares en medidas de error.

4.3-Caso 3.

En este caso la variable X_0 de referencia tiene distribución t-Student con dos grados de libertad y la variable de comparación X tiene distribución normal estándar. La variable relativa $R=F_0(X)$ tiene densidad g, bimodal, cuyo gráfico está dado en la Fig.8.

Comportamiento de los valores de $MECI_1$.

En este caso, como en los anteriores, el método núcleo mejora los resultados del método de vecinos más cercanos, para cada ventana considerada. Esto se aprecia en la Tabla 5, en la que los valores marcados en negrita fueron los más favorables y en itálica los menos favorables, a la estimación.

Para los estimadores basados en núcleos, la ventana más favorable para la estimación de g, fue la del algoritmo de VCMCS, con cota superior para sus ventanas, en segundo lugar el mismo algoritmo sin acotar los valores de h, siguiéndole la ventana óptima.

Como en los casos anteriores, para la elección de ventana de mínimos cuadrados sesgados, primero tomamos ventanas que variaran entre 0.01 y 0.99 y como un segundo procedimiento tuvimos en cuenta la propuesta de Scott((1992), pag. 166) de acotar superiormente los posibles valores de ventana con h_{os}=0.11. Con ventanas restringidas a ser menores que 0.11, obtuvimos mejores

valores de error en un gran porcentaje (Fig.23). En este caso, más que en los dos anteriores, se marca la diferencia a favor de ventanas acotadas superiormente por h_{os}. Cuando no se acotaron los valores de ventanas, las ventanas que minimizaban VCS(h) variaron entre 0.09 y 0.29. El gráfico de la Fig.22 muestra la relación entre h y VCS(h), observándose que el mínimo de VCS(h) se alcanza cuando h es mayor que 0.11.

Tabla 5-Valores descriptivos de $MECI_1$.

		Percentiles de $MECI_1$			MAD
ESTIMADOR	VENTANA	Q_1	Q_2	Q_3	
de Núcleo	Optima	0.08627	0.1196	0.164	0.057349
	VCMCS $h \leq h_{os}$	**0.05049**	**0.06641**	**0.08919**	**0.027829**
	VCMCS	0.06983	0.08294	0.09909	0.0216668
	VCMVP	0.2174	0.3421	0.551	0.2297943
de Vecino Cercano	Optima	0.1931	0.2904	0.4389	0.1729478
	V.C.M.C.	0.1137	0.2408	0.4217	0.2129378
	VCMVP	*0.909*	*2.442*	*22.42*	*NA*

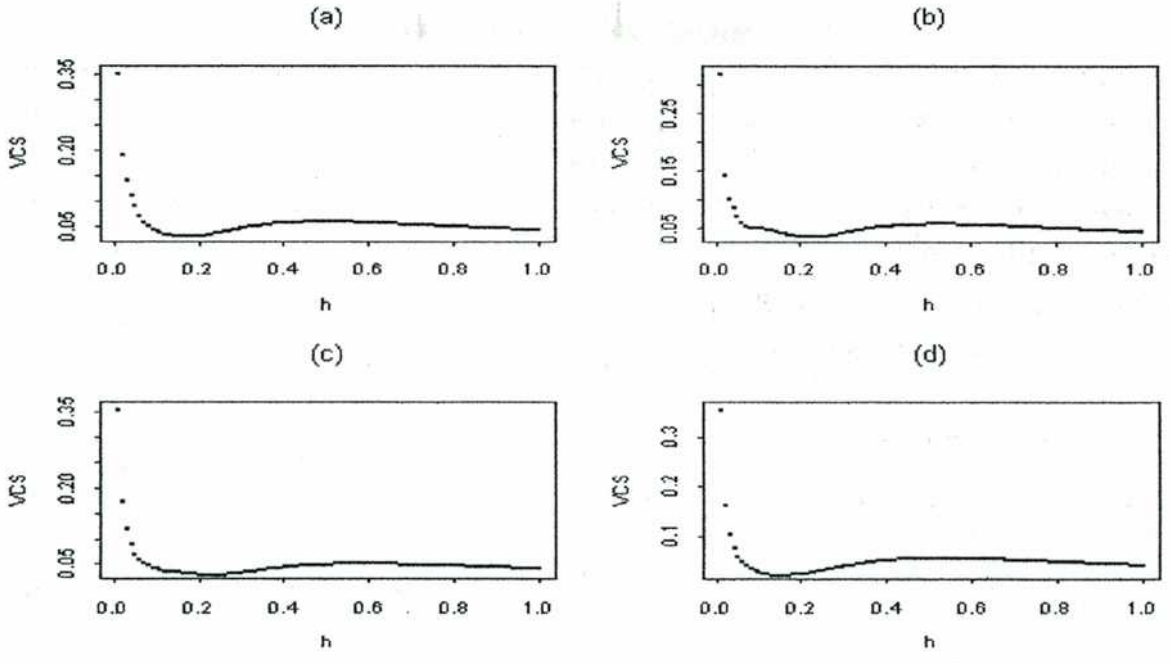

Fig. 22. F~N(0,1), F_0~t_2. Gráficos de h versus VCS(h) para cuatro replicaciones. El valor mínimo de VCS(h) se alcanzó en, (a) h_{VCS}=0.17, (b) h_{VCS}=0.22, (c) h_{VCS}=0.23, (d) h_{VCS}=0.15. Al acotar $h \leq 0.11 = h_{os}$, el valor mínimo se alcanzó, en los cuatro conjuntos de valores, en h=0.11.

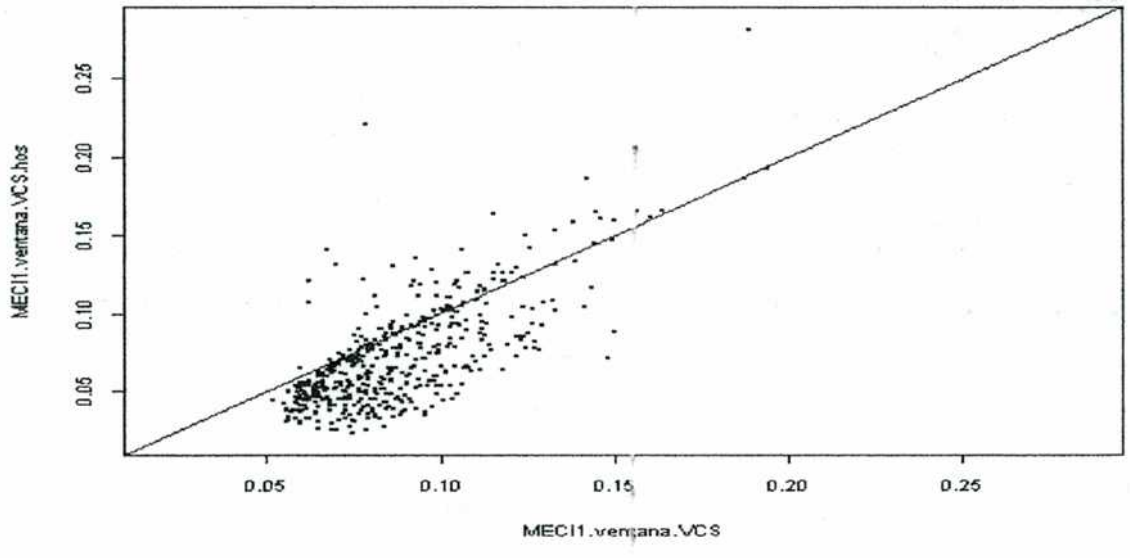

Fig. 23. F~N(0,1), F_0~t_2. Pares ($MECI_1(h_{VCS})$,$MECI_1(h_{VCS}<h_{os})$) de valores de $MECI_1$ con ventanas sin considerar la cota superior de h_{os}=0.11 y considerándola, para cada componente. En un 75% de los casos se verificó que $MECI_1(h_{VCS}<h_{os})<MECI_1(h_{VCS})$.

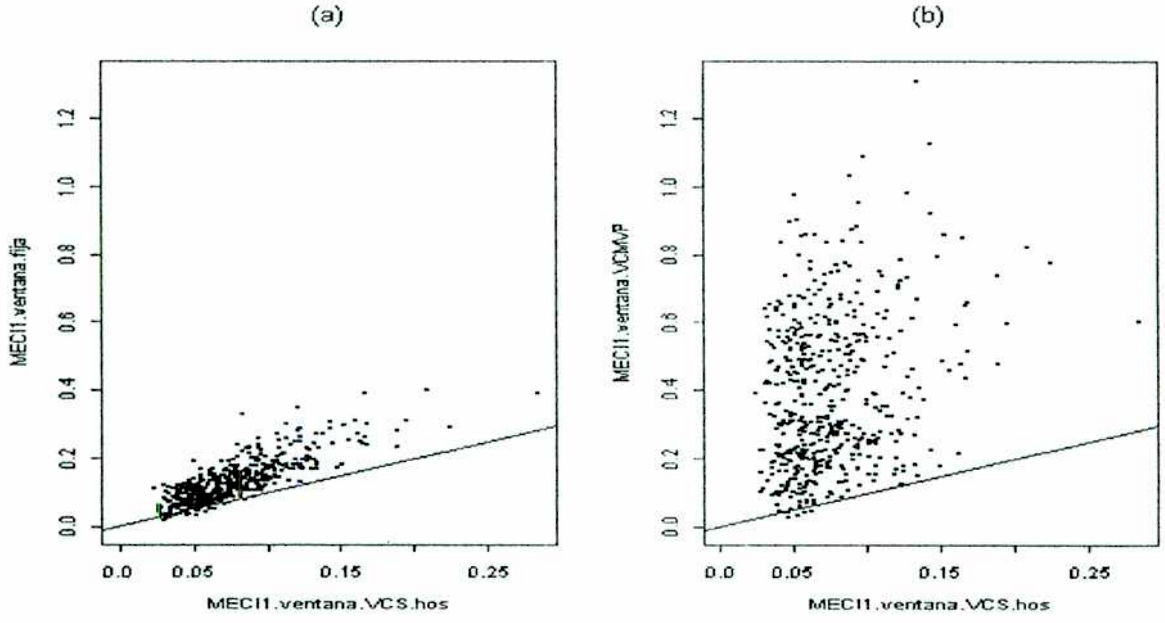

Fig. 24. F~N(0,1), F_0~t_2. Método núcleo.(a) Pares de valores de $MECI_1$, la primera componente corresponde a ventanas VCMCS con $h_{VCS} \leq h_{os}$, la segunda componente corresponde a ventana fija u óptima h_{opt}, (b)pares de valores de $MECI_1$, la primera componente con ventana $h_{VCS} \leq h_{os}$, la segunda componente con ventana h_{MVP}. La recta que figura en ambos es la bisectriz.

La Fig.24 (a) y (b), muestra los pares de $MECI_1$ obtenidos con ventana de mínimos cuadrados sesgados versus las dos restantes.
Como en las dos poblaciones anteriores, la ventana de máxima verosimilitud penalizada, genera valores mayores de ECM_1. En los dos gráficos de las Fig. 24 y Fig. 25, se pueden visualizar los mínimos y máximos de los tres conjuntos de $MECI_1$ obtenidos.
La Fig. 25 muestra un estimador de la densidad de valores de $MECI_1$, cuando estimamos g por el método de núcleos para cada una de las tres ventanas consideradas. Podemos visualizar la mayor densidad de $MECI_1$, en valores menores para la ventana de VCMCS, $h_{VCS} \leq h_{os}$, y los valores relativamente grandes que alcanza con ventanas VCMVP, con un máximo igual a 1.317.
Respecto al método de vecinos más cercanos, cuando calculamos el $MECI_1$ del estimador con número de vecinos óptimo, encontramos un valor extremadamente grande, igual a 465.7, y un 2.8% de valores mayores que 2.234. No pudiendo realizar un gráfico comparativo con ese rango, en la Fig. 26 representamos la densidad de la distribución de aquellos valores de $MECI_1$ menores o iguales a 2.234. Elegimos esa cota por ser éste el máximo valor del $MECI_1$,

cuando el número de vecinos es elegido por VCMCS. Este último procedimiento fue el que presentó los mejores resultados de estimación entre los procedimientos de vecinos más cercanos con núcleo. Respecto al número de vecinos de VCMVP, ya el tercer cuartil fue muy grande, 22.42, alcanzando un máximo de $1.3 * 10^{23}$.

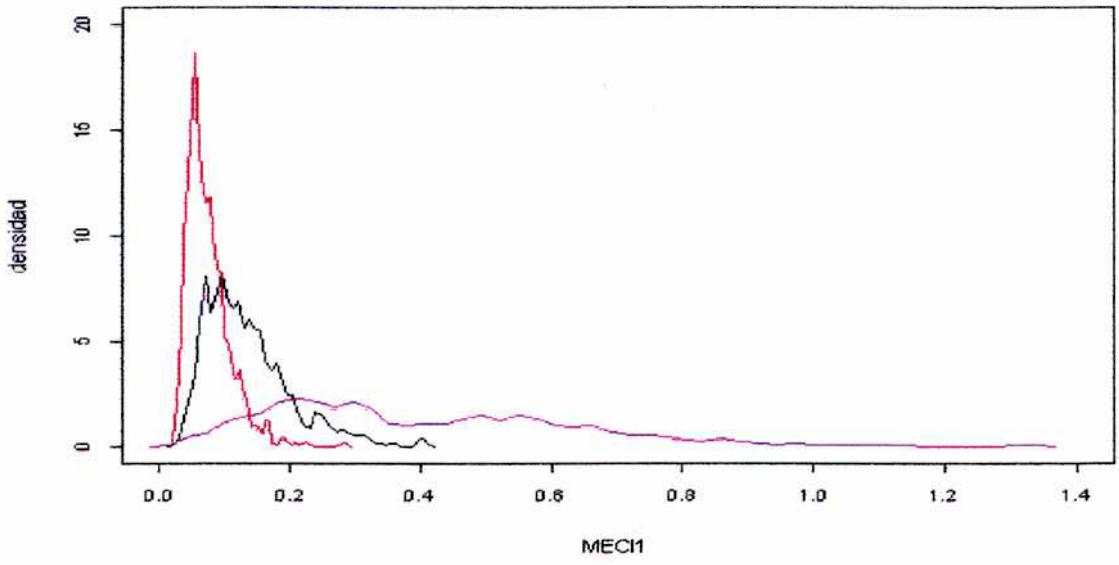

Fig.25. Gráfico de densidades de $MECI_1$ para el método de núcleos, $F_0 \sim t_2$, $F \sim N(0,1)$. La línea negra corresponde a estimador de núcleo con ventana fija, la línea roja con ventanas de validación cruzada de mínimos cuadrados sesgada (VCMCS, $h_{VCS} \leq h_{os}$) y la azul con ventanas de validación cruzada de maxima verosimilitud penalizada (VCMVP).

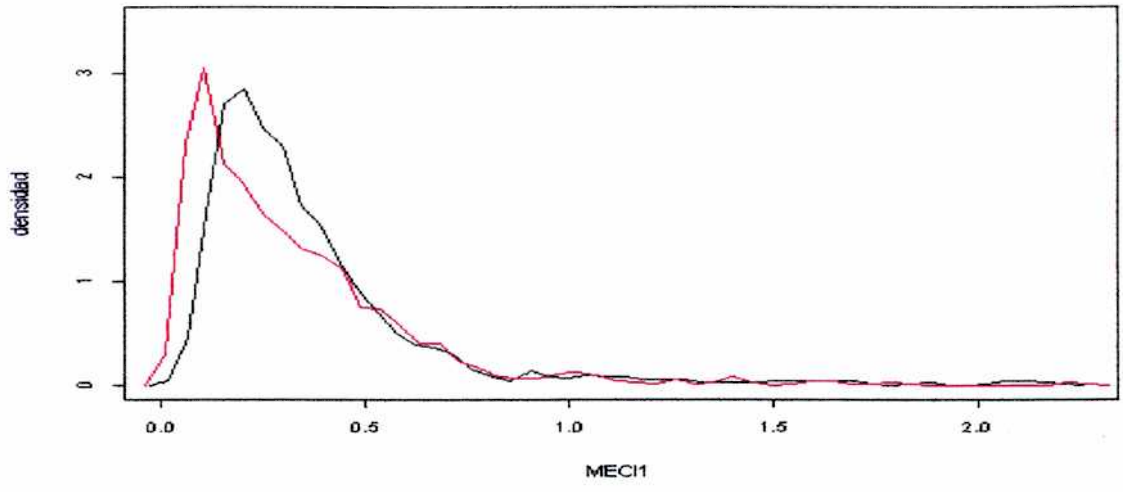

Fig.26. Gráfico de densidades de $MECI_1$ para el método de vecinos más cercanos, $F_0 \sim t_2$, $F \sim N(0,1)$. La línea negra corresponde al estimador con número de vecinos óptimo, la línea roja al obtenido con el número de vecinos de validación cruzada de mínimos cuadrados (VCMC). Para el primero se consideraron solamente los valores de MEC_1 menores que 2.25.

Comportamiento de los valores de ventanas en el método de núcleos.

La tabla siguiente muestra medidas resúmen para las ventanas para los tres procedimientos de cálculo de la ventana para el método de núcleos así como la ventana óptima.

Tabla 6. Valores descriptivos de h_{VCS} y h_{MVP}.

Percentiles de ventanas

VENTANA	Q_1	Q_2	Q_3	MAD
Optima	0.04458	0.04458	0.04458	0.
VCMCS $h \leq h_{os}=0.11$	0.11	0.11	0.11	0.
VCMCS	0.15	0.17	0.21	0.044478
VCMVP	0.01	0.02	0.02	0.014826

Los valores mayores de ventanas se presentan en el procedimiento de VCMCS, $h_{VCS} \leq h_{os}$, que generó el mejor comportamiento respecto del $MECI_1$. Según los gráficos de Fig. 27(a) y (b), la ventana $h_{os}=0.11$ es un posible valor del conjunto de ventanas h_{VCS}, mientras que el valor de ventana óptima, $h_{opt}=0.04458$, está dentro del rango de valores de ventanas h_{MVP}. Los valores mayores de $MECI_1$ se alcanzan con la ventana h_{MVP}.

Conclusión.

Como en los casos anteriores, el método de núcleos dio una mejor estimación de g, que el de vecinos cercanos, siendo la ventana de mínimos cuadrados sesgada la más conveniente, aún sin acotar los valores de ventanas por el valor de h_{os}. La ventana de máxima verosimilitud penalizada motivó valores muy grandes de $MECI_1$, tanto para método núcleo como para vecinos más cercanos.

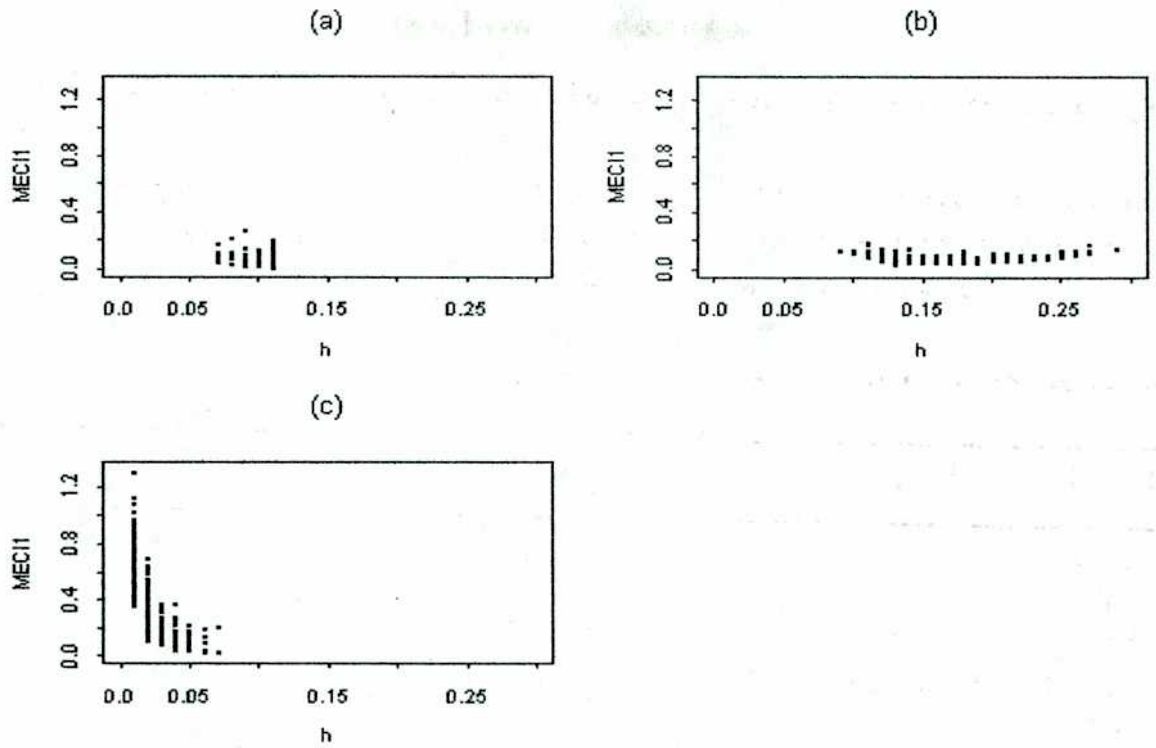

Fig. 27. Gráficos de pares (h, $MECI_1(h)$), $F \sim N(0,1)$, $F_0 \sim t_2$. (a) Ventanas $h = h_{VCS} \leq h_{os} = 0.11$, calculadas con el algoritmo de VCMCS. (b) Ventanas $h = h_{VCS}$, calculadas con el algoritmo de VCMCS, sin restringir a $h \leq h_{os}$. (c) Ventanas $h = h_{MVP}$, calculadas con el algoritmo de VCMVP.

4.4-Conclusiones.

Para los tres casos de pares de poblaciones simuladas, el método basado en núcleos se presentó como más conveniente que el método de vecinos más cercanos, para cada uno de los procedimientos de selección del parámetro de suavizado.

El parámetro de suavizado con peor comportamiento fue el de VCMVP en todos los casos y en cada uno de los dos métodos considerados en este trabajo. La ventana de VCMCS, para el método de núcleos, fue la más conveniente.

Para finalizar, desechado el método de vecinos más cercanos, presentamos los tres gráficos siguientes de boxplot comparativos para el método de núcleos, correspondientes a las tres variables relativas consideradas. En cada uno de ellos los boxplots representan los valores de $MECI_1$, para las cuatro ventanas: h_{opt}, $h_{VCS} \leq h_{os}$, h_{VCS} y h_{MVP}.

En las Fig. 28-30 (a) se presentan los cuatro boxplot comparativos. Debido a los valores grandes que asume el $MECI_1$ con

las ventanas de máxima verosimilitud, los restantes boxplot quedan desdibujados. En las Fig. 28-30 (b) los boxplots permiten comparar las ventanas fijas y de mínimos cuadrados sesgados, en un rango de valores más adecuado. Se pueden observar los valores obtenidos en cada caso de variable relativa. En el primer caso, en que R es una variable uniforme, los valores del conjunto se presentan acotados por 0.22, en el segundo caso (Fig. 29 (b)) en que R tiene una distribución sesgada los valores de $MECI_1$ son inferiores a 0.69 y en el tercer caso (Fig.30 (b)) en que R tiene una distribución bimodal y simétrica, los valores de $MECI_1$ están acotados superiormente por el valor 0.41.

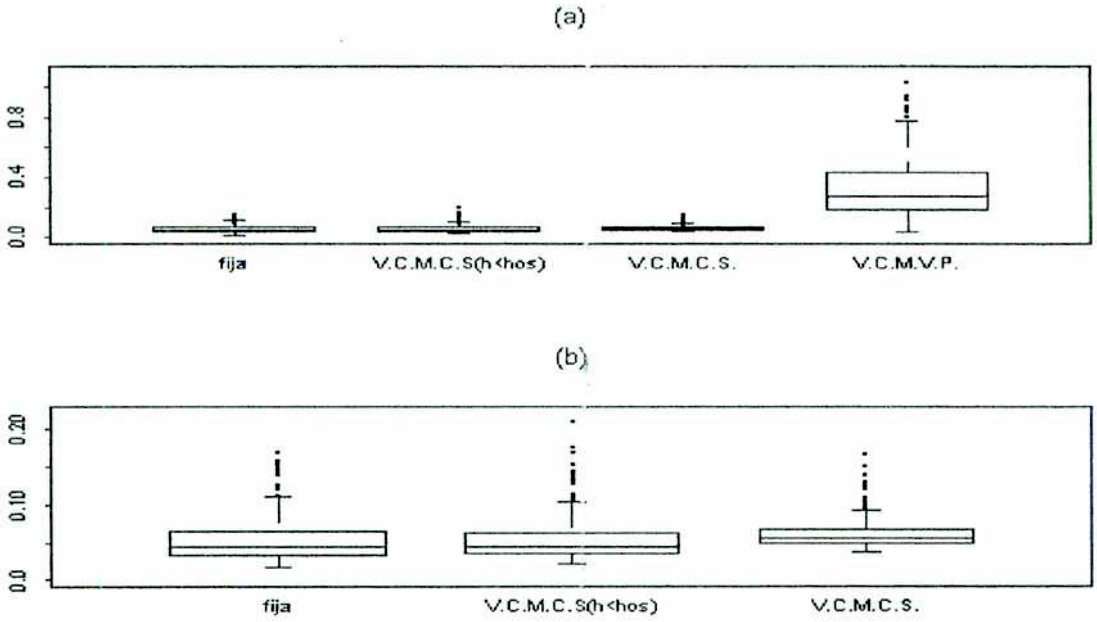

Fig.28. $F=F_0 \sim N(0,1)$. (a) Boxplot de los valores de $MECI_1$ de estimador de método núcleo, con ventanas h_{opt}, $h_{VCS}<h_{os}$, h_{VCS} sin cota y h_{VMP}. (b) Boxplot de los valores de $MECI_1$ de estimador de método núcleo, con ventanas h_{opt}, $h_{VCS}<h_{os}$ y h_{VCS} sin cota.

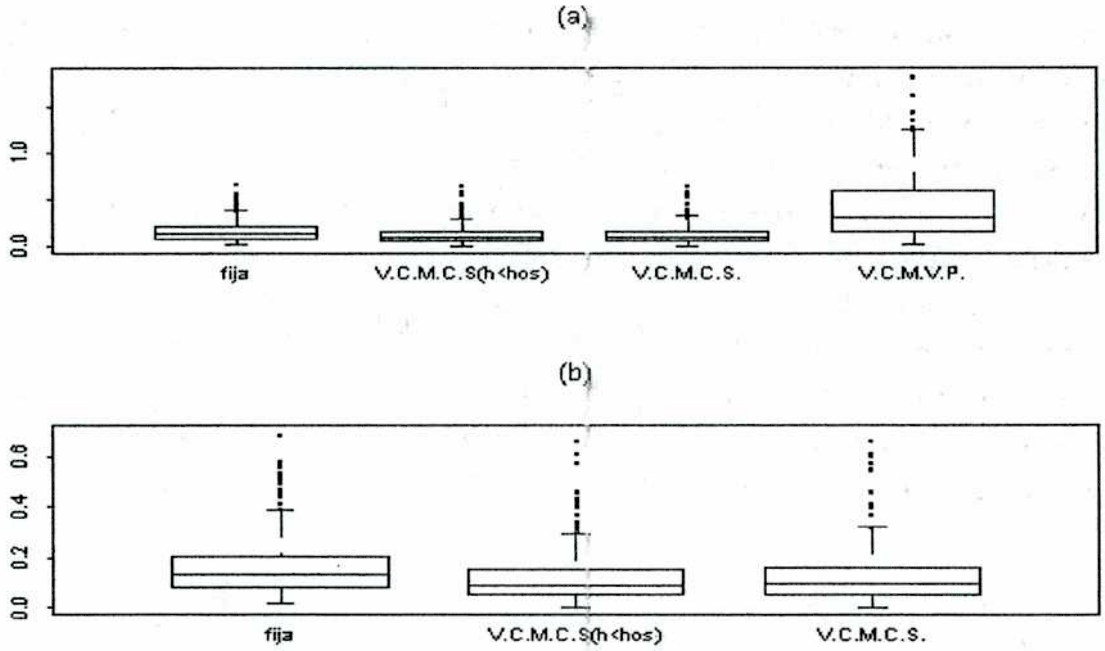

Fig.29. $F \sim N(0,1)$, $F_0 \sim N(1,2^2)$. (a) Boxplot de los valores de $MECI_1$ de estimador de método núcleo, con ventanas h_{opt}, $h_{VCS}<h_{os}$, h_{VCS} sin cota y h_{VMP}. (b) Boxplot de los valores de $MECI_1$ de estimador de método núcleo, con ventanas h_{opt}, $h_{VCS}<h_{os}$ y h_{VCS} sin cota.

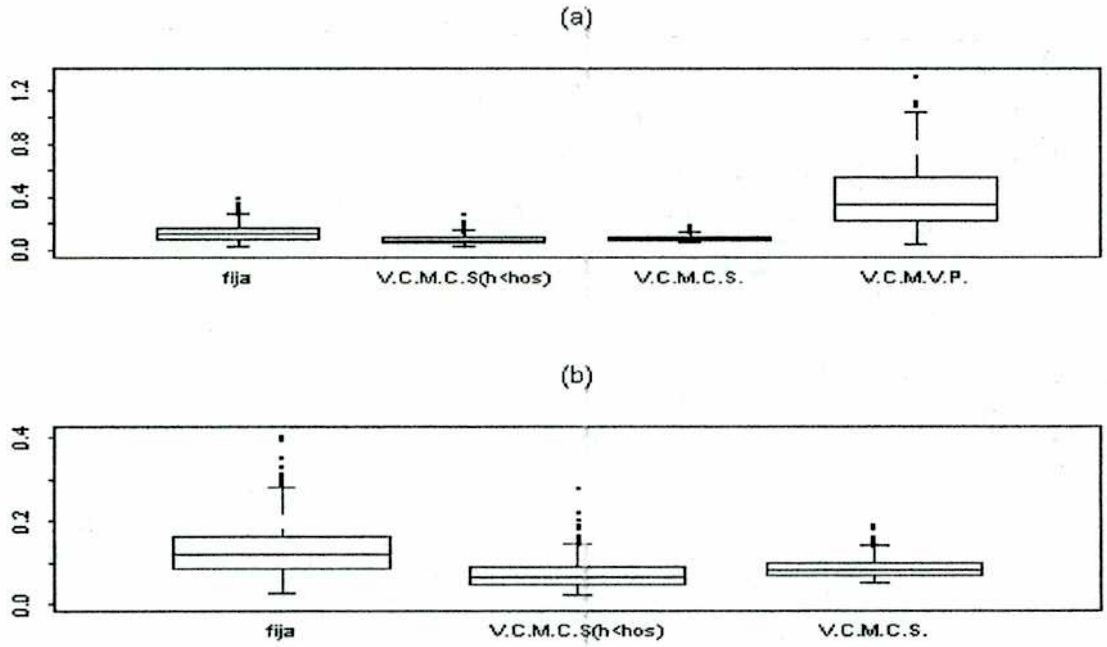

Fig.30. $F \sim N(0,1)$, $F_0 \sim t_2$. (a) Boxplot de los valores de $MECI_1$ de estimador de método núcleo, con ventanas h_{opt}, $h_{VCS}<h_{os}$, h_{VCS} sin cota y h_{VMP}. (b) Boxplot de los valores de $MECI_1$ de estimador de método núcleo, con ventanas h_{opt}, $h_{VCS}<h_{os}$ y h_{VCS} sin cota.

4.5-Ejemplos de aplicación.

Ejemplo 1.

Consideramos un conjunto de datos que es un subconjunto del estudiado en 1995 en la Tesis de Master de I. Oliveira, de la Universidad de Lisboa, Portugal, quien realizó un análisis de componentes principales. La variable que nos interesa es el largo de las hojas de castaños y se estudiaron las observaciones correspondientes al año 1988 y a las variedades *Lada* y *Longal*, del género *Castanea*. El objetivo era comparar la distribución de esta variable en las dos variedades elegidas. Para ello estimamos la densidad relativa mediante estimadores basados en núcleos, eligiendo como ventana la ventana de VCMCS, acotada superiormente por h_{os}.

Indicaremos por X la variable largo correspondiente a la variedad *Lada* y X_0 la correspondiente a la variedad *Longal*. Los tamaños de muestras de ambas variables fueron de 95 unidades.

A continuación, presentamos algunas medidas resumen para ambas variedades.

Variable	Min.	Q_1	Q_2	Media	Q_3	Max.
Lada, X	122	140	152	157	173	204
Longal, X_0	115	144	160	160.6	175	208

Con los datos observados, calculamos los predictores de la variable relativa usando la distribución empírica de X_0, calculada en los valores muestrales de la variable X, o sea, $R_{i,m}=F_{0,n}(X_i)$. El cuadro que presentamos muestra medidas resumen para $R_{i,m}$.

Variable	Min.	Q_1	Q_2	Media	Q_3	Max.
R	0.02105	0.1789	0.3895	0.4551	0.7158	0.9579

El valor de ventana sobresuavizada para la variable relativa es igual a $h_{os} = 1.144*95^{-0.2}*S_R = 0.1399 \cong 0.14$, donde S_R es el estimador del desvío poblacional σ_R.

El gráfico del estimador de la densidad relativa se presenta en Fig.31. De acuerdo a sus características podemos deducir que, aproximadamente, hasta el percentil 0.10 de la variable X_0 la densidad de X se mantiene menor que la de X_0, entre este percentil y el 0.38, la densidad de X sobrepasa a la de X_0, y a partir de

este último la densidad de X_0 es mayor que la de X. De cualquier manera, las diferencias no son muy grandes, ya que $g_n(r)$ está bastante cercana a 1. La ventana automática considerada fue igual a 0.1311 menor, por supuesto que h_{os}=0.14.

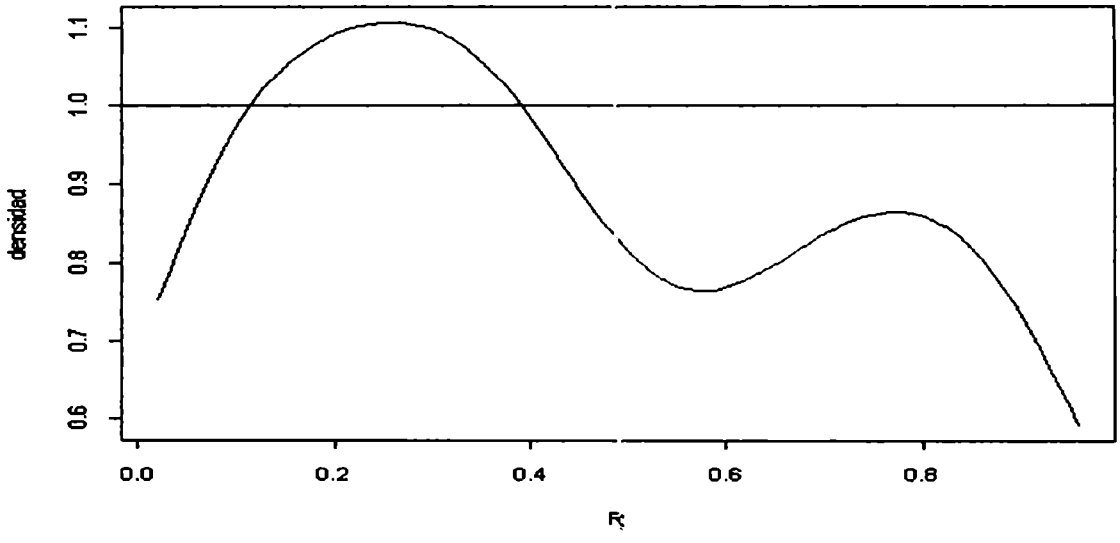

Fig.31. Estimador basado en núcleos de la densidad relativa con núcleo gaussiano y ventana de VCMCS acotada superiormente por h_{os}.

Podemos confirmar estas relaciones observando el gráfico de los estimadores de la densidad de las variables X y X_0 que se muestran en la Fig.32. El soporte de los valores de X está contenido en el de X_0. Esto explica que el gráfico de $g_n(r)$ indique que para los valores inferiores la densidad de X sea menor que la de X_0, ya que se toma como cero. Haciendo esta salvedad, se observa que la relación entre las dos densidades estimadas, calculadas de la misma manera que para la variable relativa, está explicada por el comportamiento del estimador de la densidad relativa. Los valores de ventanas obtenidos, utilizando el método de convalidación cruzada suavizada, fueron 9.3 y 9.02 mientras que las ventanas sobresuavizadas fueron 9.68 y 9.82, para X_0 y X respectivamente.

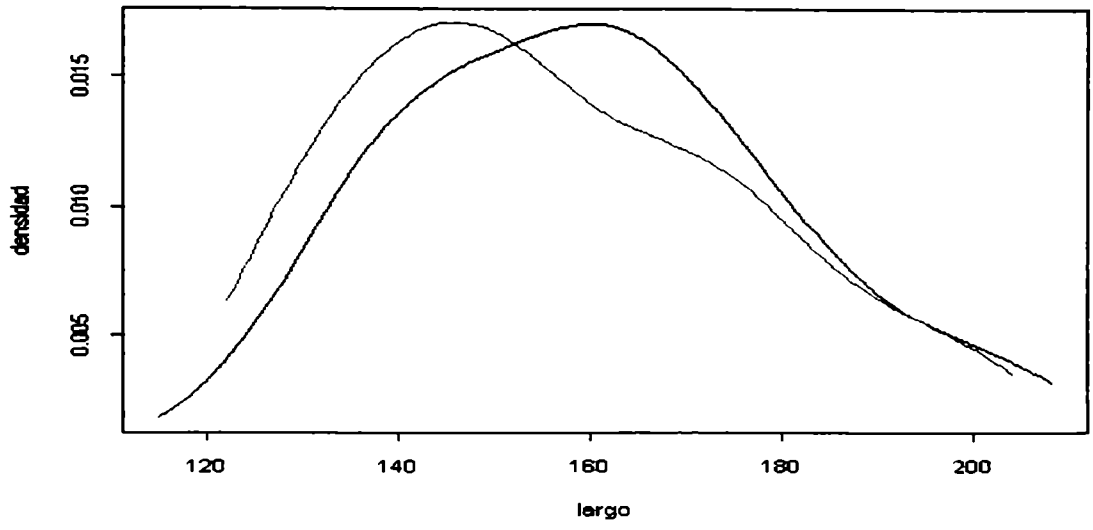

Fig.32. Estimadores de las densidades de las variables X_0 representando el largo de hoja de la variedad Longal, con trazo negro, y X representando la misma medida en la variedad Lada, con trazo rojo. Método basado en núcleos con ventanas de VCMCS acotadas superiormente por la ventana h_{os}.

Ejemplo 2.

En la Chacra Experimental de Patagones (Pcia. de Río Negro), correspondiente a la zona fitogeográfica de monte, se recogieron medidas de biomasa aérea de gramíneas, en unidades de muestreo cuadradas de $0.25m^2$. Un campo se dejó de testigo y en otro se realizó una quema en tiempo anterior. De cada campo se observaron 120 unidades experimentales. Se trata de observar las diferencias en las densidades de la variable biomasa en ambos campos. Los datos fueron recogidos por los señores H.D. Giorgetti, O.A. Montenegro y G.D. Rodriguez y cedidos para esta tesis por el Lic. R. Camina (Dto. de Matemática de la Univ. Nac. del Sur) a través del Dr. C.A. Busso (Dto. de Agronomía, Univ. Nac. del Sur). Llamaremos a la biomasa del campo testigo X_0 y a la del campo quemado X. Las medidas resumen de ambas variables se presentan a continuación. Los valores se dan en Apéndice 2.

Variable	Min.	Q_1	Q_2	Media	Q_3	Max.
Testigo	1.48	12.34	19.78	24.1	35.27	61.8
Quemado	2.63	15.82	24.42	24.98	33.69	59.4

Se calcularon los predictores de la variable relativa $R_{i,m}=F_{0,n}(X_i)$, $i=1,...n=120$, siendo $F_{0,n}$ la distribución empírica de la biomasa aérea de gramíneas del campo testigo.
Medidas resumen del conjunto $R_{i,m}$, son las siguientes,

Variable	Min.	Q_1	Q_2	Media	Q_3	Max.
R	0.01667	0.3646	0.5917	0.547	0.725	0.9917

Estimamos la densidad relativa utilizando un estimador de núcleos con núcleo gausiano y ventana elegida por convalidación cruzada sesgada acotando los valores posibles por $h_{os}=1.144*120^{-0.2}*S_R=0.097$. Como en el ejemplo anterior, $S_R=0.2215$ estima al desvío poblacional de R. El valor de la ventana automática fue igual a $h_{VCS}=0.088$. El gráfico del estimador de la densidad relativa, se presenta en la Fig.33. En él se puede observar que la densidad de biomasa en monte sin manejo es mayor que en el campo con tratamiento de fuego hasta su percentil 0.30. Luego se invierte la desigualdad hasta su percentil 0.82 aproximadamente y a partir de éste, vuelve a ser mayor la densidad de biomasa en el monte testigo. En la Fig.34 presentamos las curvas de densidad para los dos conjuntos de biomasas, que avalan lo mostrado en la curva del estimador de la densidad relativa. Estos estimadores fueron obtenidos utilizando un estimador de núcleos, con ventanas de validación cruzada sesgada acotadas por el valor h_{os} correspondiente y con núcleo gaussiano. Las ventanas del estimador de núcleo para la variable testigo y quemado fueron iguales a 4.22 y 4.54, respectivamente y las ventanas suavizadas h_{os} tomaron los valores 6.49 y 4.94, en el mismo orden.
Se puede observar que las curvas de densidad se cruzan dos veces. Todo esto indicaría que los valores menores de biomasa por unidad se presentan, con mayor densidad, en los campos testigos. El 50%, aproximadamente, de los valores centrales de biomasa se presentan con mayor frecuencia en monte quemado previamente, y solo el 18% de los valores mayores de biomasa tienen mayor densidad en el monte testigo. Esto nos llevaría a pensar que el tratamiento por fuego es efectivo en la cantidad de biomasa de gramíneas. Es interesante comparar el gráfico de $g_n(r)$ con el del ejemplo anterior, en donde el estimador toma valores más cercanos a 1 que en este caso, mostrando una mayor similitud entre el largo de

hojas de las dos variedades, que las que existen entre el valor de biomasa de los dos montes.

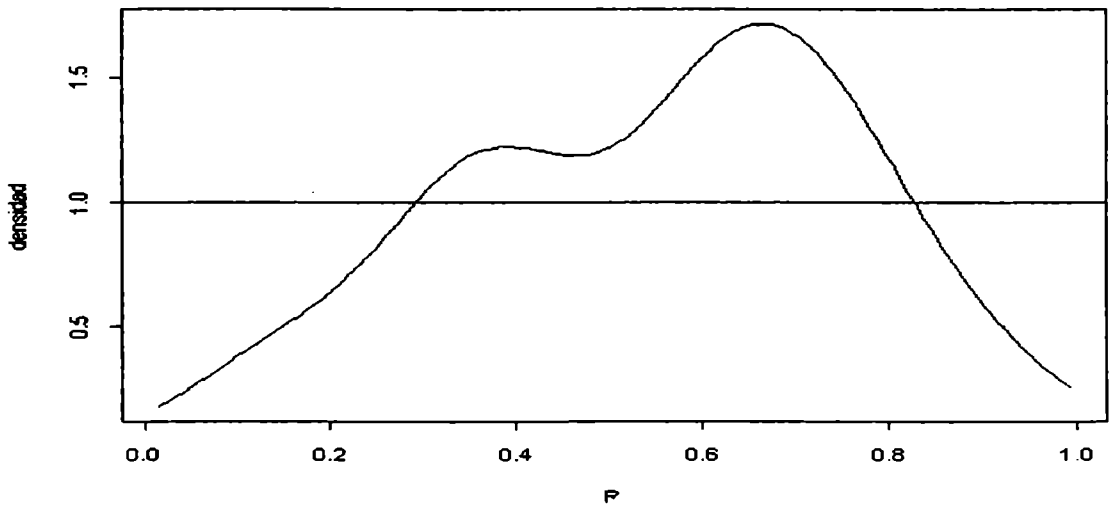

Fig.33. Estimador de núcleos de la densidad relativa, para el ejemplo de biomasa aérea en gramíneas.

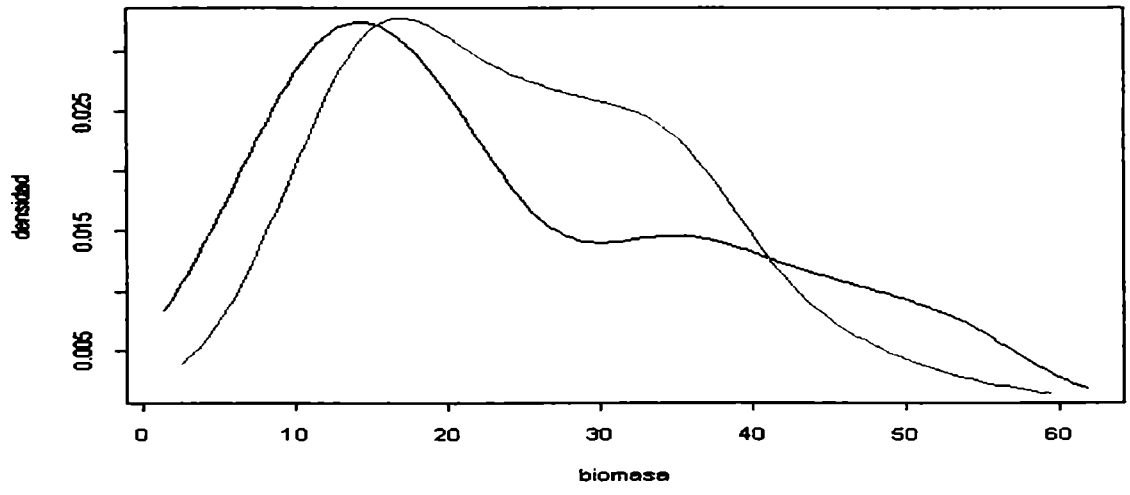

Fig.34. Estimador de núcleos de la densidad de la cantidad de biomasa aérea de gramíneas. La línea negra corresponde a campos testigos y la línea roja a campos quemados previamente. Ventana de VCMCS, acotada por h_{os}, núcleo gaussiano.

Agradecimientos. Agradecemos a la Ing. M. J. Pimentel Pereira del Departamento de Ciencias Biológicas y Ambientales de la Universidad de Trás-os-Montes y Alto Douro, Portugal, por autorizar el uso de los datos de castaño y al Lic. R. Camina y Dr.

C.A. Busso del Departamento de Matemática y Departamento de Agronomía, respectivamente, de la Universidad Nacional del Sur por el uso de los datos de biomasa aérea de gramíneas.

Bibliografía

[1] Abramson, I.S. (1982)."Arbitrariness of the pilot estimator in adaptive kernel methods". *Journal of Multivariate Analysis*, **12**, 562-567.

[2] Boyd, D.W. y Steele, J.M. (1978). "Lower bounds for nonparametric density estimation rates". *The Annals of Statistics*, **6**, 932-934.

[3] Bowman, A.W. (1984). "An alternative method of cross-validation for the smoothing of density estimates". *Biometrika*, **71**, 353-360.

[4] Cao, R. Cuevas, A. y González-Manteiga, W. (1994). "A comparative study of several smoothing methods in density estimation". *Journal of Computational and Statistical Data Analysis*, **17**, 153- 176.

[5] Cencov, N.N. (1962). "Evaluation of an unknow density from observations". *Soviet Mathematics*, **3**, 1559-1562.

[6] Chow, Y.S., Geman, S. y Wu, L.D. (1983). "Consistent cross-validated density estimation". *The Annals of Statistics*, **11**, 25-38.

[7] Devroye, L. y Penrod, C.S. (1984a). "The consistency of automatic kernel density estimates". *The Annals of Statistics*, **12**, No 4, 1231-1249.

[8] Devroye, L. y Penrod, C.S. (1984b). "Distribution-free lower bounds in density estimation". *The Annals of Statistics*, **12**, No 4, 1250-1262.

[9] Devroye, L. y Györfy, L. (1985). *Non-parametric Density Estimation: the L1 view*. John Wiley & Sons, Inc., New York.

[10] Devroye, L. (1987). *A Course in Density Estimation*. Birkhäuser, Boston.

[11] Duin, R.P.W. (1976). "On the choice of smoothing parameters for Parzen estimators of probability density functions". *IEEE Trans. Comput.*, **C-25**, 1175-1179.

[12] Epanechnikov, V.K. (1969). "Nonparametric estimation of a multivariate probabiity density". *Theor. Probab. Appl.*, **14**, 153-158.

[13] Eubank, R.L., LaRiccia, V.N. y Rosenstein, R.B.(1987). "Test statistics derived as components of Pearson's phi-squared distance measure". *Journal of the American Statistical Association*, **82**, 816-825.

[14] Farrell, R.H. (1972). "On the best obtainable asymptotic rates of convergence in estimation of the density function at a point". *The Annals of Mathematical Statisticas*, **43**, 170-180.

[15] Fernholz, L. (1991). "Alomost sure convergence of smoothed empirical distribution functions". *Scandinavian Journal of Statistics,* **18**, 255-262.

[16] Fisher, R.A. (1922). "On the mathematical foundations of theoretical statistics". *Philosophical Trans. Royal Society London Serie A*, **222**, 309-368.

[17] Fisher, R.A. (1932). *Statistical Methods for research workers*. Fourth Edition. Oliver and Boy, Edinburg.

[18] Fix, E. y Hodges, J.L. (1951). "Nonparametric discrimination: consistency properties". *Report number 4, USAF School of Aviation Medicine, Randolph Field*, Texas.

[19] Graunt, J. (1662). *Natural and Political Observations Made upon the Bills of Mortality*. Martyn, Londres.

[20] Gastwirth, J.L. (1968). "The First-Median Test: A two-sided version of the control median test". *Journal of the American Statistical Association,* **63**, 692-706.

[21] Habbema, J.D.F., Hermans, J. y van der Broek, K. (1974). "A stepwise discrimination program using density estimation". In: *Compstat Proceedings in Computational Statistics*, G. Bruckman Ed., pag.100-110, Physica-Verlag, Vienna.

[22] Hall, P. (1983). "Large sample optimality of least squares cross-validationin density estimation". *The Annals of Statistics*, **11**, 1156-1174.

[23] Hall, P. y Marron, J.S. (1987a). "On the amount of noise inherent in bandwidth selection for a kernel density estimator". *The Annals of Statistics*, **15**, 163-181.

[24] Hall, P. y Marron, J.S. (1987b). "Extent to which least-squares cross-validation minimises integrated square error in nonparametric density estimation". *Probability Theory and Related Fields*, **74**, 567-581.

[25] Hall, P, y Wand,M.P., (1988), "Minimizing L1 distance in nonparametric density estimation". *Journal of Multivariate Annalysis*, **26**, 59-88.

[26] Handcock, M.S. y Janssen,P.L. (2000). Unpublished. "Statistical inference for the relative distribution".

[27] Härdle, W. (1991). *Smoothing Techniques with implementation in S*. Springer Series in Statistics. Verlag, New York.

[28] Izenman, A.J. (1991). "Recent developments in nonparametric density estimation". *Journal of the American Statistical Association*, **86**, 205-224.

[29] Loeve, M. (1962). *Probability Theory*. D. Van Nostrand Company, Inc., Princeton, New Jersey.

[30] Loftsgaarden, D.O. y Quesenberry, C.P. (1965). "A nonparametric estimate of a multivariate density function". *The Annals of Mathematical Statistics*, **36**, 1049-1051.

[31] Marron, J.S. (1985). "An asymptotically efficient solution to the bandwidth problem of kernel density estimation". *The Annals of Statistics*, **13**, 1011-1023.

[32] Mehra, K.L. y Rama Krishnaiah, Y.S. (2002). "In pointwise laws of iterated logarithm for estimators of certain conditional functionals". A aparecer en Journal of Nonparametric Statistics.

[33] Moore, D. S., Yackel, J. W. (1976). "Large sample properties of nearest neighbor density function estimators". *Estatistical Decision Theory and Related Topics*, S.S.Gupta and D.S. Moore, Eds. Academic Press, New York.

[34] Moore, D. S., Yackel, J. W. (1977). "Consistency properties of nearest neighbor density function estimators". *The Annals of Statistics*, **5**, No 1, 143-154.

[35] Parzen, E. (1962). "On estimation of probability density function and mode". *The Annals of Mathematical Statistics*, **33**, 1065-1076.

[36] Parzen, E. (1992). "Comparison change analysis". *Nonparametric Statistics and Related Topics*, 3-15.

[37] Rosenblatt, M. (1956). "Remarks on some nonparametric estimates of a density function". *The Annals of Mathematical Statistics*, **27**, 832-837.

[38] Rudemo, M. (1982). "Empirical choice of histograms and kernel density estimators". *Scandinavian Journal of Statistics*, **9**, 65-78.

[39] Schwartz, S.C. (1967). "Estimation of a probability density by an orthogonal series". *The Annals of Mathematical Statistics*, **38**, 1262-1265.

[40] Scott, D.W., Gorry, G.A., Cole, J.S. y Gorry, G.A. (1978). "Plasma lipids as collateral risk factors in coronary artery

disease: a study of 371 males with chest pain". *Journal of Chronic diseases*, **31**, 337-345.

[41] Scott, D.W. (1985a). "Frequency polygons". *Journal of American Statistical Association,* **80**, 348-354.

[42] Scott, D.W. (1985b). "Average shifted histograms: effective nonparametric density estimators in several dimensions". *The Annals of Statistics,* **13**, 1024-1040.

[43] Scott, D.W. y Terrell, G.R. (1987). "Biased and unbiased cross-validation in density estimation". *Journal of American Statistical Association*, **82**, 1131-1146.

[44] Scott, D.W. y Wand, M.P. (1991). "Feasibility of multivariate density estimates". *Biometrika*, **78**, 197-206.

[45] Scott, D.W. (1992). *Multivariate density estimation*. John Wiley & Sons.

[46] Silverman, B.W. (1978). "Density ratios, empirical likelihood and cot death". *Applied Statistics*, **27**, 26-33.

[47] Silverman, B.W. (1992). *Density Estimation for statistics and data analysis*. Chapman & Hall, London.

[48] Stone, C.J. (1984). "An asymptotically optimal window selection rule for kernel density estimates". *The Annals of Statistics,* **12**, 1285-1297.

[49] Sturges, H.A. (1926). "The choice of a class interval". *Journal of American Statistical Association,* **21**, 65-66.

[50] Tarter, M.E. y Kronmal, R.A. (1970). "On multivariate density estimates based on orthogonal expansions". *The Annals of Mathematical Statistics*, **41**, 718-722.

[51] Thompson, J.R. y Tapia, R.A. (1990). *Nonparametric Function Estimation, Modeling, and Simulation*. SIAM, Philadelphia.

[52] Wahba, G. (1971). "A polinomial algorithm for density estimation". *The Annals of Mathematical Statistics,* **42**, 1870-1886.

[53] Wand, M. Y Jones, M. C. (1995). *Kernel Smoothing*. Chapman and Hall., London.

[54] Watson, G.S. y Leadbetter, M.R. (1963). "On the estimation of the probability density I". *The Annals of Mathematical Statistics*, **34**, 480-491.

[55] Woodroofe, M. (1970). "On choosing a delta-sequence". *The Annals of Mathematical Statistics*, **41**, 1665-1671.

Apéndice 1- Programas de computación

Programas realizados para obtener los resultados Monte Carlo

```
gaussian1<-function(x)
{
#función núcleo gausiano o normal, se llama a esta función,
#en todos los programas, por lo cual debe estar declarada
#previamente.
result <- 0.3989 * exp(-0.5 * x^2)
result
}

Program1<-function(datos, grilla)
{
#Calcula la fda empírica, de la variable datos, en valores de
# la grilla.
n <- length(datos)
m <- length(grilla)
datos <- sort(datos)
grilla <- sort(grilla)
fda <- c(1:m)
aux <- c(1:n)
for(j in 1:m) {
for(i in 1:n) {
aux[i] <- (datos[i] <= grilla[j])/n
}
fda[j] <- sum(aux)
}
#da como resultado Fdatos(grilla)
#para graficarla, considerar los pares (grilla, fda)
fda
}

Program2<-function(x, h1, a)
{
# este programa expresa las funciones (33) y (34), que se
# integrarán en el programa Program3, para encontrar el h
#óptimo del Caso 1, método núcleo.
((pnorm((a * x)/h1) - pnorm((a * (x - 1))/h1))/(2 * sqrt(pi) * h1
          ) - (pnorm(x/h1) - pnorm((x - 1)/h1))^2)/100 + (pnorm(x/
          h1) - pnorm((x - 1)/h1) - 1)^2
}

Program3<-function(h)
{
#cálculo de MECI1 para una grilla de 0.01,...,0.99, con h=100.
#segun se vió en (32),(33) y (34).
mesi <- c(1:(h - 1))
for(k in 1:(h - 1)) {
aa <- integrate(Program2, lower = 0, upper = 1, h1 = k/h, a =
1.4142)
mesi[k] <- aa$integral
```

```
}
#da los valores de MECI1, se grafican los pares (grilla,mesi)
#a fin de observar para qué valor de la grilla alcanza el mínimo.
mesi
}


Program4<-function(nsim)
{
#método núcleo, Caso 1, ventana óptima, cálculo de MECI1.
#nsim= número de simulaciones.
nucleo <- gaussian1 #nucleo gaussiano.
MECI1h <- c(1:nsim)
R<-c(1:100)
gmhm <- c(1:99)
grilla <- c(1:99)/100
auxiliar <- c(1:100)
hfijo <- 0.11  # valor obtenido en Program3
for(t in 1:nsim) {
R <- acum(rnorm(100, 0, 1), rnorm(100, 0, 1))
#esta es el estimador de g(r), que llamo gmhm(r):
for(j in 1:99) {
for(i in 1:100) {
#sobre los datos en cada v. grilla
auxiliar[i] <- nucleo((grilla[j] - R[i])/hfijo)
}
gmhm[j] <- sum(auxiliar)/(100 * hfijo)
}
#calculo el MECI1,le resto 1 porque cuando F0=F la g(r)=1:
MECI1 <- (sum((gmhm - 1)^2))/99
#lo guardo para cada simulacion:
MECI1h[t] <- MECI1
}
# da los valores de MECI1.
MECI1h
}
```

```
Program5<- function(nsim)
{
#método núcleo, Caso 1,ventana VCMCS, cálculo de MECI1.
MECI1h <- matrix(rep(c(0), 2 * nsim), ncol = 2)
#grilla de posibles h's, desde 1/100 a 60/100,acotadas para que
#corra más rápido el programa, ya que nunca superaron el
#valor 0.35:
h <- c(1:60)/100
nucleo <- gaussian1
M <- rep(c(0), 60)    #variable auxiliar
BCVh <- c(1:60) #funcion a minimizar por el ancho de ventana
#según la notacion de Scott(1992), pag.167.
gmhm <- c(1:99) #vector para el estimador calculado en
#puntos de la grilla.
Raizpi <- sqrt(pi)
grilla <- c(1:99)/100
R<-c(1:100)
for(t in 1:nsim) {
#Predictores de la variable relativa,
R <- acum(rnorm(100, 0, 1), rnorm(100, 0, 1))
#ahora comienzo el cálculo de BCV(h)
for(u in 1:60) {
for(i in 1:99) {
s <- i + 1
for(j in s:100) {
delta <- (R[i] - R[j])/h[u]
delta1 <- (delta^4 - 12 * (delta^2) + 12) * exp( - (delta^2)/4)
M[u] <- delta1 + M[u]
}
}
M[u] <- M[u]/(32 * 100)
BCVh[u] <- (1 + M[u])/(2 * 100 * h[u] * Raizpi)
}
#este es el h que minimiza el BCV(h):
hm <- order(BCVh)[1]/100
#esta es la estimacion de g(r), por metodo de nucleo:
for(j in 1:99) {
for(i in 1:100) {
#sobre los datos en cada valor de la grilla
```

```
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}
#le resto 1 porque cuando F0=F, g(r)=1:
#El MECI1 para cada muestra:
MECI1 <- (sum((gmhm - 1)^2))/99
MECI1h[t, 1] <- MECI1
MECI1h[t, 2] <- hm
}
#da la matriz de pares (MECI1(h),h)
MECI1h
}
Program6<-function(nsim)
{
#Método núcleo, Caso 1, ventana VCMCS, ventanas acotadas
#por hos, cálculo de MECI1.
MECI1h <- matrix(rep(c(0), 2 * nsim), ncol = 2)
#para grilla de posibles h's, acotadas:
nucleo<-gaussian1
acoth <- 0.13 * 100   #0.13= hos
h <- c(1:acoth)/100
M <- rep(c(0), acoth)  #variable auxiliar
BCVh <- c(1:acoth)
gmhm <- c(1:99)
Raizpi <- sqrt(pi)
grilla <- c(1:99)/100
R <- c(1:100)
for(t in 1:nsim) {
#Predictores de la variable relativa,
R <- acum(rnorm(100, 0, 1), rnorm(100, 0, 1))
#ahora comienzo el BCV(h):
for(u in 1:acoth) {
for(i in 1:99) {
s <- i + 1
for(j in s:100) {
delta <- (R[i] - R[j])/h[u]
#pag.167 de Scott
delta1 <- (delta^4 - 12 * (delta^2) + 12) * exp( - (delta^2)/4)
```

```
M[u] <- delta1 + M[u]
}
}
M[u] <- M[u]/(32 * 100)
BCVh[u] <- (1 + M[u])/(2 * 100 * h[u] * Raizpi)
}
#este es el h que minimiza el BCV(h),
hm <- order(BCVh)[1]/100
#esta es la estimacion de g(r), por método de nucleo:
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}
#le resto 1 porque cuando F0=F, g(r)=1:
#El MECI1 para cada muestra:
MECI1 <- (sum((gmhm - 1)^2))/(99)
MECI1h[t, 1] <- MECI1
MECI1h[t, 2] <- hm
}
MECI1h
#da pares de valores (MECI1(h),h)
}

Program7<-function(nsim)
{
#Método núcleo, Caso 1, ventana de VCMVP, cálculo de MECI1
nucleo <- gaussian1
gmhm <- c(1:99)
tri <- c(1:100)
p <- nucleo(0)
#acoto los valores de h para que el programa sea más rápido, ya
#que no superaron el valor 0.06.
#posibles valores de h óptimo para este procedimiento.
h <- c(1:20)/100
gsombrero <- c(1:100)
Lh <- c(1:20)
R <- c(1:100)
```

```
MECI1hm <- matrix(rep(c(0), 2 * nsim), ncol = 2)
grilla <- c(1:99)/100      #la grilla donde se calculará el
# estimador.
for(t in 1:nsim) {
R <- acum(rnorm(100, 0, 1), rnorm(100, 0, 1))
#aquí encuentro el h optimo
for(i in 1:20) {
for(j in 1:100) {
for(s in 1:100) {
tri[s] <- nucleo((R[j] - R[s])/h[i])
}
gsombrero[j] <- (sum(tri) - p)/(99 * h[i])
}
Lh[i] <- prod(gsombrero)
}
hm <- order(Lh)[20]/100
#aquí empiezo el estimador, con la ventana hm
for(j in 1:99) {
for(i in 1:100) {
#K((r-Rj)/hm), r=grilla, Rj=R[j]
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}
MECI1 <- (sum((gmhm - 1)^2))/99
MECI1hm[t, 1] <- MECI1
MECI1hm[t, 2] <- hm
}
#Da los pares (MECI1(h),h)
MECI1hm
}
Program8<- function(nsim)
{
#Calcula la ventana óptima para encontrar la versión
#suavizada del predictor, para método de vecinos más cercanos,
#según se explica en Capítulo 4, parágrafo 2-2.
h <- c(1:100)/100
s0 <- 2 * nsim
hm <- matrix(rep(c(0), s0), ncol = 2)
```

```
MECI1F0 <- c(1:100)
tri <- c(1:100)
R <- c(1:100)
for(t in 1:nsim) {
y <- rnorm(100, 0, 1)
x <- rnorm(100, 0, 1)
for(s in 1:100) {
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/h[s])
}
R[j] <- sum(tri)/100
}
MECI1F0[s] <- sum((R - pnorm(y))^2)/100
}
hm[t, 2] <- order(MECI1F0)[1]/100
hm[t, 1] <- sort(MECI1F0)[1]
}
#Da pares de valores de (h,MECI1(h)) en cada simulación
#a fin de elegir el valor de h que minimiza MECI1.
#elegimos h=0.36.
hm
}

Program9<- function(nsim)
{
#Método vecinos más cercanos, Caso 1, ventana  número
#de vecinos óptimo, cálculo de MECI1.
nucleo <- gaussian1
MECI1 <- c(1:nsim)    # aquí guardamos los resultados del MECI1
hmkopt <- c(1:99) #para cada punto de la grilla, guardamos la
#ventana optima funcion del  k=22 vecino más cercano.
gmhm <- c(1:99)
#guardamos los valores del estimador, en puntos de
#la grilla.
grilla <- c(1:99)/100
p <- nucleo(0)
tri <- c(1:100) #vector auxiliar.
R <- c(1:100)   #aquí guardamos los predictores
```

```
                # de la variable relativa.
koptimo <- 22   #100*hopt*2, hopt=0.11
for(t in 1:nsim) {
#Encontraremos los predictores Rj,m,h, según la expresion (31).
#empiezo a generar las muestras
y <- rnorm(100, 0, 1)
x <- rnorm(100, 0, 1) #Construimos los predictores:
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.36)
#h=0.36 minimiza el MECI1 del estimador suavizado (31), como
#se vió en Program8.
}
R[j] <- sum(tri)/100
}
# ahora encuentro el Hk,r.
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
hmk <- sort(tri)
hmkopt[j] <- hmk[koptimo]
#Ahora encuentro el estimador de g(x) por el metodo
# de vecinos más cercanos.
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - 1)^2)/99
}
#este es el MECI1:
MECI1
}

Program10<-
function(nsim)
{
#Método de vecinos más cercanos, Caso 1, ventana número de vecinos
```

```
# de VCMC, cálculo de MECI1.
#declaración de variables
#acotamos mínimamente el orden del vecino más cercano
k0 <- 4
k1 <- 96
nucleo <- gaussian1
R <- c(1:100)
MECI1 <- c(1:nsim)
hmkopt <- c(1:99)
gmhm <- c(1:99)
kk0 <- k1 - k0 + 1
VCh <- c(1:kk0)
grh <- c(1:99)
grilla <- c(1:99)/100
h <- c(1:100)
p <- nucleo(0)
tri <- c(1:100)
gsombrero <- c(1:100)
h1 <- c(1:100)
tri1 <- c(1:100)
for(t in 1:nsim) {
#
#empiezo a generar la muestra y las variables predictivas de la
#variable relativa R.
y <- rnorm(100, 0, 1)
x <- rnorm(100, 0, 1)
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.36)
}
R[j] <- sum(tri)/100
#R[j]= estimacion suavizada de la v. relativa.
}
#Ahora empiezo el calculo de VC(h(k)) según expresión (36)
for(k in 1:kk0) {
for(j in 1:100) {
for(i in 1:100) {
h[i] <- abs(R[j] - R[i])   #distancia del valor R[i] a los
                           #otros valores R[j]
```

```
}
m <- sort(h)     #m[1]=0
hm <- m[k + k0]  #distancia de R[j] al vecino k+k0-1

for(l in 1:100) {
tri[l] <- nucleo((R[j] - R[l])/hm)
}
gsombrero[j] <- (sum(tri) - p)/(99 * hm)
}
#este es el segundo termino de VC(h(k))
ter2 <- (2 * sum(gsombrero))/100
#ahora el primer termino,  o sea la integral
#del estimador de g(r) por vecino cercano, que la
#estimamos como se hace usualmente ,por una sumatoria.
for(i in 1:99) {
for(s in 1:100) {
h1[s] <- abs(grilla[i] - R[s])
#distancia del punto de la grilla
# a los datos R[s].
}
mm <- sort(h1)
hm1 <- mm[k + k0 - 1]
#distancia del punto de la grilla
# al dato k+k0-1 mas cercano.
for(j in 1:100) {
tri1[j] <- nucleo((grilla[i] - R[j])/hm1)
}
grh[i] <- sum(tri1)/(100 * hm1)
grh[i] <- grh[i]^2
}
ter1 <- sum(grh)/100  #el termino de la integral
VCh[k] <- ter1 - ter2
#para cada k calculamos VC(h(k))
}
#y vemos cual es el k que lo minimiza
#Ahora encuentro el k optimo de este metodo, o sea el que
#minimiza VC(h(k)).
koptimo <- order(VCh)[1] + k0 - 1
# ahora encuentro el hn(k) para cada punto de la grilla
```

```
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
#distancia del punto de la
#grilla a los R[i].
hmk <- sort(tri)      #ordeno las 100 distancias

hmkopt[j] <- hmk[koptimo]
#este es la abertura de ventana
#que usaremos y que depende del punto donde calculamos el
#estimador y del vecino "koptimo"-esimo.
#Ahora encuentro la estimación por el método
# de vecinos más cercanos
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
#g(r)=1.
MECI1[t] <- sum((gmhm - 1)^2)/99
}
#este es el MECI1:
MECI1
}


Program11<- function(nsim)
{
#Método de vecinos más cercanos, Caso 1, ventana de vecinos VCMVP.
#Cálculo de MECI1.
#acotamos el orden del vecino:
k0 <- 4
k1 <- 96
MECI1 <- c(1:nsim)
hmkopt <- c(1:99)
gmhm <- c(1:99)
kk0 <- k1 - k0 + 1
Lhk <- c(1:kk0)
```

```
grh <- c(1:99)
grilla <- c(1:99)/100
h <- c(1:100)
nucleo <- gaussian1
p <- nucleo(0)
tri <- c(1:100)
gsombrero <- c(1:100)
h1 <- c(1:100)
tri1 <- c(1:100)
R <- c(1:100)
for(t in 1:nsim) {
# Comienza el calculo de los predictores de la variable
#relativa R.
y <- rnorm(100, 0, 1)
x <- rnorm(100, 0, 1)
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.36)
#h=0.36 minimizaba el MECI1 del estimador suavizado de
# la fda F0, según la expresión (31).
}
R[j] <- sum(tri)/100
}
#
for(k in 1:kk0) {
#
#Ahora encuentro el L(h(k)), funcion de pseudo
#maxima verosimilitud, expresion (38).
for(j in 1:100) {
for(i in 1:100) {
h[i] <- abs(R[j] - R[i])
}
m <- sort(h)
hm <- m[k + k0]
#hm es la distancia del k+k0-1 vecino a R[j]
#la minima es cero, por eso sumo 1.
#Ahora estimo la g en cada R[j] usando los restantes R[i]
#
for(l in 1:100) {
```

```
tri[l] <- nucleo((R[j] - R[l])/hm)
}
gsombrero[j] <- (sum(tri) - p)/(99 * hm)
}
Lhk[k] <- prod(gsombrero)
}
#el orden de vecino que maximiza L(h(k)) lo llamo koptimo:
koptimo <- order(Lhk)[kk0] + k0 - 1
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
hmk <- sort(tri)
#el ancho de ventana, funcion del koptimo, para el punto
#j de la grilla, lo llamamos hmkopt[j]
hmkopt[j] <- hmk[koptimo]
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
#estimador por vecino cercano de g(r), en el punto j
#de la grilla:
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - 1)^2)/99
}
MECI1
}


Program12<- function(nsim)
{
#Método núcleo, Caso 2, ventana óptima.
#Comienzo calculando el hopt, tal como se explica
#en Cap. 4, parágrafo 3.1-Estimador de núcleos,
#ventana fija
sigma0 <- 2
media0 <- 1
n <- 100
sigmacuadrado <- sigma0^2
```

```
a <- (1 - sigmacuadrado) * (2 - sigmacuadrado)
b <- media0 * sigma0 * (3 - 2 * sigmacuadrado)
cc <- (media0^2) * sigmacuadrado + 1 - sigmacuadrado
d <- sqrt(sigmacuadrado - 2.5)
l <- (sigma0 * media0)/sqrt(sigmacuadrado - 2.5)
mediaS <-  - (l/d)
sigmaScuadrado <- 1/(2 * (d^2))
sigmaS <- sqrt(sigmaScuadrado)
M <- sigmacuadrado * ((2 * pi)^1.5) * exp(l^2 - media0^2)
ES4 <- mediaS^4 + 6 * mediaS^2 * sigmaS^2 + 3 * sigmaS^4
ES3 <- mediaS^3 + 3 * mediaS * sigmaS^2
ES2 <- mediaS^2 + sigmaS^2
ES1 <- mediaS
ES0 <- 1
M1 <- M * sqrt(2 * pi) * sigmaS
ter1 <- (a^2) * ES4
ter2 <- 2 * a * b * ES3
ter3 <- (b^2 + 2 * a * cc) * ES2
ter4 <- 2 * b * cc * ES1
ter5 <- (cc^2) * ES0
fin <- M1 * (ter1 - ter2 + ter3 - ter4 + ter5)
hestrella <- (1/(2 * sqrt(pi) * fin * n))^0.2
#con el hestrella=h_opt= 0.04927824 voy a calcular gmhm:
#calculo de gmhm:
R <- c(1:100)
nucleo <- gaussian1
grilla <- c(1:99)/100
tri <- c(1:100)
gmhm <- c(1:99) #valores del estimador en la grilla
gr <- vector(length = 99)
F0menos <- vector(length = 99)
MECI1h<- c(1:nsim)
#valor de g(r):
for(i in 1:99) {
F0menos[i] <- qnorm(grilla[i], 1, 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dnorm(F0menos[i],1, 2)
}
#comienzo las simulaciones:
for(t in 1:nsim) {
```

```
R <- acum(rnorm(100, 1, 2), rnorm(100, 0, 1))
for(j in 1:99) {
for(i in 1:100) {
#sobre los datos en cada v. grilla
tri[i] <- nucleo((grilla[j] - R[i])/hestrella)
}
gmhm[j] <- sum(tri)/(100 * hestrella)
}
#ya tengo gmhm[j] y gr[j], ahora MECI1:
MECI1h[t] <- (sum((gmhm - gr)^2))/99
}
#Valores de MECI1 para h=h_opt.
MECI1h
}

Program13<-function(nsim)
{
#Método núcleos, Caso 2, ventana VCMCS, cálculo de MECI1.
#Aquí se anotan los resultados: el MECI1 y el h óptimo
#correspondiente para c/simulacion
MECI1h <- matrix(rep(c(0), 2 * nsim), ncol = 2)
gr <- c(1:99)    #g(r) teorico, conocido.
tri <- c(1:100) #grilla de h's, acotada, h<0.4:
acoth <- 40      #Se tomaron como valores de ventana
#desde 0.01 hasta 0.4. Sobre 500 valores sólo 4 lo superaron.
# En este Caso 2, la acotación aconsejada fue la de hos=0.08,
# que dió valores menores del error.
h <- c(1:40)/100
M <- rep(c(0), 40)    #auxiliar, parcial
BCVh <- c(1:40)
nucleo <- gaussian1
gmhm <- c(1:99) #estimacion de g por nucleos
Raizpi <- sqrt(pi)
grilla <- c(1:99)/100
R <- c(1:100)    #calculo de g(r) teórica:
for(i in 1:99) {
F0menos[i] <- qnorm(grilla[i], 1, 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dnorm(F0menos[i], 1, 2)
}
```

```
for(t in 1:nsim) {
R <- acum(rnorm(100, 1, 2), rnorm(100, 0, 1))
#ahora comienzo el BCV(h):
for(u in 1:40) {
for(i in 1:99) {
s <- i + 1
for(j in s:100) {
delta <- (R[i] - R[j])/h[u]
delta1 <- (delta^4 - 12 * (delta^2) + 12) * exp( - (delta^2)/4)
M[u] <- delta1 + M[u]
}
}
M[u] <- M[u]/(32 * 100)
BCVh[u] <- (1 + M[u])/(2 * 100 * h[u] * Raizpi)
}
#este es el h que minimiza el BCV(h);
hm <- order(BCVh)[1]/100
#Estimador de g(r):
for(j in 1:99) {
for(i in 1:100) {
#sobre los datos en cada v. grilla
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}
MECI1 <- (sum((gmhm - gr)^2))/99
MECI1h[t, 1] <-MECI1
MECI1h[t, 2] <- hm
}
MECI1h
}

Program14<-function(nsim)
{
#Método de núcleos, Caso 2, ventana de VCMCS, h<h_os=0.08
#aquí se anotan los resultados: el MECI1 y el h óptimo
#correspondiente para c/simulación.
MECI1h <- matrix(rep(c(0), 2 * nsim), ncol = 2)
gr <- c(1:99)   #g(r) teórico.
```

```
nucleo<-gaussian1
tri <- c(1:100)
acoth <- 0.08 *100
h <- c(1:acoth)/100
M <- rep(c(0), acoth) #auxiliar, parcial
BCVh <- c(1:acoth)
gmhm <- c(1:99) #estimacion de g por nucleos
Raizpi <- sqrt(pi)
grilla <- c(1:99)/100
R<-c(1:100)
#cálculo de g(r) teorica:
for(i in 1:99) {
F0menos[i] <- qnorm(grilla[i], 1, 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dnorm(F0menos[i],1, 2)
}

for(t in 1:nsim) {
#R[i] son los predictores de la variable relativa.
R <- acum(rnorm(100, 1, 2), rnorm(100, 0, 1))
#ahora comienzo el BCV(h):
for(u in 1:acoth) {
for(i in 1:99) {
s <- i + 1
for(j in s:100) {
delta <- (R[i] - R[j])/h[u]
delta1 <- (delta^4 - 12 * (delta^2) + 12) * exp( - (delta^2)/4)
M[u] <- delta1 + M[u]
}
}
M[u] <- M[u]/(32 *100)
BCVh[u] <- (1 + M[u])/(2 * n0 * h[u] * Raizpi)
}
#este es el h que minimiza el BCV:
hm <- order(BCVh)[1]/100
#esta es la estimacion de la g(r):
for(j in 1:99) {
for(i in 1:100) {
#sobre los datos en cada v. grilla
tri[i] <- nucleo((grilla[j] - R[i])/hm)
```

```
}
gmhm[j] <- sum(tri)/(100 * hm)
}
MECI1 <- (sum((gmhm - gr)^2))/99
MECI1h[t, 1] <- MECI1
MECI1h[t, 2] <- hm
}
MECI1h
}

Program15<- function(nsim)
{
#Método de núcleos, Caso 2, ventana de VCMVP, cálculo de MECI1.
R <- c(1:100)
nucleo <- gaussian1
gr <- c(1:99)   #g(r) teorico, conocido.
gmhm <- c(1:99)
tri <- c(1:100)
p <- nucleo(0)
#acoto los valores de h, según los resultados que observé,
#para los posibles valores de h que maximizan L(h),
#y que no alcanzan a 0.1
h <- c(1:20)/100
gsombrero <- c(1:100)
Lh <- c(1:20)
MECI1hm <- matrix(rep(c(0), 2 * nsim), ncol = 2)
grilla <- c(1:99)/100  #la grilla es desde 1/100
# a 99/100
#cálculo de g(r) teorica:
for(i in 1:99) {
F0menos[i] <- qnorm(grilla[i], 1, 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dnorm(F0menos[i], 1, 2)
}
for(t in 1:nsim) {
R <- acum(rnorm(100, 1, 2), rnorm(100, 0, 1)) #predictores
#aqui encuentro la ventana
for(i in 1:20) {
for(j in 1:100) {
for(s in 1:100) {
```

```
tri[s] <- nucleo((R[j] - R[s])/h[i])
}
gsombrero[j] <- (sum(tri) - p)/(99 * h[i])
}
Lh[i] <- prod(gsombrero)
}
hm <- order(Lh)[20]/100    #aqui empiezo el estimador de g(r).
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}
MECI1 <- (sum((gmhm - gr)^2))/99
MECI1hm[t, 1] <- MECI1
MECI1hm[t, 2] <- hm
}
MECI1hm
}
```

```
Program16<-function(nsim)
{
#Método de vecinos más cercanos, Caso 2, ventana: número
#de vecinos óptimo, cálculo de MECI1.
nucleo <- gaussian1
MECI1 <- c(1:nsim)    #F0=N(1,2) F=N(0,1)
gr <- c(1:99)
hmkopt <- c(1:99)
gmhm <- c(1:99)
grilla <- c(1:99)/100
tri <- c(1:100)
R <- c(1:100)
F0menos <- c(1:99)
koptimo <- 10   #100*hopt*2, hopt=0.0493, óptimo o fijo.
#calculo de g(r) teorico en los puntos de la grilla
#estos valores no dependen de las simulaciones
for(i in 1:99) {
```

```
F0menos[i] <- qnorm(grilla[i], 1, 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dnorm(F0menos[i], 1, 2)
}
#empiezo las simulaciones
for(t in 1:nsim) {
#
#empiezo a generar las muestras de X y X0
x <- rnorm(100, 1, 2) #muestra de F0
y <- rnorm(100, 0, 1) #muestra de F
#tomo como ventana para la expresión suavizada de F0
#el hopt=0.0493.
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.0493)
#K'((yj-xi)/0.0493)
}
R[j] <- sum(tri)/100  #estimador suavizado de la fda F0(X)=R
}
# ahora encuentro el hn(k) para cada punto de la grilla
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
hmk <- sort(tri)
#k_óptimo lo calculo como  2*100*0.0493=10
#la abertura de ventana esta dada por la distancia
#del punto de la grilla al décimo vecino más cercano
#entre los Ri, ya que k_óptimo=10.
hmkopt[j] <- hmk[koptimo]
#Ahora encuentro la estimacion por el metodo
#de vecinos cercanos
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - gr)^2)/99
}
#este es el MECI1:
```

```
MECI1
}

Program17<- function(nsim)
{
#Método de vecinos más cercanos, Caso 2, ventana:
#vecinos de VCMC, cálculo de MECI1.
k0 <- 4
k1 <- 96
nucleo <- gaussian1
MECI1 <- c(1:nsim)
hmkopt <- c(1:99)
gr <- c(1:99)
gmhm <- c(1:99)
kk0 <- k1 - k0 + 1
R <- vector(length = 100)
VCh <- c(1:kk0)
grh <- c(1:99)
grilla <- c(1:99)/100
h <- c(1:100)
p <- nucleo(0)
tri <- c(1:100)
gsombrero <- c(1:100)
h1 <- c(1:100)
tri1 <- c(1:100)
F0menos <- vector(length = 99)
#valor de g(r) en puntos de la grilla
for(i in 1:99) {
F0menos[i] <- qnorm(grilla[i], 1, 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dnorm(F0menos[i], 1, 2)
}
for(t in 1:nsim) {
# Cálculo de predictores que notamos R[j].
x <- rnorm(100, 1, 2)
y <- rnorm(100, 0, 1)
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.0493)
}
```

```
R[j] <- sum(tri)/100
}
#Ahora empiezo el calculo de VC(h(k)) según expresión (36)
for(k in 1:kk0) {
for(j in 1:100) {
for(i in 1:100) {
h[i] <- abs(R[j] - R[i])
}
m <- sort(h)
hm <- m[k + k0]
for(l in 1:100) {
tri[l] <- nucleo((R[j] - R[l])/hm)
}
gsombrero[j] <- (sum(tri) - p)/(99 * hm)
}
ter2 <- (2 * sum(gsombrero))/100
#ahora el primer termino
for(i in 1:99) {
for(s in 1:100) {
h1[s] <- abs(grilla[i] - R[s])
}
mm <- sort(h1)
hm1 <- mm[k + k0 - 1]
for(j in 1:100) {
tri1[j] <- nucleo((grilla[i] - R[j])/hm1)
}
grh[i] <- sum(tri1)/(100 * hm1)
grh[i] <- grh[i]^2
}
ter1 <- sum(grh)/100
VCh[k] <- ter1 - ter2
}
koptimo <- order(VCh)[1] + k0 - 1
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
hmk <- sort(tri)
hmkopt[j] <- hmk[koptimo]
```

```
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - gr)^2)/99
}
#Vector de MECI1.
MECI1
}

Program18<- function( nsim)
{
#Método de vecinos más cercanos, Caso 2, ventana: vecinos
#más cercanos de VCMVP, cálculo de MECI1.
nucleo<-gaussian1
k0 <- 4
k1 <- 96   #F0=N(1,2),F=N(0,1)
MECI1 <- c(1:nsim)
F0menos <- c(1:99)
gr <- c(1:99)
hmkopt <- c(1:99)
gmhm <- c(1:99)
kk0 <- k1 - k0 + 1
Lhk <- c(1:kk0)
grh <- c(1:99)
grilla <- c(1:99)/100
h <- c(1:100)
p <- nucleo(0)
tri <- c(1:100)
gsombrero <- c(1:100)
h1 <- c(1:100)
R<-c(1:100)
tri1 <- c(1:100)#Ahora la g(r)
for(i in 1:99) {
F0menos[i] <- qnorm(grilla[i], 1, 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dnorm(F0menos[i], 1, 2)
}
for(t in 1:nsim) {
```

```
# Aquí va la muestra Rj, que sale integrando el nucleo:
x <- rnorm(100, 1, 2)
y <- rnorm(100, 0, 1)
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.0493)
#K'((yj-xi)/0.0493)
#0.0493 era el h fijo optimo
}
R[j] <- sum(tri)/100  #Rj=suma(K')/n0
}
#
for(k in 1:kk0) {

#Ahora encuentro el L(h(k))
for(j in 1:100) {
for(i in 1:100) {
h[i] <- abs(R[j] - R[i])
}
m <- sort(h)
hm <- m[k + k0]
#esta es la distancia del k+k0-1 vecino a r[j]
#la minima es cero, por eso sumo 1
#ahora estimo la g en cada R[j] sacando esa observacion:
for(l in 1:100) {
tri[l] <- nucleo((R[j] - R[l])/hm)
}
gsombrero[j] <- (sum(tri) - p)/(99 * hm)
}
Lhk[k] <- prod(gsombrero)
}
koptimo <- order(Lhk)[kk0] + k0 - 1
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
hmk <- sort(tri)
hmkopt[j] <- hmk[koptimo]
for(i in 1:100) {
```

```
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - gr)^2)/99
}
MECI1
}

Program19<- function(nsim)
{
#Método núcleos, Caso 3, Ventana óptima.
MECI1h <- c(1:nsim)
h <- 0.04458  #hopt=0.04458
R <- c(1:100)
F0menos <- c(1:99)
gmhm <- c(1:99)
gr <- c(1:99)
grilla <- c(1:99)/100
#Calculo la función g(r):
for(i in 1:99) {
F0menos[i] <- qt(grilla[i], 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dt(F0menos[i], 2)
}
for(t in 1:nsim) {
R <- acum(rt(100, 2), rnorm(100, 0, 1)) #Predictores de la
                                        #variable relativa.
#Cálculo del estimador de g(r).
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/h)
}
gmhm[j] <- sum(tri)/(h * 100)
}
MECI1h[t] <- (sum((gmhm - gr)^2))/99
}
MECI1h
}
```

```
Program20<-
function(nsim)
{
#Método núcleos, Caso 3, ventana de VCMCS.
MECI1h <- matrix(rep(c(0), 2 * nsim), ncol = 2)
nucleo<-gaussian1
gr <- c(1:99)
R<-c(1:100)
tri <- c(1:100) #grilla de h's, acotada, h<0.4:
acoth <- 0.4 * 100
h <- c(1:40)/100 #todos los valores de h que minimizaron la
#expresión BCV(h) fueron menores que 0.3,
#y no había mínimos locales.
M <- rep(c(0), acoth)
BCVh <- c(1:acoth)
gmhm <- c(1:99) #estimacion de g por nucleos
Raizpi <- sqrt(pi)
grilla <- c(1:99)/100
#calculo de g(r) teorica:
for(i in 1:99) {
F0menos[i] <- qt(grilla[i], 2)
#r[i] son los datos
gr[i] <- dnorm(F0menos[i], 0, 1)/dt(F0menos[i], 2)
}
for(t in 1:nsim) {
R <- acum(rt(100, 2), rnorm(100, 0, 1))
#ahora comienzo BCV(h):
for(u in 1:acoth) {
for(i in 1:99) {
s <- i + 1
for(j in s:100) {
delta <- (R[i] - R[j])/h[u]
delta1 <- (delta^4 - 12 * (delta^2) + 12) * exp( - (delta^2)/4)
M[u] <- delta1 + M[u]
}
}
M[u] <- M[u]/(32 * 100)
BCVh[u] <- (1 + M[u])/(2 * 100 * h[u] * Raizpi)
}
```

```
#este es el h que minimiza  BCV(h):
hm <- order(BCVh)[1]/100

#este es el estimador de g(r):
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}

MECI1 <- (sum((gmhm - gr)^2))/99
MECI1h[t, 1] <-MECI1
MECI1h[t, 2] <- hm
}
MECI1h
}

Program21<- function(nsim)
{
#Método núcleo, Caso 3, ventana de VCMCS, con h<h_o3=0.11
MECI1h <- matrix(rep(c(0), 2 * nsim), ncol = 2)
nucleo<-gaussian1
gr <- c(1:99)
tri <- c(1:100)
acoth <- 0.11 * 100 #h_o3=0.11
h <- c(1:acoth)/100
M <- rep(c(0), acoth)#auxiliar, parcial
BCVh <- c(1:acoth)
R<-c(1:100)
gmhm <- c(1:99)#estimacion de g por nucleos
Raizpi <- sqrt(pi)
grilla <- c(1:99)/100
#calculo de g(r) teórica:
for(i in 1:99) {
F0menos[i] <- qt(grilla[i], 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dt(F0menos[i], 2)
}
for(t in 1:nsim) {
```

```
R <- acum(rt(100, 2), rnorm(100, 0, 1))
#ahora comienzo el BCV(h):
for(u in 1:acoth) {
for(i in 1:99) {
s <- i + 1
for(j in s:100) {
delta <- (R[i] - R[j])/h[u]
delta1 <- (delta^4 - 12 * (delta^2) + 12) * exp( - (delta^2)/4)
M[u] <- delta1 + M[u]
}
}

M[u] <- M[u]/(32 * 100)
BCVh[u] <- (1 + M[u])/(2 * 100 * h[u] * Raizpi)
}
#este es el h que minimiza el BCV:
hm <- order(BCVh)[1]/100
#esta es la estimacion de la g(r):
for(j in 1:99) {
for(i in 1:100) {
#sobre los datos en cada v. grilla
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}
MECI1 <- (sum((gmhm - gr)^2))/99
MECI1h[t, 1] <- MECI1
MECI1h[t, 2] <- hm
}
MECI1h
}

Program22<- function(nsim)
{
#Método núcleos, Caso 3, ventana de VCMVP.
nucleo<-gaussian1
gr <- c(1:99)#g(r) teorico, conocido.
gmhm <- c(1:99)
tri <- c(1:100)
```

```
p <- nucleo(0)
acoth <- 0.2 * 100 #acotamos los valores de h,
h <- c(1:acoth)/100
gsombrero <- c(1:100)
Lh <- c(1:acoth)
MECI1hm <- matrix(rep(c(0), 2 * nsim), ncol = 2)
grilla <- c(1:99)/100
#calculo de g(r) teorica:
for(i in 1:99) {
F0menos[i] <- qt(grilla[i], 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dt(F0menos[i], 2)
}
for(t in 1:nsim) {
R <- acum(rt(100, 2), rnorm(100, 0, 1))
#aqui encuentro la ventana
for(i in 1:acoth) {
for(j in 1:100) {
for(s in 1:100) {
tri[s] <- nucleo((R[j] - R[s])/h[i])
}
gsombrero[j] <- (sum(tri) - p)/(99 * h[i])
}
Lh[i] <- prod(gsombrero)
}
hm <- order(Lh)[acoth]/100 #aqui empiezo el gmhm
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hm)
}
gmhm[j] <- sum(tri)/(100 * hm)
}
MECI1 <- (sum((gmhm - gr)^2))/99
MECI1hm[t, 1] <- MECI1
MECI1hm[t, 2] <- hm
}
MECI1hm
}

Program23<-function( nsim)
```

```
{
#Método de vecinos más cercanos, Caso 3,
#ventana de vecinos óptima.
nucleo<-gaussian1
MECI1 <- c(1:nsim)
gr <- c(1:99)
hmkopt <- c(1:99)
gmhm <- c(1:99)
grilla <- c(1:99)/100
tri <- c(1:100)
R <- c(1:100)
F0menos <- c(1:99)
koptimo <- 9      #100*hopt*2, hopt=0.0446
#calculo de g(r) teorico en los puntos de la grilla
#estos valores no dependen de las simulaciones
for(i in 1:99) {
F0menos[i] <- qt(grilla[i], 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dt(F0menos[i], 2)
}
#empiezo las simulaciones
for(t in 1:nsim) {
#
# Cálculo de los predictores Ri,m,h
x <- rt(100, 2)
y <- rnorm(100, 0, 1)
#tomo como ventana la óptima para este caso.
#hopt =0.0446.
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.0446)
}
R[j] <- sum(tri)/100
}
# Calculo de la ventana Hk,r, con k=9, r=grilla[j].
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
hmk <- sort(tri)
```

```
hmkopt[j] <- hmk[koptimo] #este es el valor de H_{k,r}.
#Ahora encuentro el estimador.
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - gr)^2)/99
}
#este es el MECI_1:
MECI1
}

Program24<- function(nsim)
{
#Método de vecinos más cercanos, Caso 3, ventana de vecinos
#de VCMC.
k0<-4
k1<-96
nucleo<-gaussian1
MECI1 <- c(1:nsim)
hmkopt <- c(1:99)
gr <- c(1:99)
gmhm <- c(1:99)
kk0 <- k1 - k0 + 1
VCh <- c(1:kk0)
R<-c(1:100)
grh <- c(1:99)
grilla <- c(1:99)/100
h <- c(1:100)
p <- nucleo(0)
tri <- c(1:100)
gsombrero <- c(1:100)
h1 <- c(1:100)
tri1 <- c(1:100)
# Cálculo de g(r):
for(i in 1:99) {
F0menos[i] <- qt(grilla[i], 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dt(F0menos[i], 2)
```

```
}
#Comienzan las simulaciones,
for(t in 1:nsim) {
x <- rt(100, 2) #F0~t2
y <- rnorm(100, 0, 1)  #F~N(0,1)
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.0446)
}
R[j] <- sum(tri)/100  #Predictores, versión suavizada.
}
#Cálculo de VC(h(k)), expresión (36).
for(k in 1:kk0) {
for(j in 1:100) {
for(i in 1:100) {
h[i] <- abs(R[j] - R[i])
}
m <- sort(h)
hm <- m[k + k0]
for(l in 1:100) {
tri[l] <- nucleo((R[j] - R[l])/hm)
}
gsombrero[j] <- (sum(tri) - p)/(99 *hm)
}
ter2 <- (2 * sum(gsombrero))/100 #Segundo término en (36)
#ahora el primer termino
for(i in 1:99) {
for(s in 1:100) {
h1[s] <- abs(grilla[i] - R[s])
}
mm <- sort(h1)
hm1 <- mm[k + k0 - 1]
for(j in 1:100) {
tri1[j] <- nucleo((grilla[i] - R[j])/hm1)
}
grh[i] <- sum(tri1)/(100 * hm1)
grh[i] <- grh[i]^2
}
ter1 <- sum(grh)/100
```

```
VCh[k] <- ter1 - ter2
}
koptimo <- order(VCh)[1] + k0 - 1 #vecino que minimiza (36)
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
}
hmk <- sort(tri)
hmkopt[j] <- hmk[koptimo]
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - gr)^2)/99
}
MECI1
}

Program25<- function(nsim)
{
#Método de vecinos más cercanos, Caso 3, ventana
# de vecinos de VCMVP.
nucleo<-gaussian1
k0 <- 4
k1 <- 96
MECI1 <- c(1:nsim)
F0menos <- c(1:99)
gr <- c(1:99)
hmkopt <- c(1:99)
gmhm <- c(1:99)
kk0 <- k1 - k0 + 1
Lhk <- c(1:kk0)
grh <- c(1:99)
grilla <- c(1:99)/100
h <- c(1:100)
p <- nucleo(0)
tri <- c(1:100)
gsombrero <- c(1:100)
```

```
h1 <- c(1:100)
tri1 <- c(1:100)            #Ahora  g(r)
R<-c(1:100)
for(i in 1:99) {
F0menos[i] <- qt(grilla[i], 2)
gr[i] <- dnorm(F0menos[i], 0, 1)/dt(F0menos[i], 2)
}
for(t in 1:nsim) {
# Aqu va la muestra Rj, que sale integrando el nucleo:
x <- rt(100, 2)
y <- rnorm(100, 0, 1)
for(j in 1:100) {
for(i in 1:100) {
tri[i] <- pnorm((y[j] - x[i])/0.0446)
}
R[j] <- sum(tri)/100     #R[j] predictores
}
#
for(k in 1:kk0) {
#
#Ahora encuentro el L(h(k))
for(j in 1:100) {
for(i in 1:100) {
h[i] <- abs(R[j] - R[i])
}
m <- sort(h)
hm <- m[k + k0]
#esta es la distancia del k+k0-1 vecino a r[j]
for(l in 1:100) {
tri[l] <- nucleo((R[j] - R[l])/hm)
}
gsombrero[j] <- (sum(tri) - p)/(99 *hm)
}
Lhk[k] <- prod(gsombrero)
}
koptimo <- order(Lhk)[kk0] + k0 - 1 #el k que maximiza L(h(k))
for(j in 1:99) {
for(i in 1:100) {
tri[i] <- abs(grilla[j] - R[i])
```

```
}
hmk <- sort(tri)
hmkopt[j] <- hmk[koptimo]  #ventana H_{k,r}
for(i in 1:100) {
tri[i] <- nucleo((grilla[j] - R[i])/hmkopt[j])
}
gmhm[j] <- sum(tri)/(100 * hmkopt[j])
}
MECI1[t] <- sum((gmhm - gr)^2)/99
}
MECI1
}
```

Programas para resolver los ejercicios de aplicación

```
Program26<-
function(r, numh, ngrilla)
{
#calculo del estimador de la densidad, método núcleo,
#ventana de VCMCS, h<h_os.
#r es el vector de datos, numh es el número de ventanas
#para encontrar aquella que minimice BCV(h), ngrilla el
#número de puntos en donde se estimará la densidad.
#Se calcula también la ventana sobresuavizada.
#La salida es el estimador de la densidad en ngrilla valores y
#en último lugar da la ventana h_VCS.
#hos<-1.144
nucleo <- gaussian1
n0 <- length(r)
hos <- 1.144 * (n0^(-0.2)) * sqrt(var(r))
K <- n0 - 1
grilla <- seq(min(r), max(r), length = ngrilla)
tri <- c(1:n0)
#
a <- max(r) - min(r) #grilla de h's, acotada por h_os.
acoth <- ((numh * hos)/a)
h <- (c(1:acoth) * a)/numh
#
M <- rep(c(0), acoth)    #auxiliar, parcial
BCVh <- c(1:acoth)
l <- ngrilla
gmhm <- c(1:l)      #estimacion de g por nucleos
Raizpi <- sqrt(pi)   #ahora comienzo el BCV(h):
for(u in 1:acoth) {
for(i in 1:K) {
s <- i + 1
for(j in s:n0) {
delta <- (r[i] - r[j])/h[u]
```

```
delta1 <- (delta^4 - 12 * (delta^2) + 12)*exp( - (delta^2)/4)
M[u] <- delta1 + M[u]
}
}
#hasta aca es la sumatoria en la expresion del BCV
M[u] <- M[u]/(32 * n0)
BCVh[u] <- (1 + M[u])/(2 * n0 * h[u] * Raizpi)
}
#este es el h que minimiza el BCV:
orden <- order(BCVh)[1]
hm <- h[orden]

#esta es la estimacion de la g(r):
for(j in 1:l) {
for(i in 1:n0) {
#sobre los datos, en cada v. grilla
tri[i] <- nucleo((grilla[j] - r[i])/hm)
}
gmhm[j] <- sum(tri)/(n0 * hm)
}
#vector de estimación con última componente igual
#a la ventana.
resultado <- c(gmhm, hm)
resultado
}

Program27<-
function(r, ngrilla)
{
#Da la grilla de valores donde se calcula la
#estimación de la densidad del vector r (Program26).
#Para graficar la densidad hay que plotear el
#resultado de este programa y el del programa
#anterior menos la última componente que da el
#valor de ventana
grilla <- seq(min(r), max(r), length = ngrilla)
grilla
}
```

Apéndice 2- Datos de biomasa aérea de gramíneas.

Valores de biomasa aérea de gramíneas para campos testigo

1.48	2.52	3.04	3.14	4.90	4.93	5.49	5.77	6.43	6.61	6.87
7.30	8.10	8.46	9.00	9.13	9.81	9.92	10.38	10.44	10.76	10.94
11.30	11.33	11.35	11.70	11.85	11.94	12.01	12.08	12.42	12.62	12.68
12.80	12.89	13.30	14.07	14.10	14.25	14.30	14.46	14.60	14.97	15.68
16.26	16.41	16.62	17.01	17.04	17.16	17.22	17.26	17.58	18.28	18.69
19.09	19.11	19.46	19.57	19.62	19.94	20.10	20.32	20.70	20.80	21.38
21.43	21.57	21.64	22.08	23.15	25.02	25.10	25.36	25.50	26.30	26.47
27.42	27.50	29.90	30.86	31.60	31.67	31.90	32.30	32.50	33.64	33.80
34.64	35.26	35.30	36.08	36.83	37.20	37.78	38.10	38.54	39.10	40.09
40.42	41.10	41.97	42.70	44.16	45.06	45.61	45.62	46.59	47.10	47.31
48.30	50.30	51.16	52.70	53.30	54.17	54.34	54.70	55.70	61.80	

Valores de biomasa aérea de gramíneas para campos con quema programada

2.63	6.96	7.52	7.71	9.11	9.80	10.85	10.88	11.08	11.27	11.49
11.73	12.23	12.85	13.19	13.20	13.33	13.41	14.10	14.20	14.35	14.36
14.57	14.69	14.79	14.82	15.20	15.32	15.33	15.63	15.89	16.17	16.30
16.56	16.60	16.65	17.02	17.25	17.26	17.47	17.52	17.58	18.20	18.22
18.24	18.37	18.55	19.80	20.93	20.93	20.94	20.97	21.23	21.30	21.59
22.51	22.70	22.85	24.13	24.16	24.67	24.86	24.90	24.90	25.04	25.06
25.52	25.70	25.80	26.06	26.31	27.03	27.30	27.34	27.70	27.97	28.76
28.90	28.90	29.33	29.40	30.49	30.81	31.06	31.38	32.36	32.84	33.00
33.42	33.68	33.71	34.40	34.58	34.60	34.67	35.00	35.02	35.20	35.23
35.70	35.84	35.88	36.50	36.80	36.90	37.30	37.72	38.55	38.70	40.60
42.00	43.07	43.50	44.10	46.34	46.60	49.70	50.32	54.72	59.40	