

Estimación de conjuntos de nivel

Paula Saavedra Nieves

Proyecto fin de máster tutorizado por D. Alberto Rodríguez Casal.

Firmado: Alberto Rodríguez Casal.

Índice

1. Introducción	4-8
2. Método plug-in	9-27
2.1 Método plug-in. Samworth y Wand	17-27
3. Método exceso de masa	28-30
4. Método suavizado granulométrico	31-33
5. Estudio de simulación	34-57
5.1 Comparación método Samworth-Wand con validación cruzada y Seather-Jones	34-40
5.2 Comparación método plug-in Samworth y Wand y exceso de masa	41-46
5.3 Comparación método plug-in Samworth y Wand y granulométrico	47-57
6. Apéndice	58-63

Bibliografía

1. Introducción

La estimación de conjuntos es un área de la estadística no paramétrica en la que la geometría juega un papel esencial. En general, aborda el problema de reconstrucción de un conjunto compacto R en el espacio euclídeo \mathbb{R}^d a partir de una muestra aleatoria de puntos cuya distribución está relacionada con el conjunto. Por ejemplo, R podría representar el soporte de la distribución generadora de la muestra considerada.

A menudo, será necesario imponer determinadas restricciones geométricas sobre el conjunto de partida para que la estimación pueda realizarse, debido a que la familia de conjuntos compactos euclidianos es demasiado numerosa incluyendo miembros verdaderamente extraños (*monster sets*).

Por otra parte, es posible considerar la estimación de conjuntos como la contrapartida geométrica a la teoría clásica de estimación no paramétrica de funcionales (Simonoff, 1996). En ambas teorías, los estimadores dependerán de una secuencia de parámetros de suavizado, los resultados teóricos se centran en propiedades asintóticas, especialmente consistencia y tasas de convergencia, y el principal objetivo pasa por obtener resultados tan generales como sea posible, es decir, que se mantengan para condiciones muy generales y dimensión d arbitraria.

Las principales diferencias con la estimación no paramétrica funcional surgen de la fuerte motivación geométrica escondida detrás de la estimación de conjuntos. Así, la forma de los conjuntos y las distancias entre ellos juegan un papel fundamental en el problema que nos ocupa.

En este sentido, una de las distancias más empleadas en este contexto es la distancia en medida, definida sobre los subconjuntos acotados de \mathbb{R}^d pertenecientes al σ -álgebra de Borel.

Definición 1.1 Sean A y C dos conjuntos de Borel acotados. La distancia en medida entre A y C se define como

$$d_\mu(A, C) = \mu(A \Delta C),$$

donde μ denota la medida de Lebesgue, pudiendo ser reemplazada por cualquier otra medida de interés, y Δ , la diferencia simétrica usual dada por:

$$A \Delta C = (A \setminus C) \cup (C \setminus A).$$

Nota 1.2 No resulta complicado notar que d_μ no define una métrica si μ denota la medida de Lebesgue. En efecto, si se consideran B_1 y B_2 , conjuntos de Borel acotados e idénticos salvo en un conjunto finito de puntos, entonces $B_1 \neq B_2$ y $d_\mu(B_1, B_2) = 0$.

Otra alternativa para medir la distancia entre dos conjuntos la proporciona la distancia Hausdorff, definida sobre el espacio de subconjuntos no vacíos y compactos de un espacio métrico arbitrario, en particular, considerando el espacio Euclídeo d -dimensional \mathbb{R}^d , dotado del producto escalar y de la norma, $\| \cdot \|$, correspondientes.

Definición 1.3 Sean $A, C \subset \mathbb{R}^d$ y $\lambda \in \mathbb{R}$, escribimos $\lambda A = \{\lambda a : a \in A\}$ y denotaremos por

$$A \oplus C = \{a + c : a \in A, c \in C\}$$

la adición de Minkowski, y por

$$A \oplus C = \{x : x + C \subset A\}$$

la diferencia de Minkowski, donde escribimos $x + C$ para $\{x\} \oplus C$.

Definición 1.4 Sean A y C dos subconjuntos no vacíos y compactos de \mathbb{R}^d . La distancia Hausdorff entre A y C se define como

$$d_H(A, C) = \max\left\{\sup_{a \in A} d(a, C), \sup_{c \in C} d(c, A)\right\},$$

donde

$$d(a, C) = \inf\{\|a - c\| : c \in C\}.$$

Equivalentemente:

$$d_H(A, C) = \inf\{\varepsilon > 0 : A \subset C \oplus \varepsilon B, C \subset A \oplus \varepsilon B\},$$

donde $B = \{x \in \mathbb{R}^d : \|x\| < 1\}$.

Nota 1.5 Al definir la distancia Hausdorff sobre la colección de subconjuntos no vacíos y compactos de \mathbb{R}^d se puede garantizar que d_H es una métrica.

Si X denota una variable aleatoria absolutamente continua con valores en \mathbb{R} , distribución P_X y función de densidad f , la estimación del soporte suele carecer de interés ya que una parte considerable del mismo estará prácticamente vacío en términos de probabilidad. Consecuentemente, a menudo, interesará estimar regiones de soporte substancial también denominadas conjuntos de nivel o regiones HDR (*Highest Density Regions*).

Definición 1.6 Sea $\tau \in (0, 1)$, se define el f_τ -conjunto de nivel o $100(1 - \tau)\%$ - región de elevada densidad como

$$R_\tau = \{x \in \mathbb{R} : f(x) \geq f_\tau\}$$

donde

$$f_\tau = \inf\left\{y \in (0, \infty) : \int_{-\infty}^{\infty} f(t) \mathbb{I}_{\{f(t) \geq y\}} dt \leq 1 - \tau\right\}.$$

Análogas consideraciones podrían realizarse si X representase una variable aleatoria absolutamente continua d -dimensional. En tal caso, R_τ seguiría denotando al conjunto de nivel y $\hat{R}_{\tau, n}$ al estimador considerado.

Dada una muestra $\mathcal{X}_n = \{X_1, \dots, X_n\}$ de P_X existen tres alternativas para construir un estimador, $\hat{R}_{\tau,n}$, de un conjunto de nivel, R_τ : Método plug-in, método de exceso de masa y método de suavizado granulométrico. Analicemos con algo de detalle cada una de las posibilidades.

Los **estimadores plug-in** conforman quizás la elección más natural para resolver el problema que nos ocupa. La idea consiste en substituir f por un estimador f_n no paramétrico en la expresión de R_τ ; por tanto, $\hat{R}_{\tau,n} = \{f_n \geq f_\tau\}$. Del mismo modo que en el contexto de la estimación no paramétrica de la densidad, el principal problema será la determinación del parámetro ventana o de suavizado, h .

En general, la estimación de conjuntos de nivel ha sido tratada en la literatura desde múltiples puntos de vista sin profundizar en el estudio de selección del parámetro de suavizado. De hecho, este problema no fue abordado hasta el año 2010, año en el que Samworth y Wand desarrollaron un método selector de ventana para tratar la estimación de conjuntos de nivel (plug-in) considerando una aproximación asintótica de la distribución en media entre R_τ y $\hat{R}_{\tau,n}$.

La literatura existente se centra en la determinación de tasas de convergencia en dimensión d . Recientemente, Mason and Polonik (2009) establecieron, para cualquier medida positiva dominada por la medida de Lebesgue, la normalidad asintótica de la diferencia simétrica entre el conjunto de nivel y el estimador tipo plug-in, obtenido al considerar un estimador tipo núcleo de la densidad.

Cadre (2006) proporcionó condiciones suficientes para garantizar la convergencia en probabilidad de $\sqrt{(nh^d)}\mu(\hat{R}_{\tau,n} \triangle R_\tau)$ a un valor constante en dimensión d , con μ denotando la medida de Lebesgue. En particular, bajo condiciones apropiadas, la tasa de convergencia exacta para el estimador plug-in, $\hat{R}_{\tau,n}$, es $\sqrt{nh^d}$, con respecto a la distancia en medida, d_μ .

Cuevas, González-Manteiga y Rodríguez-Casal (2006) y Molchanov (1998) coincidieron al considerar estimadores plug-in, $\{g_n \geq f_\tau\}$, para conjuntos de nivel generales, $\{g \geq f_\tau\}$, donde g es una función real de interés, no necesariamente una densidad, definida sobre un espacio métrico. Por ejemplo, g podría ser una función de regresión $g(x) = \mathbb{E}(Y | X = x)$ y la esfera unidad S^2 , el espacio métrico correspondiente. Así, mientras los primeros abordaron el problema de estimación de la frontera; los segundos, proporcionaron la tasa exacta de convergencia para $d_H(\hat{R}_{\tau,n}, R_\tau)$, donde d_H denota la distancia Hausdorff.

Cuevas y Fraiman (1997) trataron la estimación del soporte en \mathbb{R}^d empleando para ello una sucesión de estimadores de conjuntos de nivel. La idea base es bastante simple: Bajo condiciones de regularidad, el soporte S coincide esencialmente (salvo conjuntos de medida de Lebesgue, μ , nula) con el conjunto $\{f > 0\}$. Como primera aproximación, se podría pensar en estimar S reemplazando f en $\{f > 0\}$ por un estimador f_n ; sin embargo, si f_n es un estimador tipo núcleo, con kernel estrictamente positivo, entonces $\{f_n > 0\}$ coincidiría con \mathbb{R}^d . De todas formas, la aproximación plug-in permitiría considerar otra posibilidad en este contexto. El soporte S podría ser estimado a través de la sucesión de estimadores $\{f_n > c_n\}$ donde c_n representa una sucesión de parámetros con $c_n \downarrow 0$.

El **método de exceso de masa**, fue propuesto de forma independiente por Hartigan (1987) y por Müller y Sawitzki (1987) y extendido e investigado por Polonik (1995), conformando una alternativa competitiva para abordar la estimación de conjuntos de nivel, teniendo en cuenta que el conjunto de nivel R_τ maximiza el funcional

$$H_\tau(B) = P(B) - f_\tau \mu(B),$$

donde μ denota la medida de Lebesgue y P una distribución d -dimensional.

A diferencia de la aproximación plug-in, el método de exceso de masa se caracteriza por imponer restricciones geométricas sobre los estimadores centrando la búsqueda de los mismos sobre una clase de conjuntos dada. A priori, Hartigan limitó su estudio a la clase de conjuntos convexos y compactos en \mathbb{R}^2 y Nolan (1991) a la clase de elipsoides en \mathbb{R}^d .

Müller y Sawitzki (1991) generalizaron el método de exceso de masa propuesto por Hartigan considerando una clase de conjuntos más general en dimensión uno, no necesariamente convexos, que analizaremos más en profundidad.

En 1995, Polonik proporcionó resultados asintóticos para el estimador restringido a clases de conjuntos mucho menos restrictivas (incluyendo, por ejemplo, conjuntos convexos) en dimensión d empleando la teoría de procesos empíricos. Además, comprobó que el comportamiento asintótico de los estimadores dependerá de las restricciones sobre la suavidad de la densidad f considerada y de la riqueza sobre la clase de conjuntos considerada.

Al igual que en la aproximación plug-in, la metodología de exceso de masa ha tratado la estimación de conjuntos de nivel de regresión $\{x : g(x) \geq f_\tau\}$, $g(x) = \mathbb{E}(Y|X = x)$ (Polonik y Wand, 2005) y la estimación del soporte, bajo restricciones de uniformidad para la densidad f . En efecto, si f es uniforme y f_τ suficientemente pequeño, los problemas de estimación del soporte y R_τ coinciden. Así, Klemelä (1987) empleó una versión penalizada del método de exceso de masa para obtener estimadores S_n del soporte S elegidos entre una clase P_n (n aumentando). El término de penalización ha sido añadido al funcional de exceso de masa para tener en cuenta el tamaño de la clase P_n . Además, bajo la hipótesis de que la frontera del conjunto de nivel puede ser parametrizada como una función Hölder de orden s , Korostelev y Tsybakov (1993) probaron que el orden de convergencia minimax para $\mathbb{E}d_\mu(\hat{R}_{\tau,n}, R_\tau)$ son de orden $O(n^{s/(s+d-1)})$.

Finalmente, el **método de suavizado granulométrico**, propuesto por Walther (1997), permite obtener el estimador del conjunto de nivel, $\hat{R}_{\tau,n}$, como una unión de bolas cuyos centros resultan fácilmente calculables. Al igual que el método de exceso de masa, impone restricciones de tipo geométrico, concretamente r -convexidad, sobre la clase de conjuntos que considera y, de modo análogo al método plug-in, parte de una estimación no paramétrica tipo núcleo de la función de densidad f . La ventaja de este método es que si f es suave el orden de convergencia mejora al obtenido a través del método de exceso de masa propuesto por Polonik (1995).

El desarrollo de las tres aproximaciones consideradas, ha convertido a la estimación de conjuntos de nivel en una herramienta de vital interés en el análisis cluster donde Hartigan marcó el punto de partida identificando un f_τ -cluster con una componente conexa del conjunto de nivel R_τ .

Una propuesta natural para estimar los f_τ -clusters pasa por considerar los clusters empíricos, definidos como las componentes conexas de $R_{\tau,n}$. Los clusters de datos en los que la muestra, \mathcal{X}_n , es dividida podrían ser definidos clasificando las observaciones de acuerdo con los clusters empíricos construidos previamente.

Cadre, Pelletier y Pudlo (2009) propusieron un algoritmo relativamente sencillo para estimar el número de componentes conexas de un conjunto de nivel a través de la aproximación plug-in, probando la consistencia del mismo.

Klemelä (2004, 2008) se basó en la noción de cluster establecida por Hartigan para visualizar estimaciones de densidades multivariantes.

Cuevas, Febrero y Fraiman (2000) se ocuparon de estudiar los problemas teóricos y prácticos derivados de la estimación del número clusters teniendo en cuenta la forma geométrica de los mismos, evitando la estimación de clusters globulares. En Astronomía los clusters de estrellas o galaxias no son necesariamente globulares; por tanto, el análisis cluster resulta verdaderamente útil en ese contexto proporcionando múltiples aplicaciones (Janson y Hendry, 2007).

Müller y Sawitzi (1991) proporcionaron una metodología útil y alternativa para tratar el difícil problema de contrastar multimodalidad en dimensión uno, identificando las modas a través de elevadas concentraciones de masa de probabilidad o clusters y no como máximos locales de la función de densidad.

El análisis cluster conforma, quizás, el área a la que la estimación de conjuntos de nivel ha aportado más contribuciones. Sin embargo, existen numerosas aplicaciones en otros contextos. Por ejemplo, ciertos problemas en citometría involucran la estimación de conjuntos de nivel (Roederer y Hardy, 2001). Theiler y Cai (2003) describieron la utilidad de la estimación de conjuntos de nivel en el análisis multiespectral de imágenes y Prastawa (2003) hizo lo mismo en el campo de la medicina.

En este trabajo, comentaremos los fundamentos de cada una de las técnicas citadas. En la Sección 2, abordaremos la estimación de conjuntos de nivel a través del método plug-in, tratando el problema de selección del parámetro de suavizado a partir de la comparación de dos ventanas concretas y de la presentación detallada del algoritmo de selección automática del parámetro ventana propuesto por Samworth y Wand (2010). En la Sección 3, consideraremos el método de exceso de masa, particularizaremos el algoritmo propuesto por Hartigan para dimensión 1 y analizaremos brevemente el procedimiento de Müller y Sawitzi (1991) para testar multimodalidad. En la Sección 4, estudiaremos el método de suavizado granulométrico propuesto por Walther, planteando la generalización del mismo para densidades discontinuas. Finalmente, en la Sección 5, presentaremos los resultados obtenidos a partir de la realización de varios estudios de simulación. Primero, compararemos tres métodos plug-in considerando tres selectores de ventana bien distintos: Samworth y Wand, Seather y Jones y validación cruzada. A continuación, compararemos el método plug-in propuesto por Samworth y Wand con exceso de masa, considerando la particularización presentada para dimensión uno del algoritmo propuesto por Hartigan (1987). Finalmente, compararemos el método de suavizado granulométrico con el método plug-in de Samworth y Wand (2010).

2. Método plug-in

La aproximación plug-in es quizás la idea más directa para estimar el conjunto R_τ . Tal y como hemos comentado previamente, se basa en reemplazar la densidad f por un estimador no paramétrico de la misma, f_n . Habitualmente, y para el caso 1–dimensional, el estimador de la densidad empleado es el estimador tipo núcleo definido por

$$f_n(t) = \frac{1}{n} \sum_{i=1}^n K_h(t - X_i), \quad (1)$$

donde $K_h(z) = (1/h)K(z/h)$, $K : \mathbb{R} \rightarrow \mathbb{R}$ función de densidad simétrica fija, llamada núcleo y $h > 0$, parámetro ventana o de suavizado.

El estimador tipo núcleo de la densidad, f_n , hereda las propiedades de suavidad de la función núcleo considerada, sin jugar ésta un papel determinante en la estimación de f . Sin embargo, la elección del parámetro de suavizado conforma el principal problema en el contexto de estimación no paramétrica de la densidad. Así, valores de h demasiado grandes producen sobresuavización en la estimación y valores de h demasiado pequeños producen el efecto contrario, infrasuavización. Para conseguir que la varianza y el sesgo no sean demasiado grandes se necesita que $h \rightarrow 0$ y $nh \rightarrow \infty$.

Según lo anterior, el estimador propuesto para el conjunto de nivel, R_τ , será del tipo:

$$\hat{R}_{\tau,n} = \{x \in \mathbb{R} : f_n(x) \geq \hat{f}_\tau\}, \quad (2)$$

donde $\hat{f}_\tau = f_\tau(f_n)$ denota el estimador plug-in de f_τ . Análogamente, podría considerarse la generalización para el caso d –dimensional.

Desde un punto de vista intuitivo, resulta muy sencillo motivar la construcción del estimador propuesto en (2); sin embargo, existen dos inconvenientes fundamentales. El método plug-in, resulta poco útil si deseamos asumir ciertas restricciones a priori sobre la forma de R_τ y, al igual que sucede en el contexto de la estimación no paramétrica de la densidad, el punto fundamental volverá a ser la elección del parámetro ventana dado que no está claro como elegirlo. En general, la elección del parámetro de suavizado se centra en estimar bien f , no R_τ . Pero, ¿cómo influye h en la estimación de R_τ ? ¿Una buena ventana para estimar f es la mejor opción para estimar R_τ ? Para estudiarlo con algún detalle, analizaremos el efecto del parámetro de suavizado en el contexto de la estimación de conjuntos de nivel considerando como referencia dos ventanas concretas. La primera, h_{ISE} , cuyo cálculo se ha efectuado empleando la librería *ks* de R, surge de la minimización, para una muestra, del Error Cuadrático Integrado

$$ISE(h) = \int (f_n(t) - f(t))^2 dt,$$

La segunda ventana, h_{μ_f} , se obtiene de minimizar, para la misma muestra,

$$\mu_f(R_\tau \Delta \hat{R}_{\tau,n}) = \int_{R_\tau \Delta \hat{R}_{\tau,n}} f(t) dt,$$

donde μ denota la medida de Lebesgue. Así, mientras la ventana h_{ISE} persigue minimizar la diferencia puntual de la densidad a la estimación, la ventana h_{μ_f} obedece a criterios bien distintos, centrándose

en la estimación de R_τ y otorgando más peso en aquellas regiones en las cuales los datos tienden a ser más densos. Como modelos de prueba se considerarán las quince densidades descritas en Marron y Wand (1992) las cuales poseen la peculiaridad de poder ser escritas como mixturas de normales (ver Apéndice).

Con carácter ilustrativo, se exponen a continuación las tablas del cociente de las ventanas obtenidas para distintos valores de τ y n considerados:

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\frac{h_{ISE}}{h_{\mu_f}}$	0.61	1.15	1.57	0.41	0.76	1.11	0.79	0.94	0.91	0.81	0.92	0.59	0.7	0.93	1.32

Tabla 1: Cociente ventanas h_{ISE} y h_{μ_f} para $n = 1000$ y $\tau = 0.5$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\frac{h_{ISE}}{h_{\mu_f}}$	0.84	0.7	0.54	0.68	0.73	0.89	1.03	0.76	0.95	1.41	1.02	0.69	2.85	0.73	0.76

Tabla 2: Cociente ventanas h_{ISE} y h_{μ_f} para $n = 400$ y $\tau = 0.5$.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$\frac{h_{ISE}}{h_{\mu_f}}$	1.83	1.46	0.53	0.65	7	1.4	1.97	0.47	0.77	0.84	0.84	1.08	0.55	1.07	1.04

Tabla 3: Cociente ventanas h_{ISE} y h_{μ_f} para $n = 1000$ y $\tau = 0.2$.

Según las tablas expuestas, parece no existir relación, en términos comparativos, entre las dos ventanas consideradas. La figura siguiente muestra la fuerte influencia del parámetro τ en la comparación de ambas ventanas, representando para distintos valores de τ , el cociente considerado, $\frac{h_{ISE}}{h_{\mu_f}}$, para las densidades 1, 5 y 10 de Marron y Wand, respectivamente.

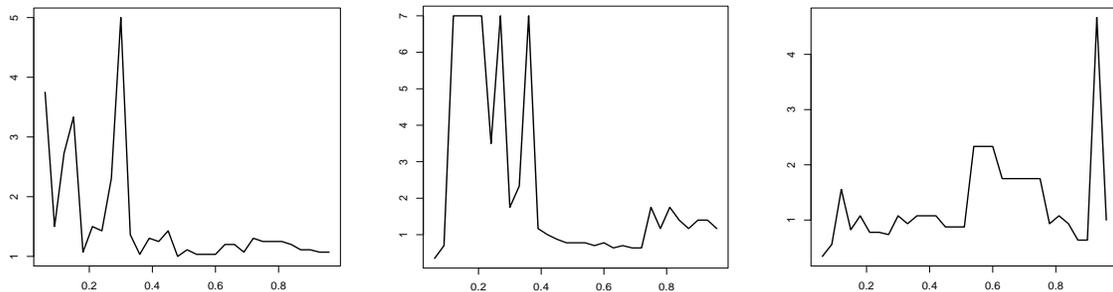
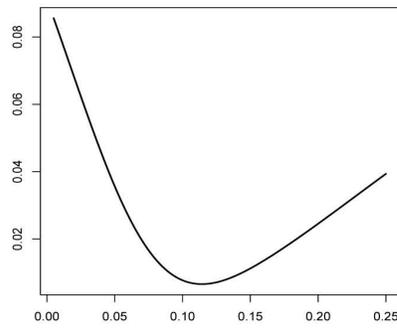
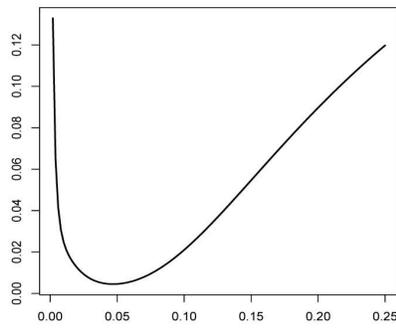
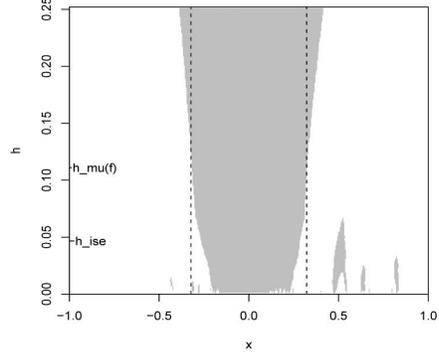
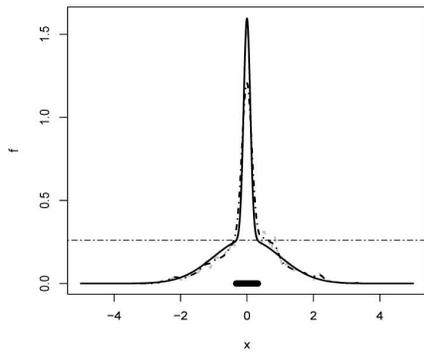
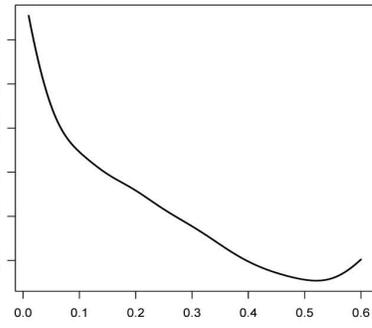
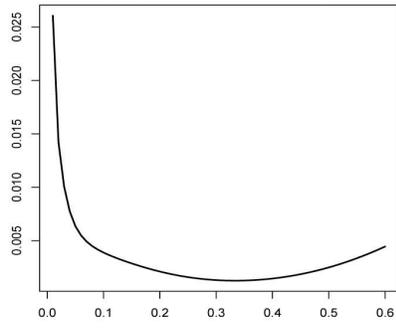
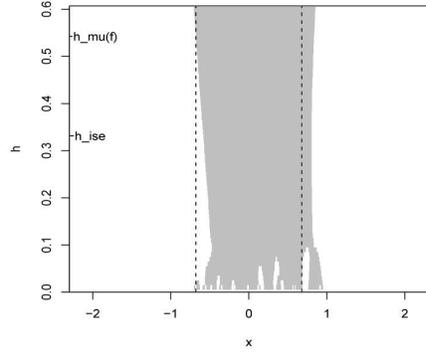
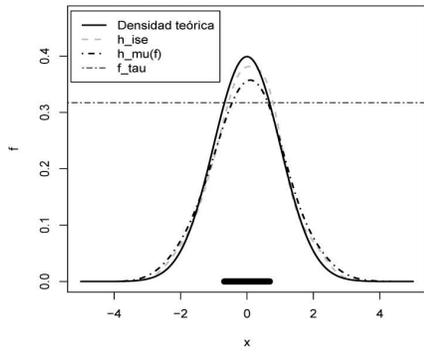


Figura 1: Cociente ventanas h_{ISE} y h_{μ_f} para distintos valores de τ y $n = 1000$.

A la vista de la Figura 1, nuevamente constatamos la ausencia de relación, incluso para una misma muestra, entre las ventanas h_{ISE} y h_{μ_f} . Por tanto, la consideración de la ventana h_{ISE} pudiera resultar desastrosa para estimar conjuntos de nivel.

Otra cuestión interesante es ver como el estimador $\hat{R}_{\tau,n}$ depende de h . ¿Cómo es la curva del error? La figura siguiente, para $\tau = 0.5$ y $n = 1000$, consta de cuatro gráficos para cada una de las densidades representadas. El primero (izquierda) muestra las estimaciones de la densidad correspondiente para las ventanas consideradas, h_{ISE} y h_{μ_f} , así como f_τ y el conjunto de nivel teórico sobre el eje de abscisas. El segundo (derecha) representa las estimaciones de los conjuntos de nivel para un rango de valores de h marcando los extremos del conjunto de nivel teórico a través de líneas discontinuas. Los dos últimos gráficos representan las curvas de error ISE (izquierda) y μ_f (derecha). Seguidamente, se realizan consideraciones similares para $\tau = 0.5$ y $n = 400$ y $\tau = 0.2$ y $n = 1000$.



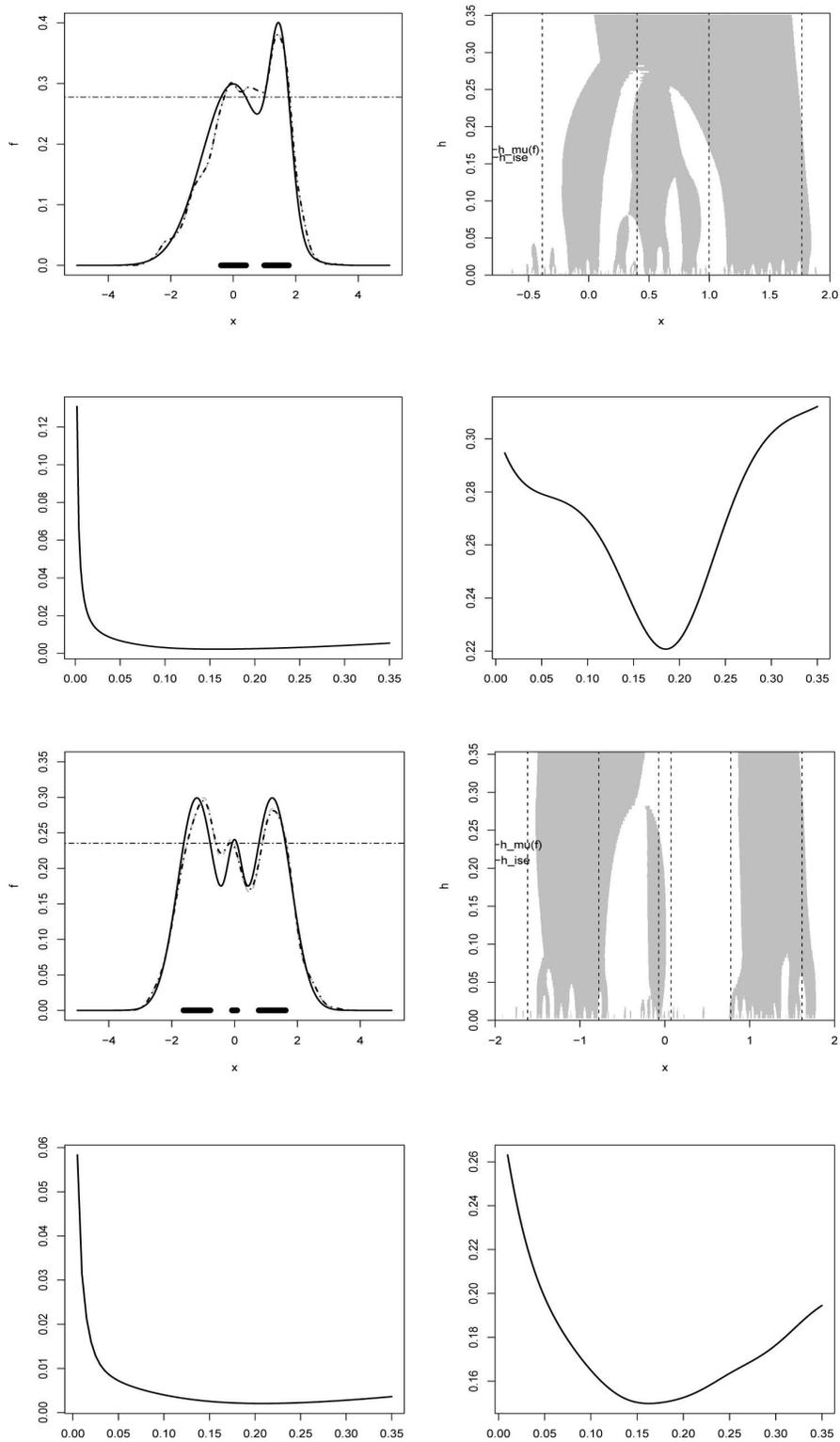


Figura 2: Estimaciones densidad y conjuntos de nivel para $\tau = 0.5$ y $n = 1000$.

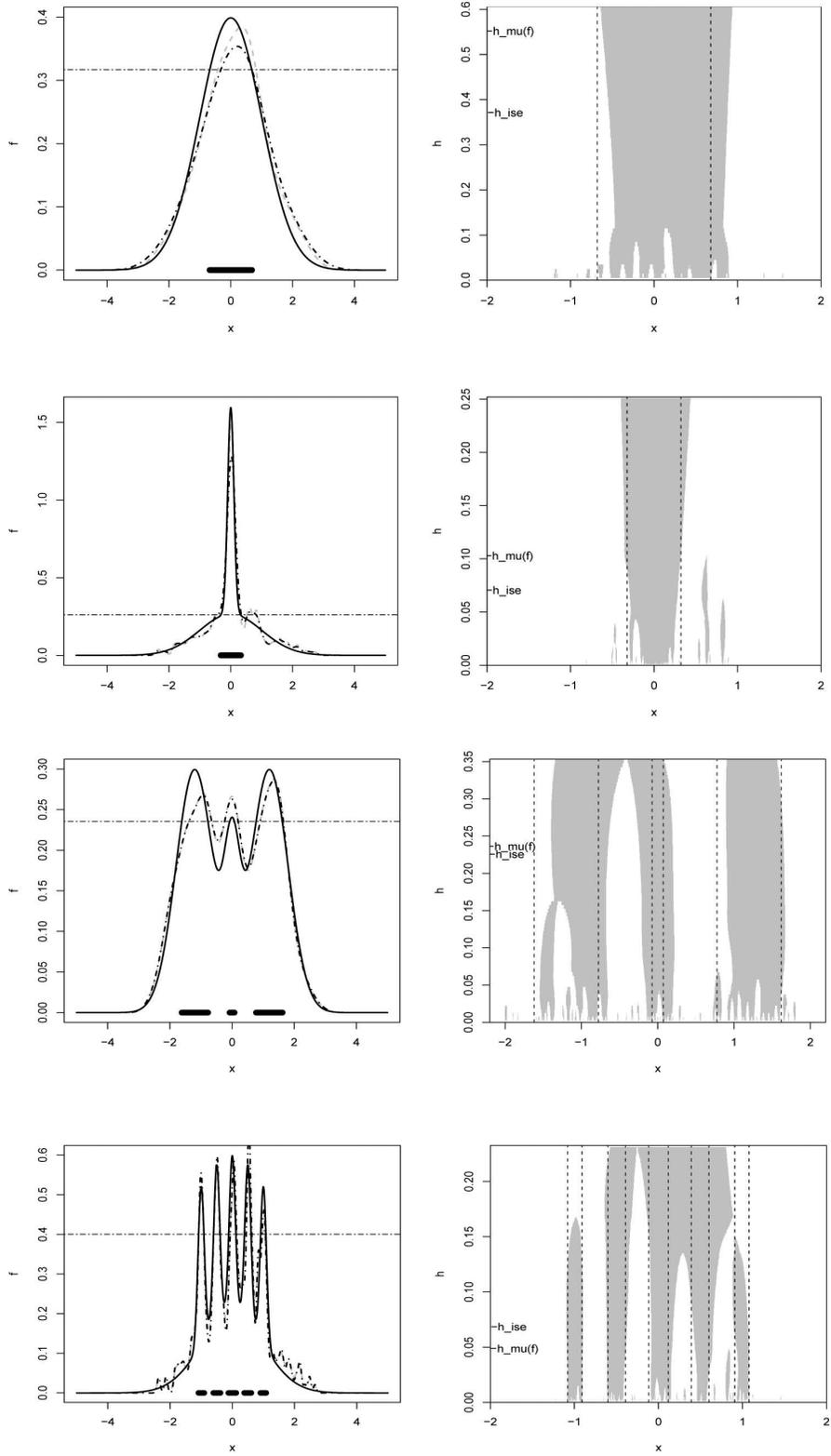


Figura 3: Estimaciones densidad y conjuntos de nivel para $\tau = 0.5$ y $n = 400$.

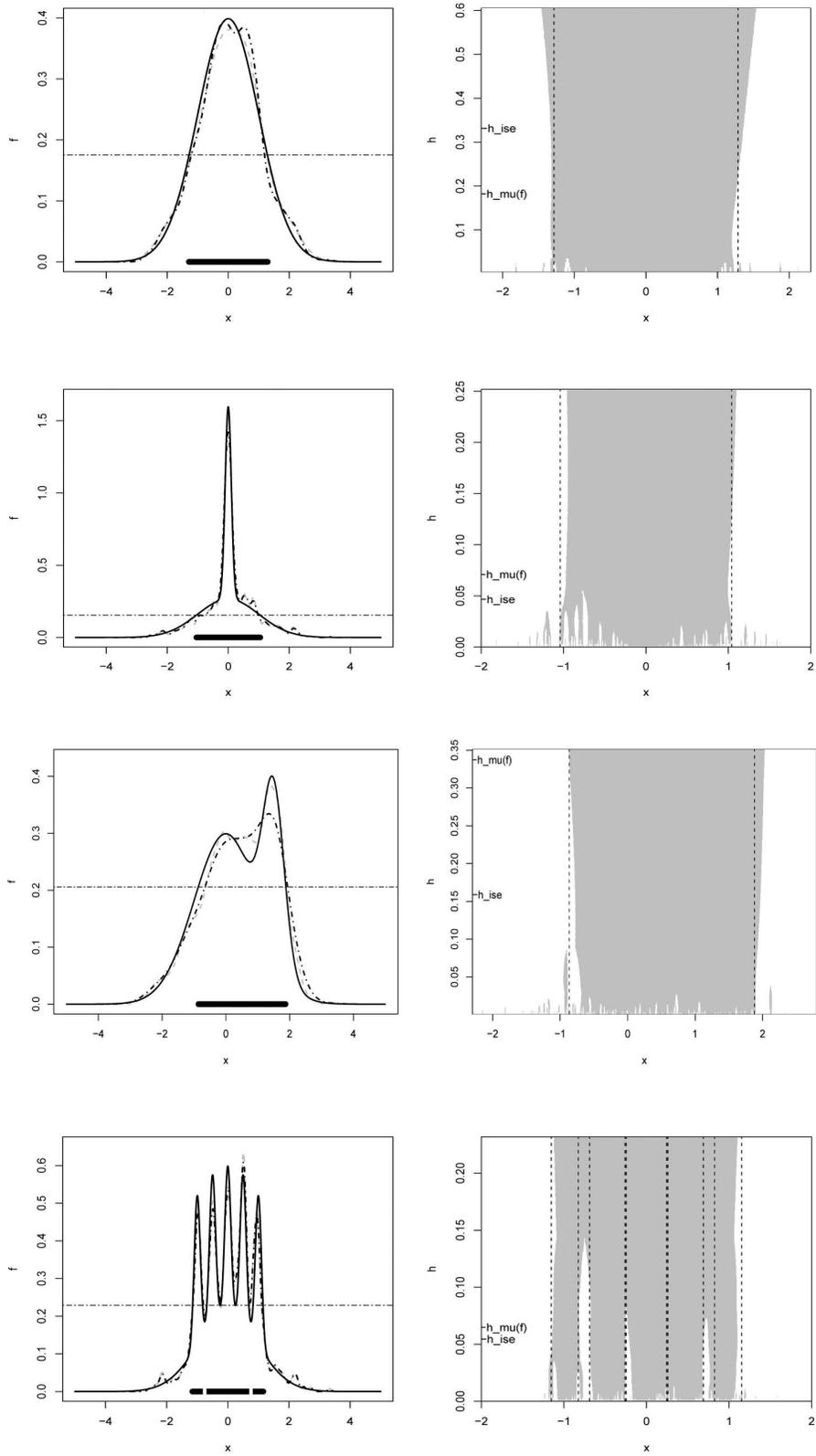


Figura 4: Estimaciones densidad y conjuntos de nivel para $\tau = 0.2$ y $n = 1000$.

La observación conjunta de las Figuras 2 y 3, realizadas considerando $\tau = 0.5$, permite analizar la influencia del tamaño muestral en la estimación de los conjuntos de nivel. Así, para $n = 400$ las estimaciones resultan bastante peores que las obtenidas para $n = 1000$. Obsérvense, por ejemplo, los resultados obtenidos para el modelo 9 de Marron y Wand.

Por otra parte, si analizamos las Figuras 2 y 4, queda constancia del relevante papel que el parámetro τ ejerce sobre en la estimación de conjuntos de nivel. Para $\tau = 0.2$, se persigue encerrar bajo la curva de la densidad considerada un área igual a 0.8; en cambio, para $\tau = 0.5$, el área considerada pasa a valer 0.5. Por tanto, $f_{0.2}$ será, en general, menor que $f_{0.5}$ reduciendo la complejidad en la estimación. Al observar las densidades con más de una moda, es fácilmente deducible que para valores de f_τ suficientemente bajos el conjunto de nivel se reduce a un intervalo. El aumento progresivo del parámetro f_τ provocaría la consideración de varios intervalos siendo precisa una buena estimación no paramétrica de las modas, lo cual complica substancialmente la situación. Por otra parte, valores de f_τ grandes exigen una buena estimación de f sólo en una región pequeña, con elevada densidad de datos, lo cual hace que criterios globales como el *ISE* resulten claramente inadecuados.

Además, un hecho relevante es observable a partir de las Figuras 2, 3 y 4. Fijado un modelo determinado, si ambas ventanas, h_{ISE} y h_{μ_f} , toman valores similares (ver Tablas 1, 2 y 3), éstas proporcionarán las estimaciones del conjunto de nivel óptimas. En caso contrario, la ventana h_{μ_f} obtendrá mejores estimaciones (obsérvese el modelo 4 de Marron y Wand en la Figura 3 o los modelos 1 y 10 en la Figura 4). En resumen, no cabe duda de que nos encontramos ante un problema distinto al de determinar el parámetro ventana para la estimación no paramétrica de la densidad. La estimación de conjuntos de nivel ha sido abordado en la literatura desde múltiples puntos de vista sin considerar, en general, la conexión con el problema de selección de la ventana que nos ocupa. Sin embargo, en 2010, Samworth y Wand desarrollaron un método selector de ventana para abordar la estimación de conjuntos de nivel a partir de la aproximación plug-in considerando una aproximación al riesgo asintótico que consideraremos con detalle en la siguiente sección.

2.1. Método plug-in. Samworth y Wand

Samworth y Wand (2010) consideraron como medida de error $\mu_f(R_\tau \Delta \hat{R}_{\tau,n})$, con μ_f definida previamente. Intuitivamente, $\mu_f(R_\tau \Delta \hat{R}_{\tau,n})$ representa la probabilidad de que una observación de f pertenezca a uno y sólo a uno de los conjuntos R_τ y $\hat{R}_{\tau,n}$. A diferencia con la medida de Lebesgue, μ_f otorga más peso en aquellas regiones en las cuales los datos tienden a ser más densos.

Asumamos,

1. f uniformemente continua en \mathbb{R} . Existe un número finito de puntos $x_1 < \dots < x_{2r}$ tales que

$$f(x_j) = f_\tau$$

para $j = 1, 2, \dots, 2r$ y existe $\delta > 0$ con f dos veces continuamente diferenciable en

$$\cup_{j=1}^r [x_{2j-1} - \delta, x_{2j} + \delta]$$

con $f'(x_{2j-1}) > 0$ y $f'(x_{2j}) < 0$ para $j=1, \dots, 2r$.

2. Sean $h^- = h_n^-$ y $h^+ = h_n^+$ sucesiones no negativas con $h^- \leq h^+$, tales que

$$n(h^-)^4 / \sqrt{\log(1/h^-)} \rightarrow \infty \text{ y } h^+ \rightarrow 0$$

cuando $n \rightarrow \infty$. Entonces, $h = h_n$ es una sucesión con $h_n^- \leq h_n \leq h_n^+$ para todo n .

3. El núcleo K es no negativo, continuamente diferenciable y además satisface:

$$\int xK(x) dx = 0, \quad \mu_2(K) \equiv \int x^2K(x) dx < \infty \text{ y } \int K'^2(x) dx < \infty.$$

La primera de las restricciones garantiza que f_τ es el único número real positivo que verifica

$$\int f(t) \mathbb{I}_{\{f(t) \geq f_\tau\}} dt = 1 - \tau.$$

Bajo las suposiciones anteriores con probabilidad próxima a uno para n grande, $\hat{f}(x)$ posee una propiedad análoga. En otras palabras, \hat{f}_τ es el único número real positivo verificando:

$$\int \hat{f}_n(t) \mathbb{I}_{\{\hat{f}_n(t) \geq \hat{f}_\tau\}} dt = 1 - \tau.$$

Samworth y Wand analizaron el comportamiento asintótico de $\mathbb{E}\{\mu_f(\hat{R}_{\tau,n}\Delta R_\tau)\}$ bajo 1, 2 y 3. Antes de detallar las conclusiones obtenidas, debemos introducir notación conveniente. Si denotamos por ϕ y Φ la función de densidad y distribución de la normal estándar, respectivamente, y $R(K) = \int K^2(t) dt$, es posible definir:

$$D_1 = \frac{1}{2}\mu_2(K) \left\{ \sum_{j=1}^{2r} \frac{1}{|f'(x_j)|} \right\}^{-1} \left[\sum_{j=1}^{2r} \frac{f''(x_j)}{|f'(x_j)|} + \frac{1}{f_\tau} \sum_{j=1}^r \{f'(x_{2j}) - f'(x_{2j-1})\} \right],$$

$$D_2 = R(K)f_\tau \left\{ \sum_{j=1}^{2r} \frac{1}{|f'(x_j)|} \right\}^{-2} \sum_{j=1}^{2r} \frac{1}{f'(x_j)^2}$$

y

$$D_{3,j} = \frac{R(K)f_\tau}{|f'(x_j)|} \left\{ \sum_{k=1}^{2r} \frac{1}{|f'(x_k)|} \right\}^{-1}, \quad j = 1, 2, \dots, 2r.$$

Ahora, estamos en condiciones de enunciar el teorema de Samworth y Wand que será fundamental para proponer un método automático de selección del parámetro ventana en el contexto de la estimación plug-in de los conjuntos de nivel.

Teorema 2.1 *Bajo las condiciones de regularidad establecidas previamente,*

$$\mathbb{E}\{\mu_f(\hat{R}_{\tau,n}\Delta R_\tau)\} = \sum_{j=1}^{2r} \left[\frac{B_{1,j}\phi(B_{2,j}n^{1/2}h^{5/2})}{(nh)^{1/2}} + B_{3,j}h^2\{2\Phi(B_{2,j}n^{1/2}h^{5/2}) - 1\} \right] + o\left(\frac{1}{(nh)^{1/2}} + h^2\right)$$

cuando $n \rightarrow \infty$, uniformemente para $h \in [h^-, h^+]$ donde

$$B_{1,j} = 2f_\tau \frac{\{R(K)f_\tau - 2D_{3,j} + D_2\}^{1/2}}{|f'(x_j)|}, \quad B_{2,j} = \frac{|1/2\mu_2(K)f''(x_j) - D_1|}{\{R(K)f_\tau - 2D_{3,j} + D_2\}^{1/2}}$$

y

$$B_{3,j} = f_\tau \frac{|1/2\mu_2(K)f''(x_j) - D_1|}{|f'(x_j)|}.$$

El siguiente corolario conforma el primer paso para encontrar la ventana óptima citada. Impone condiciones, relativamente restrictivas, para garantizar que existe un único mínimo en la expresión asintótica del error.

Corolario 2.2 *Bajo las restricciones primera y tercera impuestas previamente, considerando $r = 1$ y f simétrica sobre algún punto de la recta real. Entonces, existe un único $c_{opt} \in (0, \infty)$ dependiendo de f y de K pero no de n , tal que cualquier sucesión de ventanas h_{opt} que minimiza $\mathbb{E}\{\mu_f(R_{\tau,n}\Delta R_\tau)\}$ verifica*

$$h_{opt} = c_{opt}n^{-1/5}\{1 + o(1)\}$$

cuando $n \rightarrow \infty$.

Aproximación numérica riesgo.

El Teorema 2.1 garantiza la siguiente aproximación para el riesgo asintótico:

$$\mathbb{E}\{\mu_f(\hat{R}_{\tau,n}\Delta R_\tau)\} \simeq \sum_{j=1}^{2r} \left[\frac{B_{1,j}\phi(B_{2,j}n^{1/2}h^{5/2})}{(nh)^{1/2}} + B_{3,j}h^2\{2\Phi(B_{2,j}n^{1/2}h^{5/2}) - 1\} \right]. \quad (3)$$

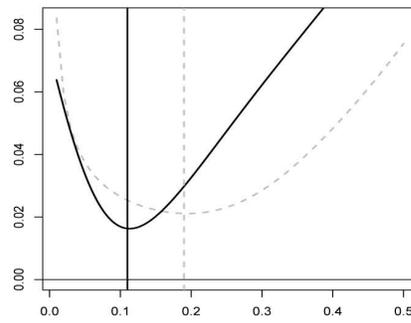
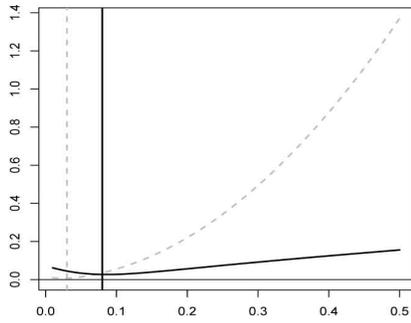
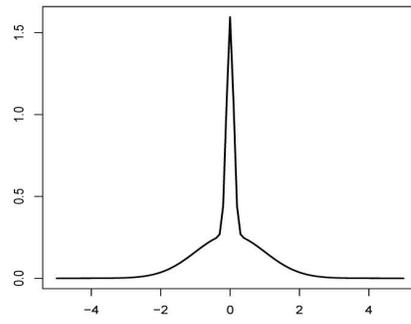
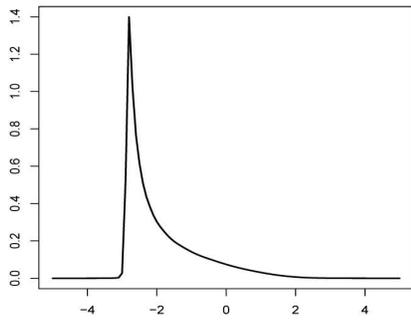
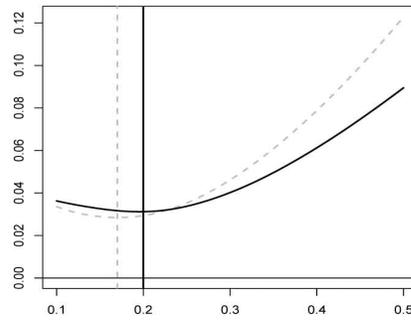
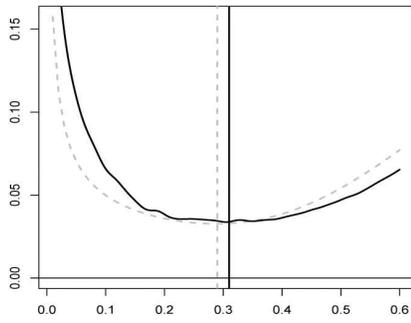
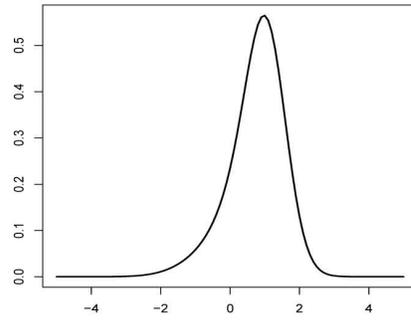
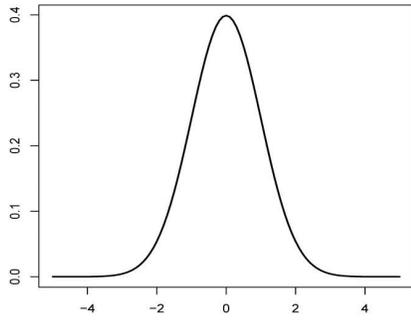
El lado derecho de la expresión (3) resultará de gran utilidad en los procedimientos de selección de ventana ya que se podrá estimar a partir de una muestra. Para poder usarla, debemos garantizar que resulta una buena aproximación al riesgo.

En general, dadas f , h y τ resulta complicado calcular el riesgo de forma exacta, $\mathbb{E}\{\mu_f(R_{\tau,n}\Delta R_\tau)\}$, de manera exacta. Por tanto, será necesario considerar una aproximación Monte Carlo,

$$\frac{1}{M} \sum_{i=1}^M \mu_f(\hat{R}_{\tau,n}^{[i]}\Delta R_\tau), \quad (4)$$

donde $\hat{R}_{\tau,n}^{[1]}, \dots, \hat{R}_{\tau,n}^{[M]}$ son M realizaciones de $\hat{R}_{\tau,n}$.

A continuación, exponemos el error asintótico (línea discontinua) y la aproximación Monte Carlo del riesgo exacto (línea continua) considerando como modelos de prueba las densidades de Marron y Wand, $M = 500$ muestras de tamaño $n = 1000$, $\tau = 0.5$ y $\tau = 0.8$.



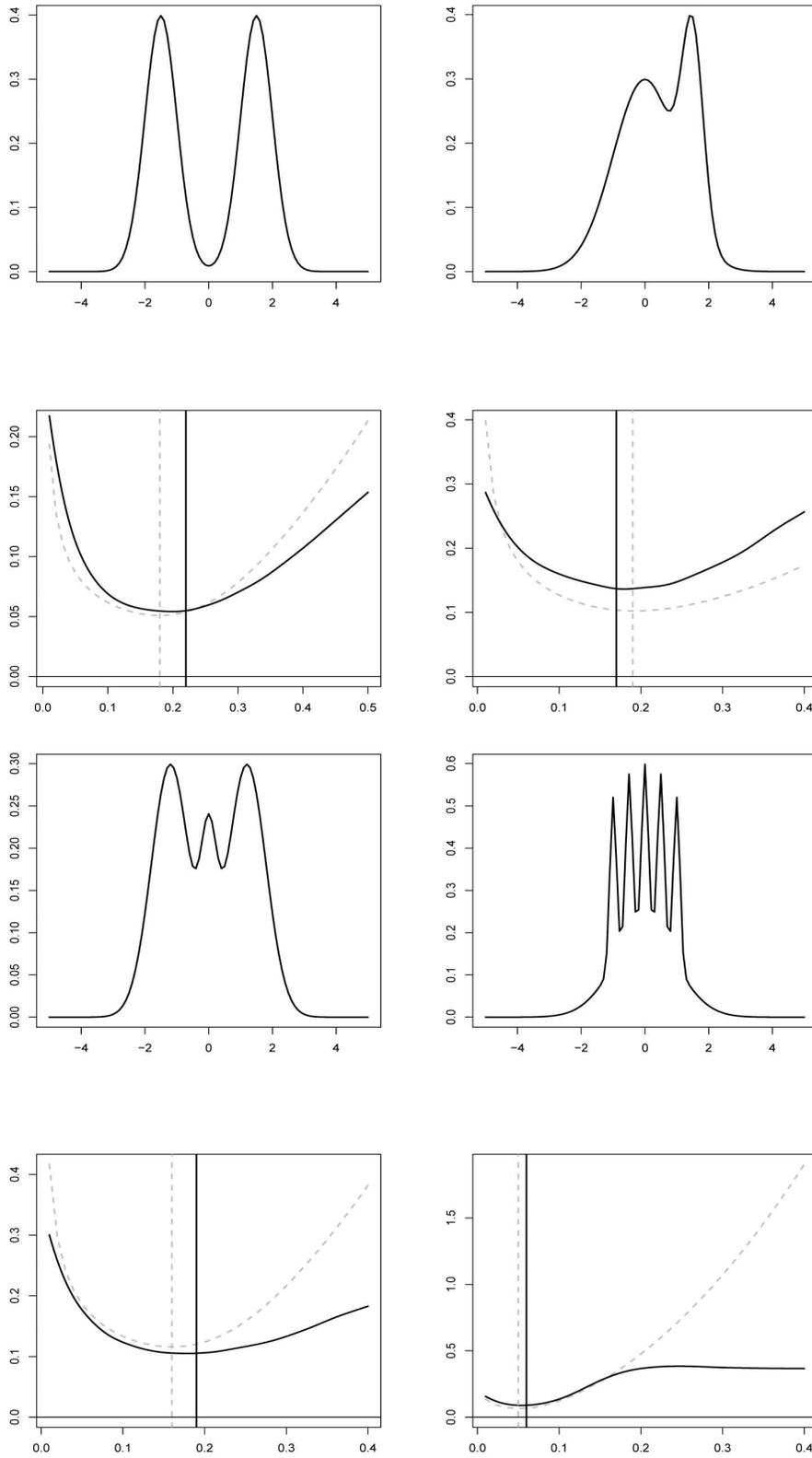


Figura 5: Error asintótico y aproximación Monte Carlo $n = 1000$, $M = 500$ y $\tau = 0.5$.

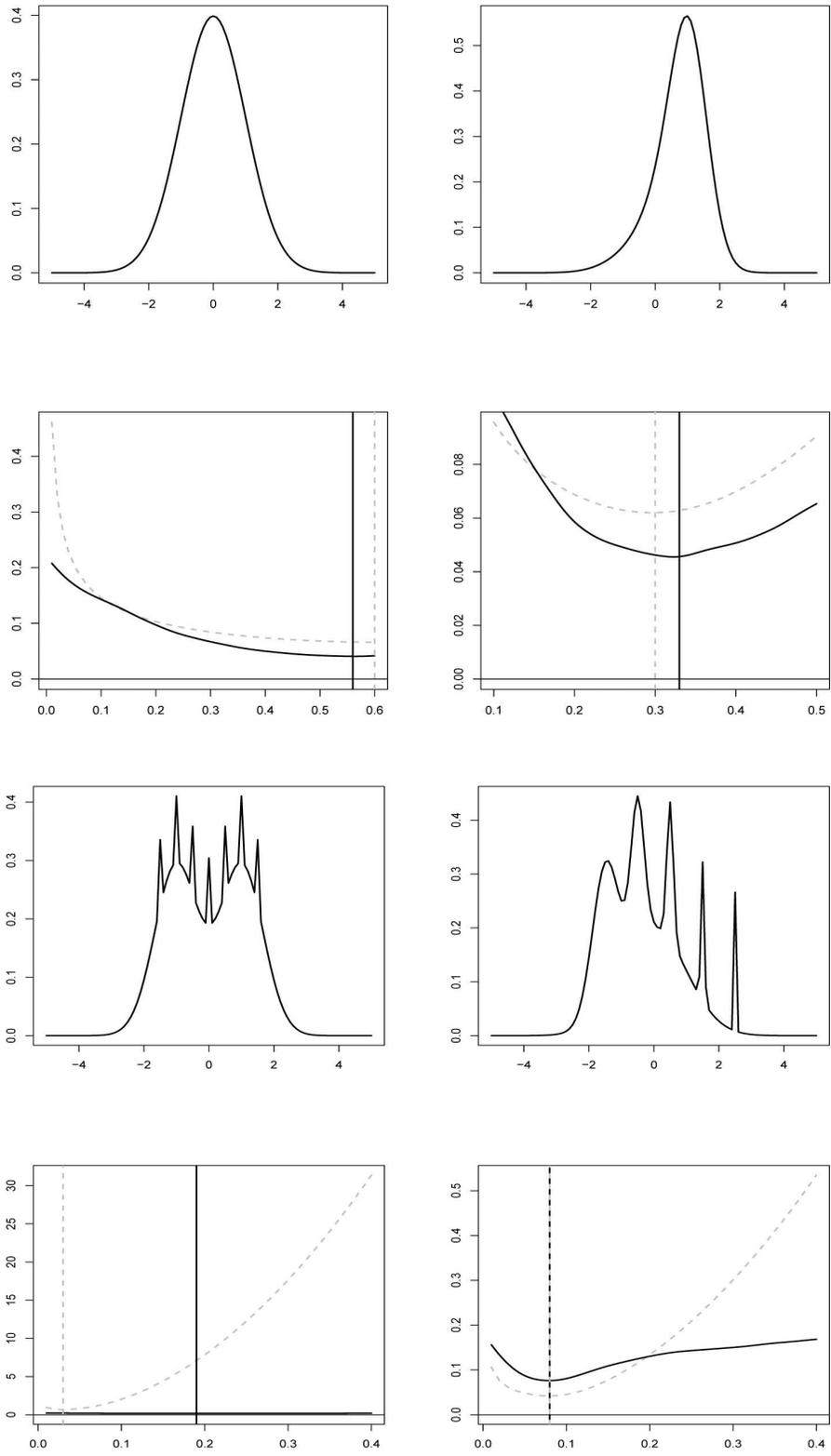


Figura 6: Error asintótico y aproximación Monte Carlo $n = 1000$, $M = 500$ y $\tau = 0.8$.

Según la Figura 5, para $\tau = 0.5$, el error asintótico se aproxima relativamente bien al exacto (estimado por Monte Carlo) en los modelos 1 y 2 de Marron y Wand. Idénticas conclusiones pueden ser extraídas para los modelos 6 y 7, caracterizados por su bimodalidad. En cuanto al modelo 10, al menos en el entorno de interés, podemos garantizar una buena aproximación. Los modelos 3, 4 y 5, a pesar de ser unimodales, presentan al error asintótico como una mala aproximación del exacto. Concretamente, la densidad 4 de Marron y Wand, para nivel $f_{0.5}$, proporciona valores elevados para $|f''|$ debido a la rápida transición que la curva sufre en ese entorno. A partir de los resultados obtenidos, no resulta sencillo determinar si el efecto del parámetro τ juega un papel determinante en la aproximación del error.

Del mismo modo, que hemos aproximado por Monte Carlo $\mathbb{E}\{\mu_f(\hat{R}_{\tau,n} \triangle R_\tau)\}$, es posible considerar la aproximación de otras medidas de riesgo, $\mathbb{E}\{d_H(\hat{R}_{\tau,n}, R_\tau)\}$ y $\mathbb{E}\{\mu(\hat{R}_{\tau,n} \triangle R_\tau)\}$ donde d_H denota la distancia Hausdorff y μ la medida de Lebesgue, que al considerar el caso 1-dimensional, coincide con la longitud de la diferencia simétrica. A continuación, se muestran los resultados obtenidos para las aproximaciones Monte Carlo realizadas para ambos criterios de error: $\mathbb{E}\{d_H(\hat{R}_{\tau,n}, \hat{R}_\tau)\}$ (línea continua) y $\mathbb{E}\{\mu(\hat{R}_{\tau,n} \triangle R_\tau)\}$ (línea discontinua). Para el cálculo de $\mathbb{E}\{\mu(\hat{R}_{\tau,n} \triangle R_\tau)\}$ se ha tenido en cuenta que

$$\mu(\hat{R}_{\tau,n} \triangle R_\tau) = \mu(\hat{R}_{\tau,n}) + \mu(R_\tau) - 2\mu(\hat{R}_{\tau,n} \cap R_\tau).$$

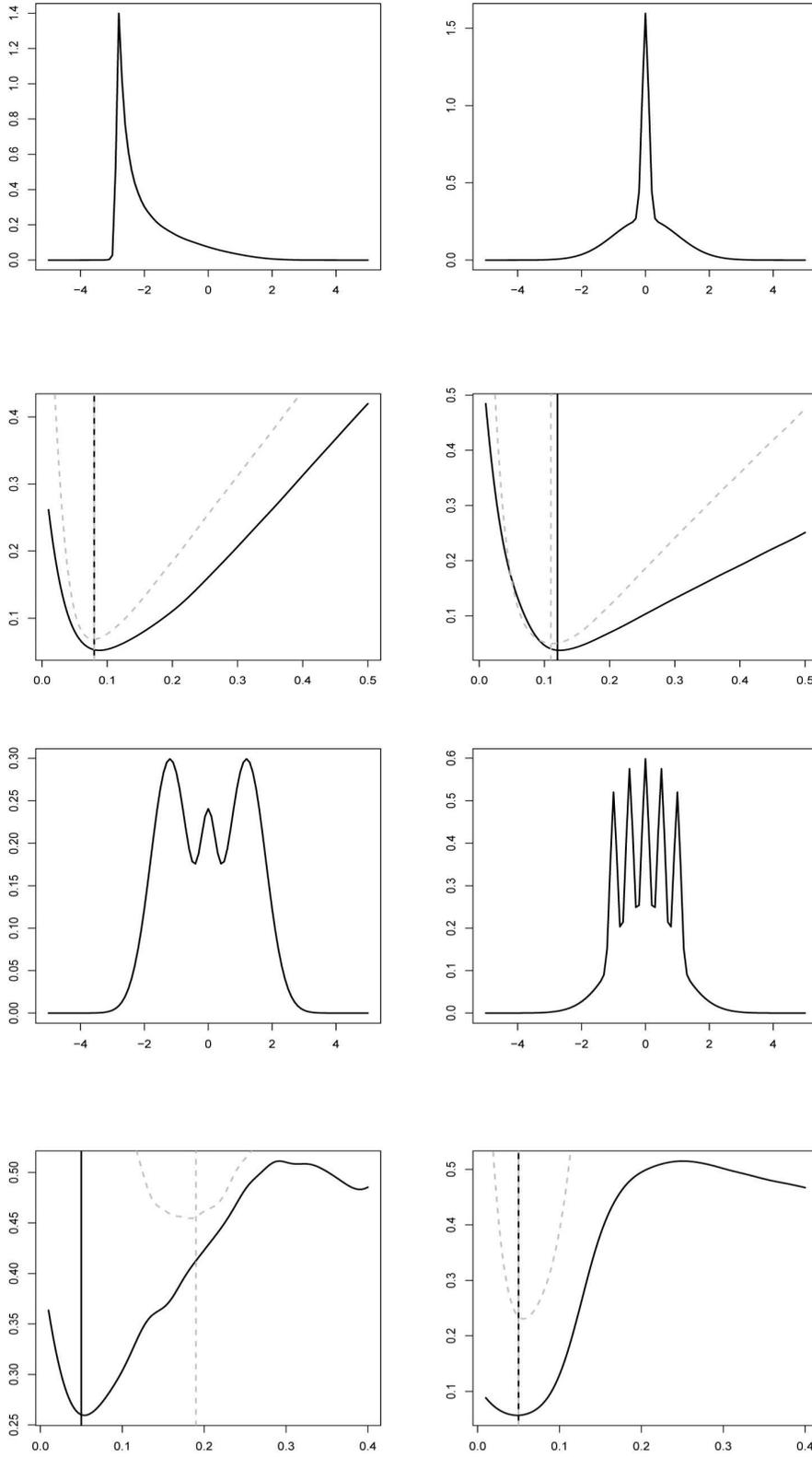


Figura 7: Aproximación Monte Carlo $n = 1000$, $M = 500$ y $\tau = 0.5$.

Selección ventana.

El Corolario (2.2) permite asumir que existe un único

$$c_{opt} \in (0, \infty) \text{ tal que } h_{opt} = c_{opt}n^{-1/5}\{1 + o(1)\}.$$

En este caso, c_{opt} minimiza el riesgo asintótico dado por

$$AR(c) = \frac{1}{n^{2/5}} \sum_{i=1}^{2r} \left[\frac{B_{1,j}}{c^{1/2}} \phi(B_{2,j}c^{5/2}) + B_{3,j}c^2 \{2\Phi(B_{2,j}c^{5/2}) - 1\} \right].$$

En la práctica, para encontrar un selector de ventana, resulta necesario obtener un estimador \hat{c}_{opt} de c_{opt} . El modo natural de construir tal estimador pasa por considerar estimadores \hat{D}_1, \hat{D}_2 y $\hat{D}_{3,j}$ de D_1, D_2 y $D_{3,j}$, respectivamente para obtener estimadores plug-in $\hat{B}_{1,j}, \hat{B}_{2,j}$ y $\hat{B}_{3,j}$ de $B_{1,j}, B_{2,j}$ y $B_{3,j}$. Entonces,

$$\hat{c}_{opt} = \arg \min_{c \in (0, \infty)} \hat{AR}_n(c),$$

donde

$$\hat{AR}_n(c) = \frac{1}{n^{2/5}} \sum_{i=1}^{2r} \left[\frac{\hat{B}_{1,j}}{c^{1/2}} \phi(\hat{B}_{2,j}c^{5/2}) + \hat{B}_{3,j}c^2 \{2\Phi(\hat{B}_{2,j}c^{5/2}) - 1\} \right].$$

Con probabilidad uno, la solución a este problema de minimización será única para n grande pudiendo ser aproximada numéricamente. Así, la ventana final,

$$\hat{h}_\tau = \hat{c}_{opt}n^{-1/5}.$$

Para construir los estimadores \hat{D}_1, \hat{D}_2 y $\hat{D}_{3,j}$ deben considerarse estimadores plug-in de $f_\tau, f'(x_j)$ y $f''(x_j)$ para $j = 1, \dots, 2r$. Asumiendo núcleo K suave podemos construir estimadores núcleo $f_{n,h_0}(\hat{x}_{j,h_0}), f'_{n,h_1}(\hat{x}_{j,h_0})$ y $f''_{n,h_2}(\hat{x}_{j,h_0})$ de $f_\tau, f'(x_j)$ y $f''(x_j)$, respectivamente, donde \hat{x}_{j,h_0} es un estimador de x_j y $h_i, i = 0, 1, 2$ denota las diferentes ventanas que pueden ser usadas para estimar f y sus derivadas.

La estimación de $x_j, j = 1, \dots, 2r$ resulta sencilla. Una vez estimada f no paramétricamente a través de f_{n,h_0} y fijado j, \hat{x}_{j,h_0} puede calcularse resolviendo la ecuación dada por $f_{n,h_0} - \hat{f}_\tau$, donde \hat{f}_τ denota el estimador de f_τ obtenido a partir de la muestra.

El siguiente teorema garantiza que el procedimiento es consistente.

Teorema 2.3 *Bajo condiciones de regularidad (Samworth y Wand, 2010), asumiendo que c_{opt} es único y $AR''(c_{opt}) > 0$,*

$$\frac{\hat{h}_\tau}{h_{opt}} = 1 + O_P(n^{-2/9})$$

cuando $n \rightarrow \infty$. Además, dado que $\hat{h}_\tau = \hat{c}_{opt}n^{-1/5}$,

$$\frac{\hat{AR}_n(\hat{c}_{opt})}{AR(c_{opt})} = 1 + O_P(n^{-2/9}).$$

Selector de ventana.

En la práctica, las ventanas piloto h_0 , h_1 y h_2 son estimadas usando estrategias plug-in con dos niveles de estimación tipo núcleo. Para h_0 , el procedimiento es similar al propuesto por Park y Marron (1990) y Sheather y Jones (1991), involucrando funcionales de la forma:

$$\psi_r = \int_{-\infty}^{\infty} f^{(r)}(x) dx.$$

Los estimadores tipo núcleo para ψ_r ,

$$\hat{\psi}_r(g) = n^{-2} g^{-r-1} \sum_{i=1}^n \sum_{j=1}^n L^{(r)}\{(X_i - X_j)/g\},$$

donde L representa una función núcleo suficientemente suave y $g > 0$ un parámetro ventana. Para comenzar con el algoritmo, se asume f es una densidad $N(\mu, \sigma^2)$. Por tanto, partimos de

$$\hat{\psi}_r^{NS} = \frac{(-1)^{r/2} r!}{(2\hat{\sigma})^{r+1} (r/2)! \pi^{1/2}}.$$

Algoritmo completo. Tomemos $L = K = \phi$,

1. Como argumentos de entrada consideramos $\mathcal{X}_n = \{X_1, \dots, X_n\}$ y el parámetro $0 < \tau < 1$.
2. Sea $\hat{\sigma} = \text{mín}(\text{desviación estándar muestral}, \text{rango intercuartílico muestral} / 1.349)$.
3. Estimación ψ_8 , ψ_{10} y ψ_{12} como $\hat{\psi}_8^{NS}$, $\hat{\psi}_{10}^{NS}$ y $\hat{\psi}_{12}^{NS}$.
4. Estimación ψ_6 , ψ_8 y ψ_{10} usando estimaciones tipo núcleo $\hat{\psi}_6(g_{0,1})$, $\hat{\psi}_8(g_{1,1})$ y $\hat{\psi}_{10}(g_{1,1})$ donde

$$g_{0,1} = \{30/(\hat{\psi}_8^{NS} n)\}^{1/9}, \quad g_{1,1} = \{-210/(\hat{\psi}_{10}^{NS} n)\}^{1/11} \quad \text{y} \quad g_{1,2} = \{1890/(\hat{\psi}_{12}^{NS} n)\}^{1/13}.$$

5. Estimación ψ_4 , ψ_6 y ψ_8 usando estimaciones tipo núcleo $\hat{\psi}_4(g_{0,2})$, $\hat{\psi}_6(g_{1,2})$ y $\hat{\psi}_8(g_{2,2})$ donde

$$g_{0,2} = \{6/(\hat{\psi}_8(g_{0,1}) n)\}^{1/7}, \quad g_{1,2} = \{-30/(\hat{\psi}_{10}(g_{1,1}) n)\}^{1/9} \quad \text{y} \quad g_{2,2} = \{210/(\hat{\psi}_{12}(g_{1,2}) n)\}^{1/11}.$$

6. Obtención ventanas plug-in $h^{(r)}$ para estimar $f^{(r)}$. Las expresiones explícitas,

$$\hat{h}_0 = [1/\{2\pi^{1/2} \hat{\psi}_4(g_{0,2}) n\}]^{1/5}, \quad \hat{h}_1 = [-3/\{4\pi^{1/2} \hat{\psi}_6(g_{1,2}) n\}]^{1/7} \quad \text{y} \quad \hat{h}_2 = [15/\{8\pi^{1/2} \hat{\psi}_8(g_{2,2}) n\}]^{1/9}.$$

7. Obtención estimaciones piloto de f , f' y f'' empleando núcleo gaussiano y las ventanas \hat{h}_0 , \hat{h}_1 y \hat{h}_2 .
8. Emplear estimación no paramétrica de f con ventana \hat{h}_0 para estimar f_τ , r y x_i , $i = 1, \dots, 2r$.
9. Obtención estimadores $\hat{B}_{1,j}$, $\hat{B}_{2,j}$ y $\hat{B}_{3,j}$.

10. La ventana seleccionada para realizar la estimación de la densidad con núcleo gaussiano,

$$\hat{h}_\tau = \hat{c}_{opt} n^{-1/5}$$

con

$$\hat{c}_{opt} = \arg \min_{c \in (0, \infty)} \hat{A}R_n(c).$$

El objetivo principal de este trabajo es programar este método y comprobar su comportamiento comparándolo con las alternativas existentes en la literatura. En la Sección 5.1 compararemos Samworth y Wand con validación cruzada y Seather y Jones. En la Sección 5.2, compararemos Samworth y Wand con exceso de masa. Finalmente, en la Sección 5.3, compararemos Samworth y Wand con el método de suavizado granulométrico.

3. Método exceso de masa

El método de exceso de masa, propuesto de forma independiente por Hartigan (1987) y por Müller y Sawitzki (1987), conforma una alternativa competitiva para abordar la estimación de R_τ en el caso d -dimensional. Esta aproximación se basa en que el funcional

$$H_\tau(B) = \int_B (f(x) - f_\tau) dx = P(B) - f_\tau \mu(B),$$

es maximizado en el conjunto de nivel, donde μ denota la medida de Lebesgue y P una distribución sobre \mathbb{R}^d .

Entonces, si \mathfrak{B} denota una clase de conjuntos dada, un estimador natural del conjunto de nivel, bajo la restricción de forma de que debe pertenecer a \mathfrak{B} , sería el máximo en \mathfrak{B} del exceso de masa empírico

$$H_{\tau,n}(B) = P_n(B) - f_\tau \mu(B). \quad (5)$$

El método exceso de masa proporciona un modo natural para incorporar restricciones de forma en el problema que nos ocupa. De hecho, estas restricciones son requeridas para garantizar la aplicabilidad del método ya que, si no se realiza ningún tipo de asunción sobre \mathfrak{B} , el máximo de (5) sería alcanzado en \mathcal{X}_n . Inicialmente, Hartigan consideró \mathfrak{B} la clase de conjuntos convexos y compactos en \mathbb{R}^2 teniendo en cuenta que si P está determinada por n puntos los únicos posibles vértices del conjunto a estimar son los átomos de P .

En el caso unidimensional, bastaría restringir el estudio al conjunto de intervalos de la forma

$$[X_i, X_j], \quad X_i < X_j, \quad i, j = 1, \dots, n$$

como posibles candidatos a maximizar $(P_n - \lambda)S$. En efecto, fijando $X_{(0)} = -\infty$, $X_{(n+1)} = \infty$ y denotando la muestra ordenada por

$$X_{(0)}, X_{(1)}, \dots, X_{(n)}, X_{(n+1)},$$

consideraremos el intervalo $[a, b]$ verificando para $i = 1, \dots, n+1$, $j = 0, \dots, n$ e $i \leq j$,

$$X_{(i-1)} < a < X_{(i)},$$

y

$$X_{(j)} < b < X_{(j+1)}.$$

Bajo la notación anterior,

$$P_n[X_{(i)}, X_{(j)}] = F_n(X_{(j)}) - F_n(X_{(i-1)}) = \frac{j - i + 1}{n}$$

y

$$P_n[a, b] = F_n(X_{(j)}) - F_n(X_{(i-1)}) = \frac{j - i + 1}{n},$$

donde F_n denota la función de distribución empírica asociada a la muestra de partida. Así,

$$H_{\tau,n}([X_{(i)}, X_{(j)}]) = \frac{j-i+1}{n} - f_{\tau}(X_{(j)} - X_{(i)}),$$

$$H_{\tau,n}([a, b]) = \frac{j-i+1}{n} - f_{\tau}(b-a).$$

Además,

$$[X_{(i)}, X_{(j)}] \subset [a, b] \Rightarrow (X_{(j)} - X_{(i)}) \leq (b-a).$$

Por tanto,

$$H_{\tau,n}([X_{(i)}, X_{(j)}]) \geq H_{\tau,n}([a, b]).$$

Müller y Sawitzki (1991) generalizaron el algoritmo anterior, para dimensión uno, demostrando que la función $f_{\tau} \mapsto E(f_{\tau})$ donde $E(f_{\tau}) = \sup_B \{H_{\tau}(B)\}$ puede resultar útil á la hora de proporcionar una metodología para abordar el difícil problema de contrastar multimodalidad en dimensión uno distinguiendo dos problemas básicos: localizar y estimar el número de modas.

Desde un punto de vista analítico, una moda no es más que un máximo local de la correspondiente densidad. Así, una distribución determinada pudiera tener una moda en un punto y otorgar probabilidad baja a ciertas vecindades que lo contienen. Müller y Sawitzki proporcionan una definición alternativa: una moda se presentará donde se concentre un exceso de masa de probabilidad.

Bajo el contexto anterior, las componentes conexas de R_{τ} se denominarán f_{τ} -clusters. Además, a medida que f_{τ} aumente, los f_{τ} -clusters se concentrarán sobre las modas (máximos locales de la densidad). Por tanto, es posible reemplazar la noción usual de moda por el concepto intrínseco de f_{τ} -cluster, definido como una componente conexa de R_{τ} .

Si una familia tuviese exactamente m f_{τ} -clusters el funcional exceso de masa podría reescribirse como:

$$E(f_{\tau}) = \sup \sum \int_{C_j} (f(x) - f_{\tau}) dx,$$

donde el supremo es calculado todas las familias $\{C_j : j = 1, \dots, m\}$ de conjuntos conexos disjuntos. En general, se define,

$$E_m(f_{\tau}) = \sup \sum_{j=1}^m \int_{C_j} (f(x) - f_{\tau}) dx.$$

Por tanto, podemos escribir,

$$E_m(f_{\tau}) = \sup \sum_{j=1}^m H_{\tau}(C_j).$$

Desde el punto de vista empírico, dada una muestra \mathcal{X}_n de observaciones independientes, podemos considerar

$$E_{n,M}(f_{\tau}) = \sup \sum_{j=1, \dots, M} H_{\tau,n}(C_j) \tag{6}$$

donde M se asume como el máximo número de modas.

En general, el estadístico de contraste empleado para testar multimodalidad puede definirse como

$$\Delta_{n,M} = \max_{f_\tau} D_{n,M}(f_\tau)$$

donde

$$D_{n,M}(f_\tau) = E_{n,M}(f_\tau) - E_{n,1}(f_\tau).$$

La interpretación resulta bastante sencilla, pues valores grandes de $D_{n,M}(f_\tau)$ denotan una violación de la hipótesis de unimodalidad.

Las componentes C_j maximizando (6) de denoniman f_τ -clusters empíricos. En el caso 1-dimensional, si son no vacíos, se corresponden con intervalos cerrados de extremos puntos muestrales los cuales estimarán consistentemente los f_τ -clusters.

El algoritmo propuesto por Müller y Sawitzi estima el exceso de masa y encuentra los f_τ -clusters empíricos, $C_{n,1}^M, \dots, C_{n,M}^M$ ($M \leq n$).

1. Para $M = 1$, buscaremos los índices de orden i y j que maximicen $H_{\tau,n}[X_{(i)}, X_{(j)}]$ considerando entonces $C_{n,1}^1 = [X_{(i)}, X_{(j)}]$.
2. Para abordar el paso de M a $M + 1$, se usa la siguiente relación:
Si $C_{n,1}^M, \dots, C_{n,M}^M$ son intervalos disjuntos entonces

$$E_{n,M}(f_\tau) = \sum_j H_{\tau,n}(C_{n,j}^M) \text{ y } E_{n,M+1}(f_\tau) = \sum_j H_{\tau,n}(C_{n,j}^{M+1})$$

para un conjunto de intervalos disjuntos $C_{n,1}^{M+1}, \dots, C_{n,M+1}^{M+1}$ que satisfacen

$$\{C_{n,1}^{M+1}, \dots, C_{n,M+1}^{M+1}\} = \{C_{n,1}^M, \dots, C_{n,M}^M\} \cup \{C\}$$

con $C \cap C_{n,j}^M = \emptyset$ para todo j , o

$$\{C_{n,1}^{M+1}, \dots, C_{n,M+1}^{M+1}\} = \{C_{n,1}^M, \dots, C_{n,i-1}^M, C_{n,i}^M, \dots, C_{n,M}^M\} \cup \{C', C''\}$$

con C' y C'' disjuntos con la envoltura convexa $C_{n,i}^M$ para algún i .

Entonces, para realizar el paso de M a $M + 1$ debe ser encontrado un intervalo C con $C \cap C_{n,j}^M$ para todo j maximizando $H_{f_\tau,n}$ y un intervalo \bar{C} con $\bar{C} \subset C_{n,i}^M$ para un cierto i maximizando $-H_{f_\tau,n}$. Si $H_{f_\tau,n}(C) > |H_{f_\tau,n}(\bar{C})|$ el nuevo intervalo C es añadido; en caso contrario, $C_{n,i}^M$ es reemplazado por los dos intervalos que conforman el complementario $C_{n,i}^M \setminus \bar{C}$.

3. C será encontrado realizando una búsqueda exhaustiva restringiéndose a los intervalos que forman $(\cup C_{n,j}^M)^c$ y \bar{C} , por búsqueda en los intervalos $C_{n,i}^M$.

4. Método suavizado granulométrico

El método de suavizado granulométrico fue propuesto por Walther en 1997 con una fuerte componente geométrica y buenas propiedades estadísticas y computacionales. A continuación, recordaremos la definición de la adición y diferencia de conjuntos, herramientas básicas en la construcción del estimador:

Definición 4.1 Sean $A, C \subset \mathbb{R}^d$ y $\lambda \in \mathbb{R}$, escribimos $\lambda A = \{\lambda a : a \in A\}$ y denotaremos por

$$A \oplus C = \{a + c : a \in A, c \in C\}$$

la adición de Minkowski, y por

$$A \ominus C = \{x : x + C \subset A\}$$

la diferencia de Minkowski, donde escribimos $x + C$ para $\{x\} \oplus C$.

Bajo este contexto, asumiremos que el conjunto de nivel a estimar R_τ satisface la propiedad de rodamiento. Así, un conjunto cerrado se dice que cumple la propiedad de rodamiento libre si una bola puede rodar libremente por su borde, desde el interior, manteniendo contacto con todos los puntos del borde.

Propiedad 4.1: Un conjunto cerrado R se dice que verifica la propiedad de rodamiento libre (interior) si existe $r > 0$ tal que para todo $x \in \partial R$ existe una bola $B(a, r)$ tal que $x \in B(a, r)$ y $B(a, r) \subset R$. Análogamente, podría definirse la propiedad de rodamiento libre exterior.

La condición anterior asume suavidad de la frontera de R y la establece en términos geométricos pudiendo ser interpretada en términos de diferenciabilidad. Además guarda gran relación con la propiedad de r -convexidad:

Definición 4.2 Un conjunto A se dice r -convexo si coincide con su envoltura r -convexa,

$$C_r(A) = \bigcap_{B_r(x) \cap A = \emptyset} (B_r(x))^c,$$

donde $B_r(x)$ denota la bola abierta de centro x y radio r . Si r tendiese a infinito, $C_r(A)$ aproximaría la envoltura convexa de A .

En efecto, Walther demostró que todo subconjunto $R \subset \mathbb{R}^d$ compacto, no vacío, verifica la r_0 -rolling property si y sólo si R y $\overline{R^c}$ son r_0 -convexos y cada componente conexa de R posee interior no vacío. Esto resulta equivalente a que $\Psi_r(R) = R$ para todo $(-r_0, r_0]$, donde para un $r \geq 0$,

$$\Psi_r(R) = \bigcup_{\overline{B}_r(x) \subset R} \overline{B}_r(x), \quad \Psi_{-r}(R) = (R \ominus r\overline{B}) \oplus r\overline{B},$$

con \overline{B} denota la bola cerrada de radio 1 centrada en 0.

Así, la transformación $r \mapsto \Psi_r(R)$, conocida como granulometría de R , proporciona información sobre los cambios en R tras ser dilatado o contraído empleando \overline{B} . Así, la invarianza $\Psi_r(R) = R$ puede ser interpretada como una propiedad de suavidad de R .

Según lo anterior, Walther propuso estimar conjuntos de nivel R_τ los cuales verificaran la r_0 -rolling property teniendo en cuenta dos pasos. Primero, considerar un estimador de la densidad f_n , por ejemplo el tipo kernel presentado en (1). Entonces es posible definir:

$$\mathcal{X}_n^+(f_\tau) = \{X \in \mathcal{X}_n : f_n(X) \geq f_\tau\}, \quad \mathcal{X}_n^-(f_\tau) = \mathcal{X}_n \setminus \mathcal{X}_n^+(f_\tau).$$

Segundo, el estimador de R_τ se define como una aproximación empírica de $\Psi_r(R_\tau)$, $\hat{\Psi}_r(R_\tau)$, dada por

$$R_{\tau,n}(f_{\tau,n}) = (\mathcal{X}_n^+(f_{\tau,n}) \cap (\mathcal{X}_n^-(f_{\tau,n}) \oplus r_n \overline{B})^c) \oplus r_n \overline{B},$$

$$f_{\tau,n} = \text{máx}\{f_\tau : F_n(R_{\tau,n}(f_\tau)) \geq 1 - \tau\},$$

donde r_n es una sucesión de números convergiendo a cero a medida que el tamaño muestral aumenta que representa los radios de las bolas que ruedan en el interior de la frontera.

En definitiva, el estimador es construido a partir de una unión de bolas cuyos centros son los puntos muestrales de \mathcal{X}_n^+ que distan de cada punto de \mathcal{X}_n^- al menos r_n . Su obtención es factible y el número de pasos empleados para calcular los centros correspondientes desde un punto de vista computacional son $O(dn^2)$, resultando lineal en la dimensión.

Respecto al orden de convergencia, bajo ciertas condiciones de regularidad, si f se asume suave es más rápido que el obtenido a través del método de exceso de masa propuesto por Polonik.

Además, Walther generalizó el algoritmo presentado previamente para el caso en que la densidad presente una discontinuidad. Como antes, deben calcularse \mathcal{X}_n^+ , una estimación tipo núcleo de la densidad f_n , con parámetro ventana verificando $h_n \rightarrow 0$, $h_n \gg (\frac{\log n}{n})^{1/d}$, y $\mathcal{X}_n^+(f_\tau)$, $\mathcal{X}_n^-(f_\tau)$.

Así, es posible definir:

$$\widetilde{\mathcal{X}}_n^+ = \mathcal{X}_n^+(f_\tau) \cap (\mathcal{X}_n^-(f_\tau) \oplus 2\sigma_n B)^c, \quad \widetilde{\mathcal{X}}_n^- = \mathcal{X}_n^-(f_\tau) \cap (\mathcal{X}_n^+(f_\tau) \oplus 2\sigma_n B)^c$$

y

$$L = \mathcal{X}_n \setminus (\widetilde{\mathcal{X}}_n^+ \cup \widetilde{\mathcal{X}}_n^-),$$

pudiendo garantizar para n grande, $\widetilde{\mathcal{X}}_n^+ \subset R_\tau$ y $\widetilde{\mathcal{X}}_n^- \subset R_\tau^c$. A continuación, se suavizarán los conjuntos $\widetilde{\mathcal{X}}_n^+ \cup L_i$, para varios $L_i \subset L$ empleando suavización granulométrica dos veces. Los conjuntos L_i , $i = 1, \dots, I$ son obtenidos por procesos de aleatorización, por ejemplo, considerando todos los posibles subconjuntos $L_i \subset L$ si $|L|$ es pequeño. Así, obtenemos como candidatos,

$$Z_{n,i} = \left\{ X \in \mathcal{X}_n : |X - Y| > 2r_n \text{ para todo } Y \in ((\widetilde{\mathcal{X}}_n^+ \cup L_i) \oplus r_n B)^c \cap \widetilde{\mathcal{X}}_n^- \right\} \oplus r_n B.$$

El estimador será nuevamente una unión de bolas, siendo los centros fácilmente calculables. Finalmente, $\hat{R}_{n,\tau}$ es elegido entre todos los $Z_{n,i}$ como el óptimo de ellos si consideramos el método de exceso de masa. Es decir,

$$P_n(\hat{R}_{n,\tau}) - f_\tau(\hat{R}_{n,\tau}) \geq P_n(Z_{n,i}) - f_\tau(Z_{n,i}), \text{ para todo } i = 1, \dots, I.$$

El método descrito resulta interesante por dos razones bien diferentes: permite visualizar geométricamente la construcción del estimador y resulta competitivo en casos muy generales que consideran conjuntos de nivel no convexos.

5. Estudio simulación

En esta sección se presentan los resultados obtenidos a partir de los estudios de simulación realizados para comparar algunos de los algoritmos descritos a través del estudio de los errores obtenidos, considerando tres criterios diferentes $d_H(\hat{R}_{\tau,n}, R)$, $\mu(\hat{R}_{\tau,n} \Delta R_\tau)$ y $\mu_f(\hat{R}_{\tau,n} \Delta R_\tau)$, donde μ denota la medida de Lebesgue. Generaremos $M = 250$ muestras de tamaño $n = 1000$ fijando como modelos de prueba las 15 densidades de Marron y Wand y la densidad exponencial de parámetro 1, discontinua en cero.

Primero, nos centraremos en los métodos plug-in. Compararemos los errores obtenidos al considerar tres selectores de ventana bien diferentes: el propuesto por Samworth y Wand (2010), validación cruzada y Seather y Jones. De modo análogo, compararemos el método propuesto por Samworth y Wand (2010) con el método exceso de masa propuesto por Hartigan para el caso unidimensional, donde el parámetro f_τ ha sido estimado usando la ventana de Seather y Jones. Finalmente, repetiremos el proceso para analizar el método de Samworth y Wand (2010) frente al método granulométrico asumiendo f suave.

5.1. Comparación método Samworth-Wand con validación cruzada y Seather-Jones

La comparación de los errores obtenidos al considerar los tres selectores Samworth y Wand (2010), validación cruzada y Seather y Jones, se resume en las tablas siguientes que contienen las medias (M) y desviaciones típicas respectivas (DT) para los tres criterios de error, para cada uno de los modelos de prueba establecidos y $\tau = 0.5$:

	Método SW		Método CV		Método SJ	
	M_{d_H}	DT_{d_H}	M_{d_H}	DT_{d_H}	M_{d_H}	DT_{d_H}
1	0.079	0.049	0.087	0.041	0.077	0.045
2	0.052	0.028	0.065	0.029	0.053	0.028
3	0.062	0.070	0.056	0.064	0.054	0.058
4	0.092	0.107	0.125	0.126	0.156	0.132
5	0.009	0.005	0.010	0.004	0.008	0.004
6	0.158	0.131	0.175	0.090	0.146	0.075
7	0.075	0.031	0.083	0.028	0.080	0.032
8	0.260	0.182	0.224	0.109	0.209	0.111
9	0.631	0.186	0.487	0.218	0.442	0.266
10	0.164	0.164	0.065	0.063	0.217	0.170
11	0.415	0.119	0.368	0.071	0.400	0.075
12	0.959	0.182	0.944	0.166	0.944	0.157
13	0.407	0.326	0.282	0.093	0.293	0.084
14	1.478	0.496	0.857	0.251	1.069	0.327
15	1.356	0.033	1.118	0.418	1.352	0.033
Exp(1)	0.084	0.069	0.093	0.074	0.089	0.060

Tabla 4: Resumen descriptivo errores distancia Hausdorff para $\tau = 0.5$, $M = 250$ y $n = 1000$.

	Método SW		Método CV		Método SJ	
	M_μ	DT_μ	M_μ	DT_μ	M_μ	DT_μ
1	0.121	0.084	0.121	0.056	0.116	0.078
2	0.076	0.047	0.087	0.037	0.076	0.044
3	0.078	0.132	0.069	0.079	0.068	0.038
4	0.088	0.056	0.086	0.052	0.117	0.112
5	0.013	0.008	0.013	0.005	0.011	0.007
6	0.354	0.206	0.375	0.210	0.331	0.176
7	0.167	0.069	0.182	0.062	0.179	0.072
8	0.581	0.344	0.552	0.308	0.517	0.300
9	0.501	0.210	0.546	0.197	0.468	0.159
10	0.389	0.207	0.244	0.073	0.599	0.313
11	0.962	0.705	1.149	0.732	1.144	0.761
12	0.516	0.247	0.478	0.218	0.493	0.230
13	1.322	0.874	1.809	0.772	1.611	0.800
14	0.612	0.114	0.466	0.097	0.608	0.112
15	0.532	0.053	0.484	0.074	0.553	0.053
Exp(1)	0.089	0.120	0.117	0.191	0.098	0.113

Tabla 5: Resumen descriptivo errores en longitud para $\tau = 0.5$, $M = 250$ y $n = 1000$.

	Método SW		Método CV		Método SJ	
	M_{μ_f}	DT_{μ_f}	M_{μ_f}	DT_{μ_f}	M_{μ_f}	DT_{μ_f}
1	0.039	0.027	0.037	0.018	0.037	0.026
2	0.032	0.020	0.036	0.015	0.032	0.019
3	0.024	0.015	0.025	0.018	0.025	0.015
4	0.023	0.015	0.022	0.014	0.028	0.016
5	0.036	0.022	0.034	0.015	0.030	0.018
6	0.089	0.053	0.092	0.051	0.082	0.044
7	0.053	0.023	0.057	0.020	0.057	0.024
8	0.161	0.092	0.150	0.083	0.141	0.081
9	0.116	0.049	0.123	0.044	0.108	0.038
10	0.140	0.055	0.091	0.029	0.185	0.063
11	0.125	0.043	0.130	0.044	0.121	0.036
12	0.140	0.057	0.130	0.051	0.134	0.051
13	0.163	0.047	0.151	0.041	0.145	0.041
14	0.198	0.039	0.149	0.031	0.190	0.037
15	0.169	0.018	0.154	0.025	0.175	0.018
Exp(1)	0.038	0.025	0.042	0.027	0.043	0.022

Tabla 6: Resumen descriptivo errores (μ_f) para $\tau = 0.5$, $M = 250$ y $n = 1000$.

La figura siguiente será observada en vertical. Para tres densidades concretas, hemos representado los diagramas de cajas para los errores obtenidos con cada uno de los tres métodos considerados, Samworth y Wand, validación cruzada y Seather y Jones. En la primera fila, se considera $d_H(\hat{R}_{\tau,n}, R)$; en la segunda, $\mu(\hat{R}_{\tau,n} \triangle R_\tau)$, donde μ denota la medida de Lebesgue (longitud) y en la tercera y última, $\mu_f(\hat{R}_{\tau,n} \triangle R_\tau)$.

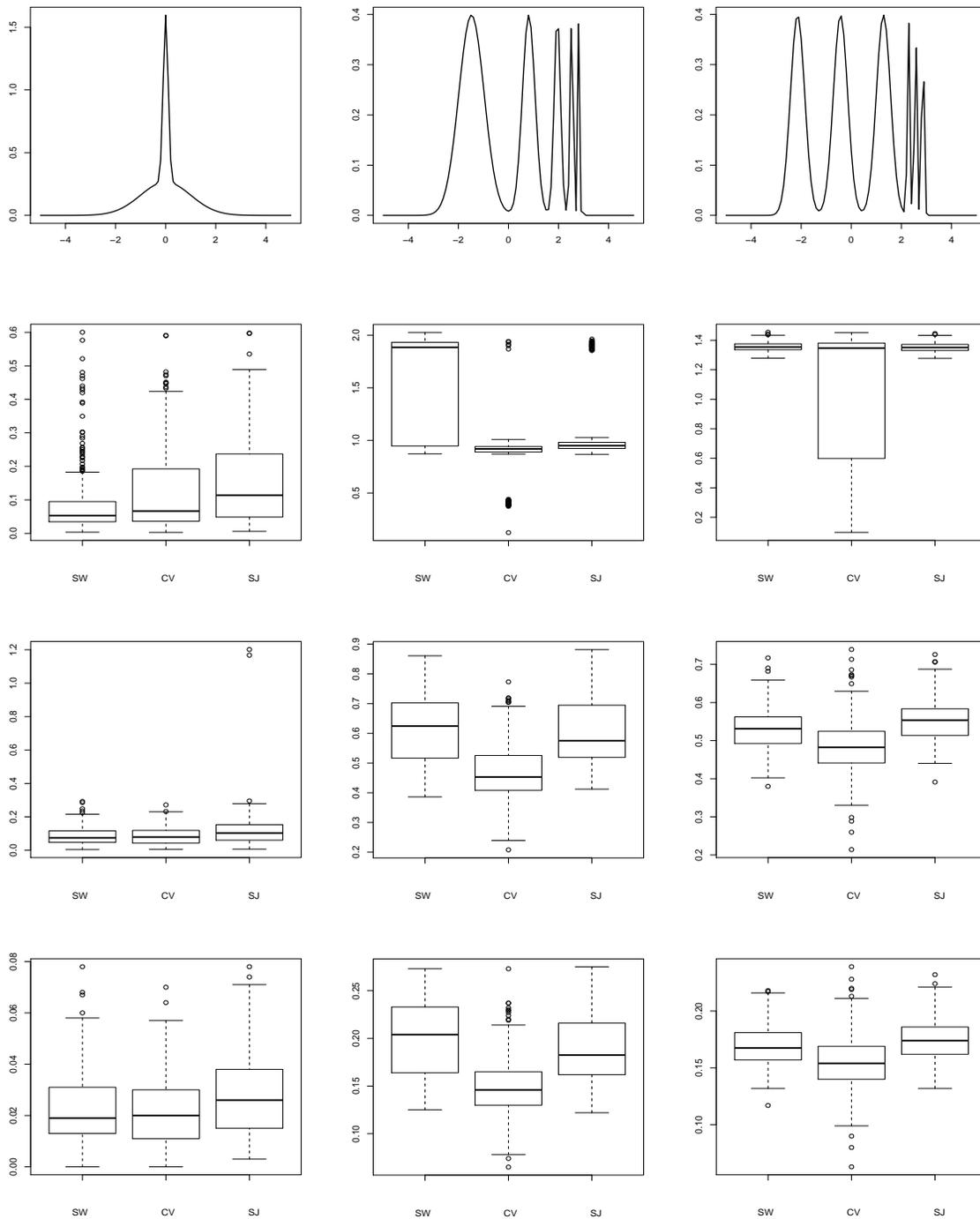


Figura 8: Boxplots errores métodos plug-in para $\tau = 0.5$ y $M = 250$ muestras con $n = 1000$.

Análogamente, se exponen los resultados obtenidos para $\tau = 0.2$:

	Método SW		Método CV		Método SJ	
	M_{d_H}	DT_{d_H}	M_{d_H}	DT_{d_H}	M_{d_H}	DT_{d_H}
1	0.073	0.038	0.110	0.040	0.074	0.037
2	0.058	0.031	0.080	0.034	0.060	0.030
3	0.184	0.159	0.232	0.191	0.202	0.174
4	0.133	0.080	0.156	0.098	0.174	0.105
5	0.009	0.005	0.014	0.006	0.010	0.005
6	0.087	0.035	0.072	0.022	0.082	0.033
7	0.074	0.026	0.092	0.026	0.075	0.025
8	0.067	0.044	0.074	0.046	0.068	0.037
9	0.113	0.044	0.118	0.037	0.118	0.038
10	0.063	0.015	0.051	0.016	0.068	0.012
11	0.080	0.038	0.062	0.030	0.075	0.033
12	1.505	0.332	1.408	0.352	1.484	0.323
13	0.117	0.035	0.123	0.045	0.108	0.035
14	0.180	0.105	0.124	0.061	0.172	0.064
15	0.184	0.194	0.079	0.049	0.116	0.051
Exp(1)	0.144	0.122	0.093	0.074	0.130	0.096

Tabla 7: Resumen descriptivo errores distancia Hausdorff para $\tau = 0.2$, $M = 250$ y $n = 1000$.

	Método SW		Método CV		Método SJ	
	M_μ	DT_μ	M_μ	DT_μ	M_μ	DT_μ
1	0.104	0.054	0.166	0.064	0.106	0.051
2	0.081	0.043	0.121	0.051	0.084	0.041
3	0.256	0.301	0.250	0.150	0.230	0.138
4	0.469	0.936	1.060	1.478	1.492	1.613
5	0.014	0.008	0.022	0.011	0.014	0.008
6	1.404	1.499	1.750	1.554	0.976	1.328
7	0.160	0.054	0.223	0.063	0.164	0.052
8	0.090	0.045	0.142	0.366	0.093	0.048
9	1.953	1.116	1.407	0.923	1.096	0.971
10	2.515	0.824	1.297	0.653	2.280	0.606
11	1.473	1.471	1.826	1.526	1.257	1.407
12	3.750	1.533	3.716	1.570	3.984	1.317
13	0.652	0.960	0.574	0.785	0.387	0.547
14	0.817	0.168	0.418	0.385	0.868	0.170
15	0.881	0.170	0.351	0.343	0.904	0.152
Exp(1)	0.176	0.093	0.117	0.191	0.169	0.078

Tabla 8: Resumen descriptivo errores en longitud para $\tau = 0.2$, $M = 250$ y $n = 1000$.

	Método SW		Método CV		Método SJ	
	M_{μ_f}	DT_{μ_f}	M_{μ_f}	DT_{μ_f}	M_{μ_f}	DT_{μ_f}
1	0.018	0.010	0.027	0.011	0.018	0.009
2	0.015	0.009	0.022	0.010	0.017	0.009
3	0.025	0.017	0.034	0.020	0.030	0.020
4	0.034	0.022	0.042	0.024	0.047	0.024
5	0.016	0.009	0.022	0.011	0.015	0.009
6	0.046	0.016	0.039	0.011	0.045	0.018
7	0.026	0.010	0.036	0.011	0.026	0.009
8	0.018	0.010	0.021	0.014	0.019	0.010
9	0.083	0.022	0.087	0.023	0.083	0.025
10	0.058	0.020	0.045	0.016	0.064	0.021
11	0.048	0.016	0.042	0.012	0.047	0.015
12	0.076	0.017	0.074	0.019	0.079	0.017
13	0.057	0.021	0.063	0.030	0.052	0.023
14	0.093	0.023	0.060	0.017	0.098	0.020
15	0.102	0.039	0.056	0.079	0.091	0.023
Exp(1)	0.020	0.014	0.042	0.027	0.018	0.016

Tabla 9: Resumen descriptivo errores (μ_f) para $\tau = 0.2$, $M = 250$ y $n = 1000$.

A continuación, con criterios análogos a los comentados previamente, exponemos los diagramas de cajas, para ciertos modelos, de los errores obtenidos con cada uno de los tres métodos considerados:

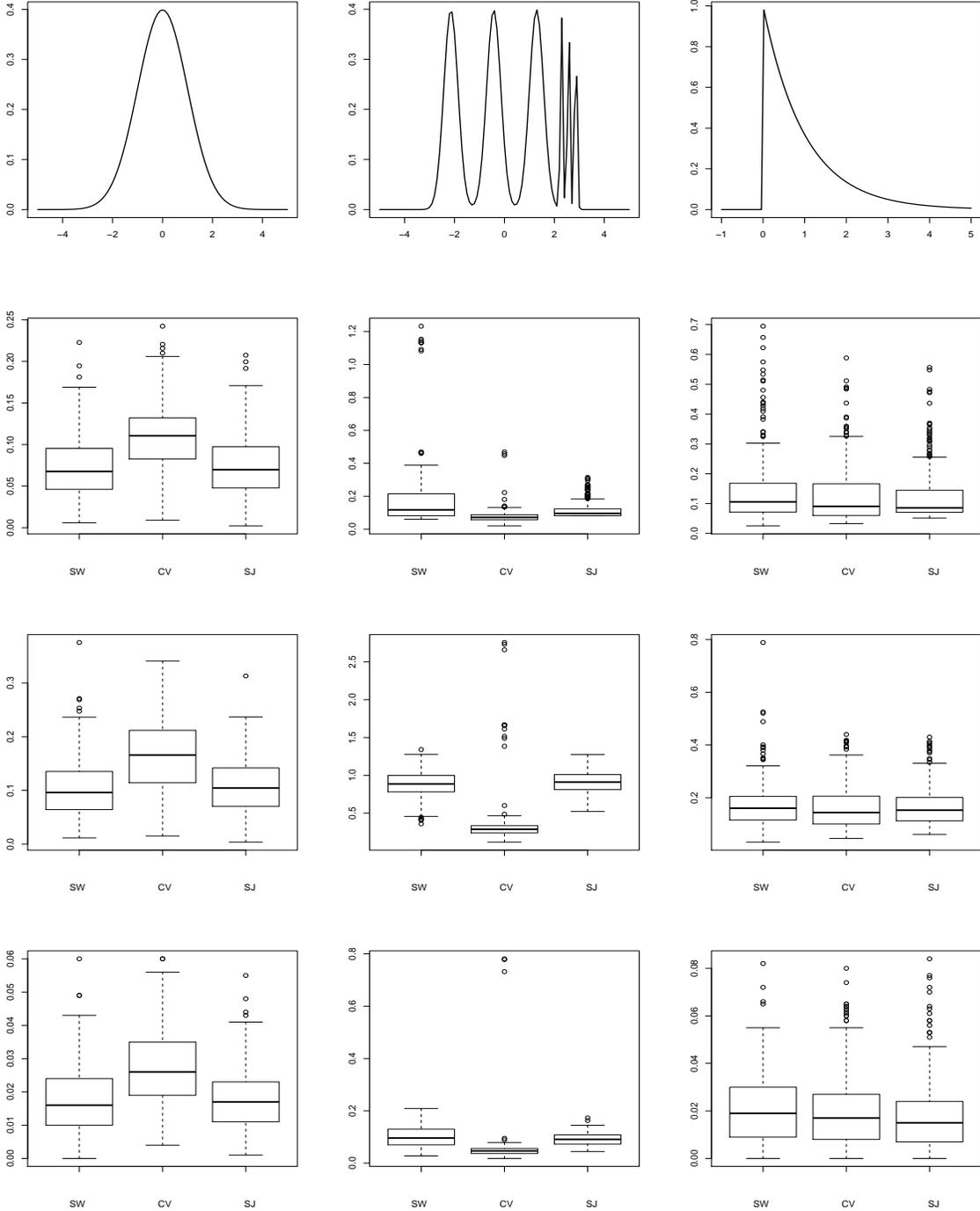


Figura 9: Boxplots errores métodos plug-in para $\tau = 0.2$ y $M = 250$ muestras con $n = 1000$.

A partir de las tablas 4, 5 y 6, para $\tau = 0.5$, es posible analizar la importancia de elección de una medida de riesgo u otra. Así, para el modelo 14 de Marron y Wand al considerar la distancia Hausdorff, el método de Samworth y Wand proporciona errores considerablemente mayores que los obtenidos mediante validación cruzada o, incluso, Seather y Jones. En cambio, las diferencias se reducen de manera notable al considerar la longitud o μ_f . Para el modelo exponencial de parámetro uno, el método de Samworth y Wand resulta más competitivo en cualquiera de las tres situaciones consideradas.

La Figura 8 permite observar de manera gráfica el comportamiento de cada una de las clases de error fijadas. Para el modelo 4, μ_f reduce notablemente el número de atípicos, respecto a los obtenidos al considerar la distancia Hausdorff. Para el modelo 15, la distancia Hausdorff en validación cruzada proporciona errores, en distribución, muy distintos a los obtenidos con los otros dos métodos considerados. En cambio, μ_f proporciona errores con distribuciones más similares a pesar de que el método más competitivo continuaría siendo validación cruzada.

5.2. Comparación método plug-in Samworth y Wand y exceso de masa

En este apartado, compararemos el método propuesto por Samworth y Wand (2010) con exceso de masa para el caso unidimensional, donde f_τ ha sido estimado usando la ventana de Seather y Jones.

El análisis descriptivo de los errores obtenidos se detalla en las tablas siguientes, las cuales contienen las medias (M) y desviaciones típicas (DT) para los seis primeros modelos de Marron y Wand y la densidad Exp(1). La no consideración de los restantes densidades de Marron y Wand se debe a que el método de exceso de masa considerado establece como restricción de forma la convexidad. En este sentido, exponemos los resultados obtenidos para el modelo sexto de Marron y Wand (caso más sencillo de bimodalidad) para observar como, efectivamente, el método exceso de masa proporciona errores considerablemente mayores que el método plug-in considerado.

	Método EM						Método SW					
	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}
1	0.152	0.078	0.217	0.114	0.067	0.036	0.080	0.046	0.125	0.077	0.039	0.025
2	0.098	0.050	0.143	0.079	0.060	0.033	0.046	0.028	0.068	0.045	0.029	0.019
3	0.058	0.046	0.069	0.047	0.029	0.020	0.065	0.074	0.079	0.105	0.027	0.018
4	0.081	0.032	0.117	0.051	0.033	0.016	0.097	0.111	0.093	0.091	0.022	0.013
5	0.015	0.008	0.022	0.012	0.061	0.034	0.009	0.005	0.013	0.008	0.036	0.023
6	1.294	0.664	1.454	0.685	0.294	0.037	0.170	0.131	0.363	0.165	0.091	0.042
Exp(1)	0.085	0.063	0.087	0.063	0.042	0.029	0.078	0.059	0.130	0.255	0.039	0.023

Tabla 10: Resumen descriptivo errores para $\tau = 0.5$, $M = 250$ y $n = 1000$.

	Método EM						Método SW					
	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}
1	0.132	0.070	0.189	0.100	0.033	0.019	0.075	0.041	0.106	0.061	0.018	0.011
2	0.100	0.055	0.140	0.078	0.027	0.015	0.058	0.034	0.082	0.047	0.015	0.009
3	0.128	0.106	0.140	0.108	0.022	0.018	0.187	0.166	0.289	0.333	0.026	0.019
4	0.145	0.084	0.211	0.120	0.034	0.021	0.127	0.076	0.379	0.791	0.032	0.020
5	0.014	0.007	0.020	0.010	0.023	0.013	0.010	0.005	0.014	0.008	0.015	0.009
6	0.131	0.063	1.631	1.555	0.058	0.022	0.082	0.0320	1.370	1.506	0.044	0.015
Exp(1)	0.108	0.083	0.109	0.083	0.022	0.016	0.143	0.108	0.179	0.090	0.020	0.015

Tabla 11: Resumen descriptivo errores para $\tau = 0.2$, $M = 250$ y $n = 1000$.

A continuación, exponemos de forma gráfica comparaciones de los errores obtenidos para algunos de los modelos considerados. Las figuras siguientes se estructuran de manera análoga a las expuestas en la sección previa. En este caso, consideramos $\tau = 0.5$ y $\tau = 0.2$.

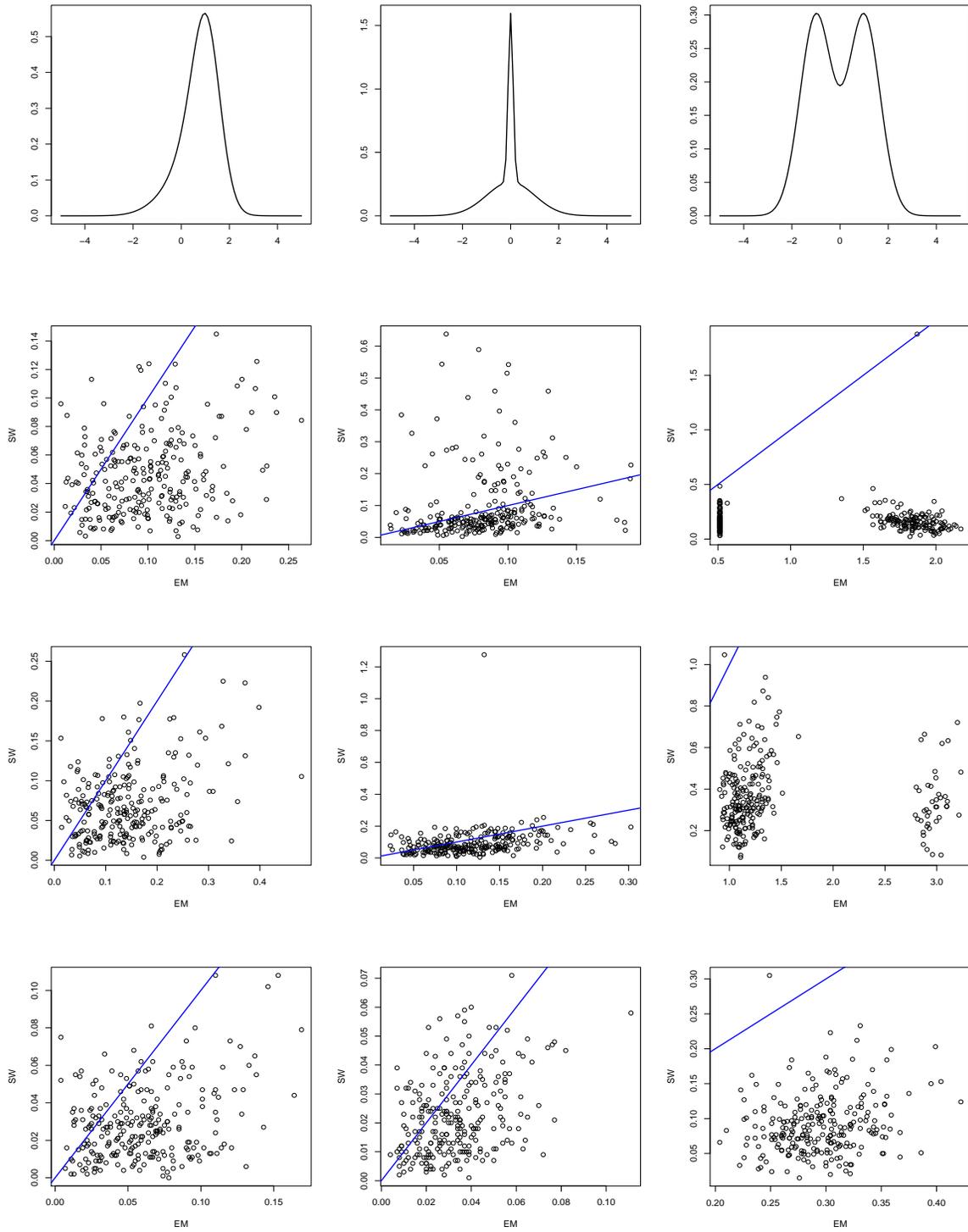


Figura 10: Comparación errores métodos plug-in y exceso masa, $\tau = 0.5$, $M = 250$ y $n = 1000$.

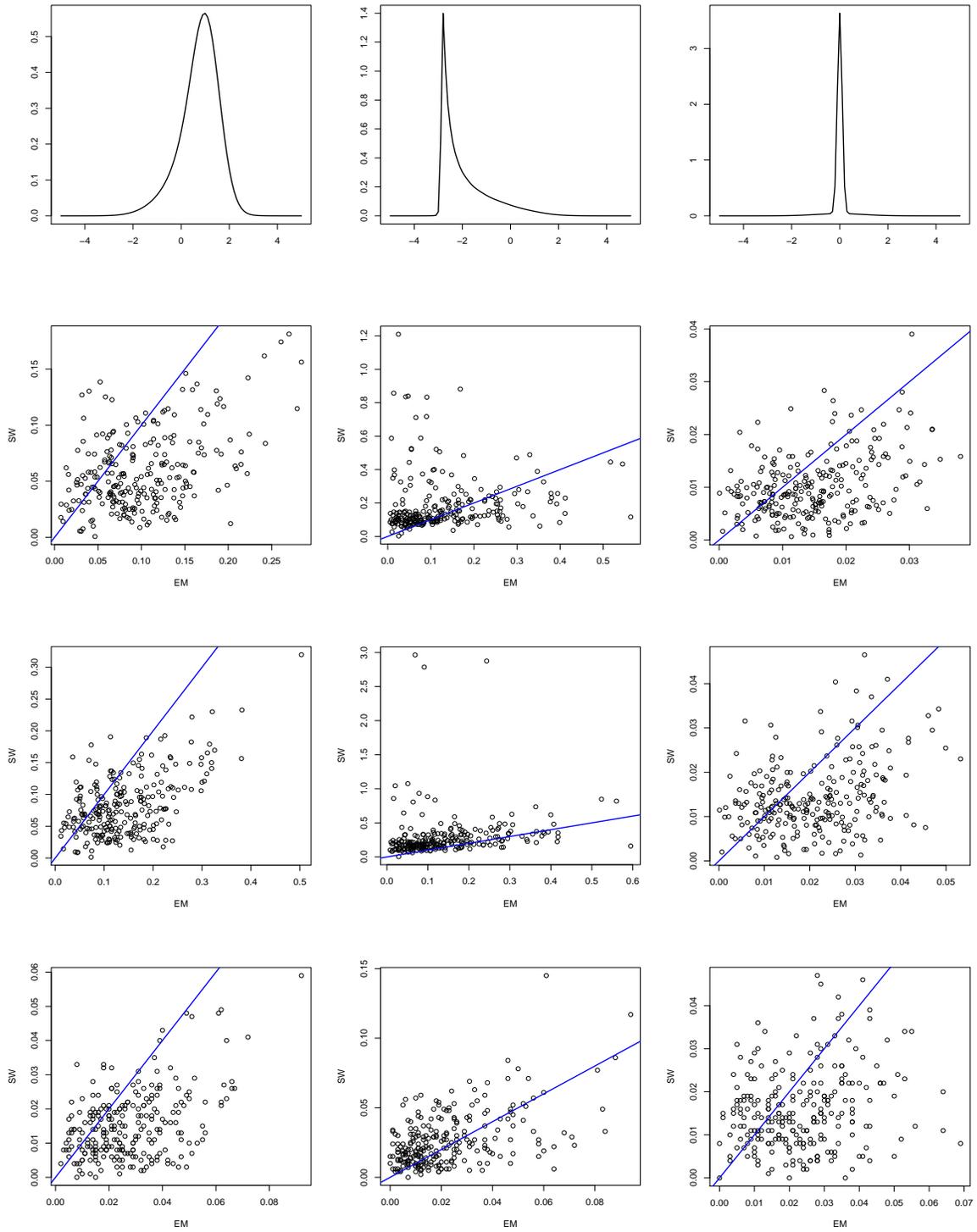


Figura 11: Comparación errores métodos plug-in y exceso masa, $\tau = 0.2$, $M = 250$ y $n = 1000$.

Con criterios gráficos análogos, para $\tau = 0.5$ e $\tau = 0.2$, exponemos los respectivos boxplots para los errores obtenidos a través de la consideración de ambos métodos.

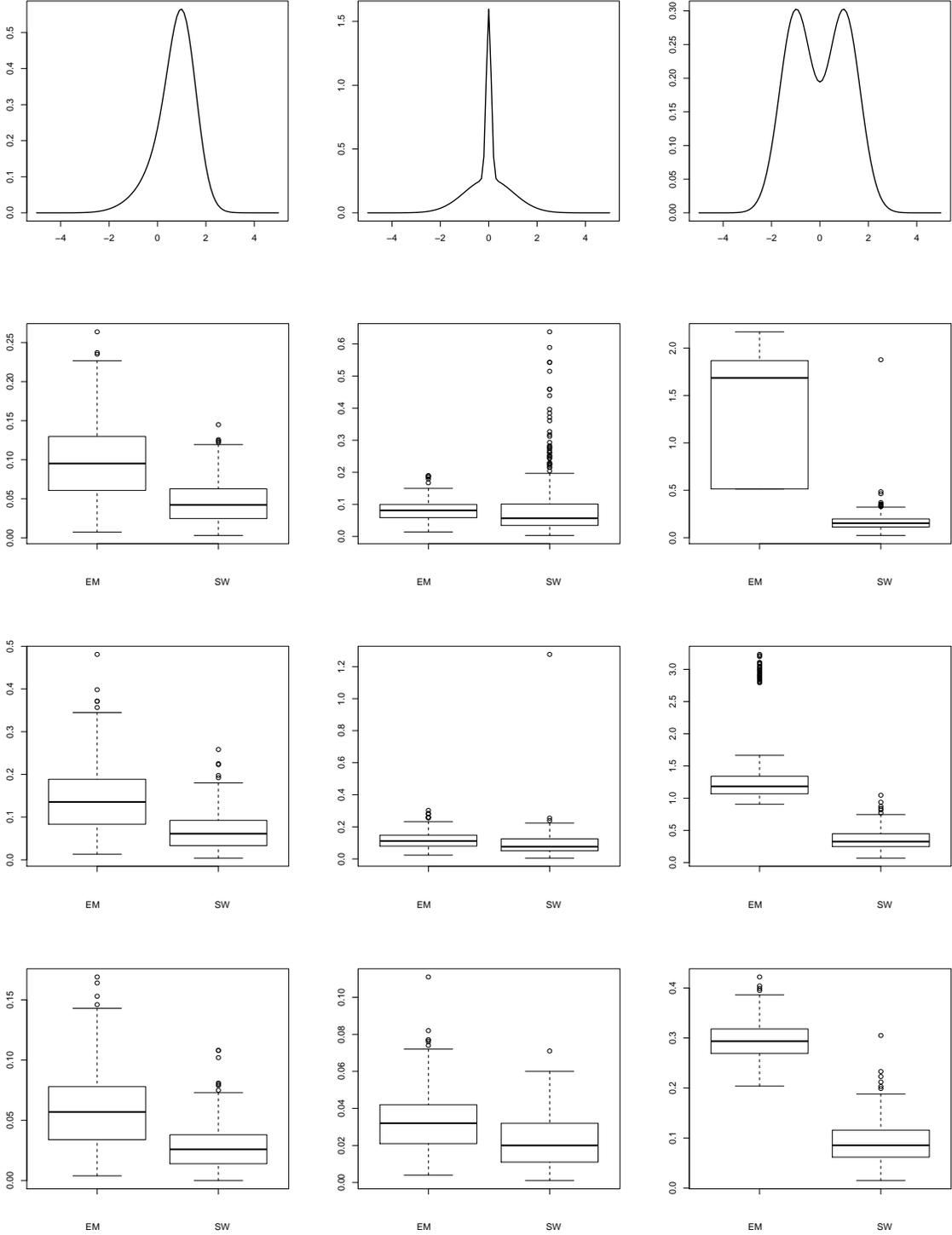


Figura 12: Boxplots errores para $\tau = 0.5$ y $M = 250$ muestras de tamaño $n = 1000$.

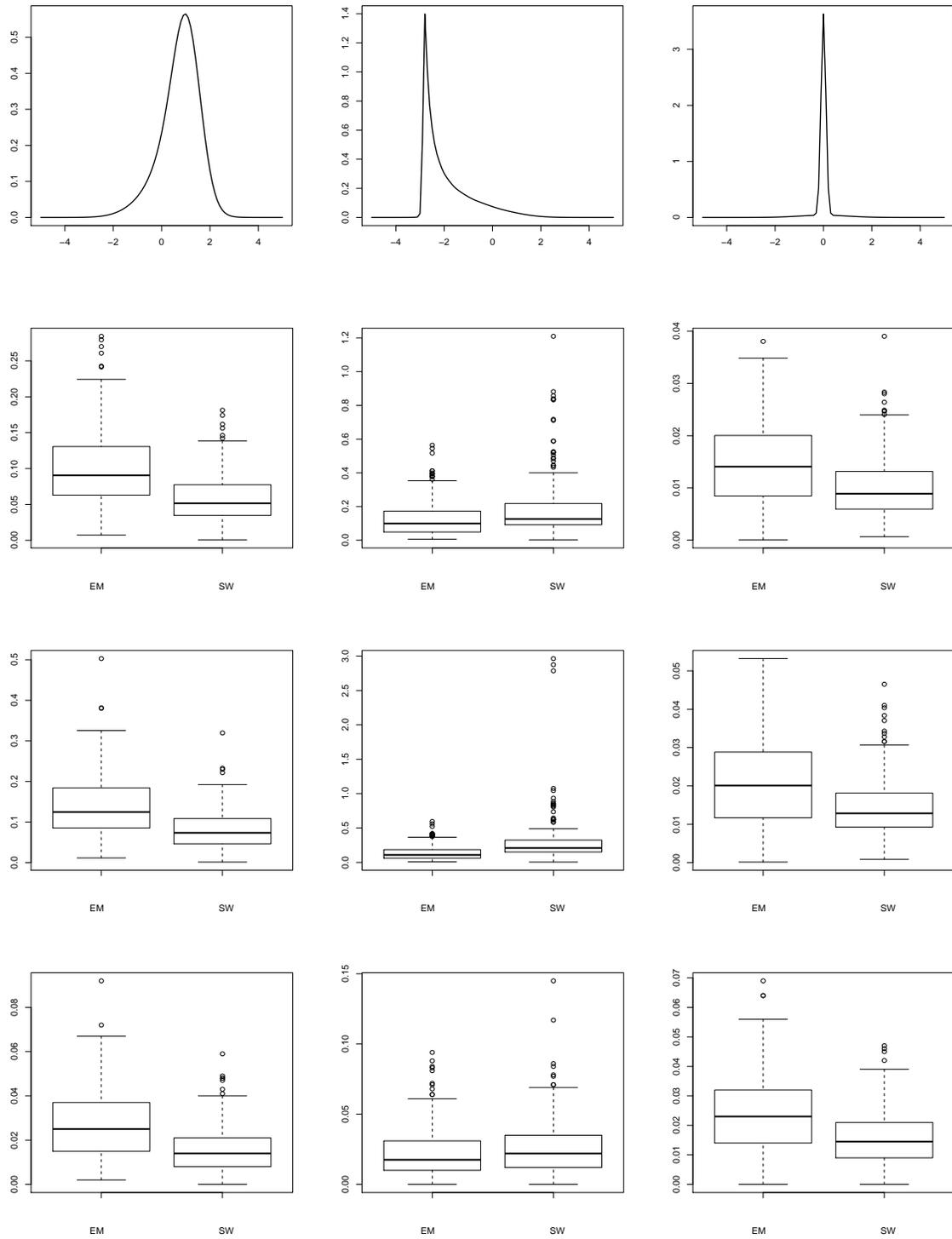


Figura 13: Boxplots errores para $\tau = 0.2$ y $M = 250$ muestras de tamaño $n = 1000$.

En relación con lo comentado previamente, la Figura 10 permite observar el comportamiento de ambos métodos para el modelo 6 de las densidades de Marron y Wand para $\tau = 0.5$. Obviamente, el método plug-in resulta mucho más competitivo. En cambio, para $\tau = 0.2$, lógicamente las diferencias entre los errores no resultan tan evidentes (ver Tabla 11).

De nuevo, se observa la diferencia entre los tres criterios de error tenidos en cuenta. Así, la Tabla 10 refleja como sólo para las densidades 3 y 4 exceso de masa proporciona mejores resultados si consideramos la distancia Hausdorff (recuérdese comportamiento error asintótico, Figura 5); concretamente, la Figura 12 refleja el elevado número de atípicos que el método plug-in obtiene para la densidad 4. Idénticas conclusiones pueden ser extraídas a partir de la Tabla 11. El método de exceso de masa proporciona para las densidades 3 y 4 mejores resultados si consideramos, esta vez, como criterios de error la distancia Hausdorff o la longitud de la diferencia simétrica. Para el segundo de los modelos, exceso de masa resulta más competitivo incluso considerando el criterio de error propuesto por Samworth y Wand.

5.3. Comparación método plug-in Samworth y Wand y granulométrico

En esta sección, compararemos el método propuesto por Samworth y Wand (2010) con el método granulométrico para el caso unidimensional considerando, como antes, tres criterios de error bien diferentes:

$$d_H(\hat{R}_{\tau,n}, R), \mu(\hat{R}_{\tau,n} \triangle R_\tau) \text{ y } \mu_f(\hat{R}_{\tau,n} \triangle R_\tau),$$

donde μ denota la medida de Lebesgue.

Las tablas siguientes resumen el análisis descriptivo realizado para los distintos modelos y criterios de error a través del cálculo de las medias (M) y desviaciones típicas (DT) considerando distintos valores del parámetro τ , $\tau = 0.5$ y $\tau = 0.2$, y r_n , $r_n = 0.02$ y $r_n = 0.1$.

	Método W						Método SW					
	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}
1	0.166	0.046	0.237	0.059	0.071	0.018	0.076	0.044	0.117	0.075	0.036	0.025
2	0.127	0.035	0.191	0.045	0.075	0.018	0.050	0.029	0.073	0.046	0.031	0.020
3	0.229	0.085	0.244	0.048	0.085	0.016	0.060	0.061	0.077	0.101	0.026	0.017
4	0.443	0.130	0.438	0.063	0.085	0.016	0.082	0.082	0.091	0.058	0.026	0.017
5	0.084	0.006	0.160	0.011	0.301	0.019	0.008	0.005	0.013	0.008	0.036	0.023
6	0.202	0.085	0.412	0.155	0.100	0.039	0.171	0.080	0.385	0.194	0.097	0.051
7	0.089	0.027	0.200	0.064	0.061	0.021	0.073	0.031	0.168	0.071	0.053	0.024
8	0.214	0.084	0.523	0.312	0.136	0.075	0.252	0.158	0.587	0.343	0.162	0.095
9	0.386	0.231	0.536	0.164	0.121	0.037	0.627	0.185	0.528	0.265	0.120	0.050
10	0.208	0.162	0.694	0.421	0.191	0.068	0.167	0.168	0.403	0.215	0.143	0.054
11	0.362	0.073	1.819	0.776	0.133	0.032	0.442	0.161	1.080	0.778	0.123	0.047
12	0.906	0.063	0.591	0.331	0.141	0.042	0.944	0.148	0.533	0.306	0.140	0.059
13	0.236	0.085	2.070	0.830	0.151	0.037	0.389	0.293	1.351	0.872	0.165	0.047
14	0.994	0.230	0.639	0.119	0.198	0.037	1.526	0.513	0.614	0.113	0.199	0.038
15	1.335	0.074	0.618	0.061	0.191	0.021	1.355	0.035	0.535	0.048	0.171	0.017
Exp(1)	0.252	0.071	0.244	0.041	0.101	0.019	0.079	0.057	0.106	0.187	0.040	0.023

Tabla 12: Resumen descriptivo errores para $\tau = 0.5$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

	Método W						Método SW					
	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}
1	0.411	0.080	0.743	0.140	0.100	0.016	0.0748	0.041	0.104	0.059	0.018	0.010
2	0.469	0.097	0.721	0.129	0.104	0.016	0.059	0.033	0.084	0.049	0.016	0.010
3	1.138	0.233	0.962	0.145	0.105	0.016	0.181	0.168	0.319	0.413	0.028	0.021
4	0.847	0.154	1.322	0.207	0.136	0.017	0.141	0.083	0.427	0.843	0.034	0.020
5	2.609	0.387	2.280	0.169	0.168	0.012	0.009	0.005	0.013	0.008	0.015	0.009
6	0.229	0.054	3.472	0.668	0.088	0.017	0.082	0.031	1.306	1.481	0.044	0.014
7	0.164	0.036	0.483	0.120	0.070	0.017	0.075	0.026	0.170	0.061	0.028	0.011
8	0.384	0.082	0.535	0.101	0.089	0.016	0.061	0.036	0.085	0.048	0.017	0.010
9	0.111	0.033	3.817	1.172	0.083	0.014	0.110	0.038	1.999	1.095	0.083	0.020
10	0.271	0.095	4.108	0.178	0.115	0.015	0.063	0.011	2.521	0.803	0.058	0.017
11	0.223	0.051	3.546	0.567	0.086	0.016	0.078	0.039	1.566	1.491	0.046	0.015
12	0.919	0.042	4.750	0.636	0.115	0.018	1.483	0.335	3.803	1.498	0.076	0.018
13	0.182	0.049	3.275	0.816	0.089	0.016	0.118	0.034	0.624	0.918	0.058	0.020
14	0.169	0.040	0.899	0.163	0.111	0.017	0.173	0.095	0.814	0.159	0.093	0.023
15	0.113	0.025	0.764	0.127	0.093	0.018	0.174	0.182	0.873	0.175	0.098	0.038
Exp(1)	1.110	0.272	0.984	0.170	0.121	0.016	0.147	0.113	0.173	0.087	0.020	0.015

Tabla 13: Resumen descriptivo errores para $\tau = 0.2$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

	Método W						Método SW					
	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}	M_{d_H}	DT_{d_H}	M_μ	DT_μ	M_{μ_f}	DT_{μ_f}
1	0.153	0.048	0.215	0.062	0.065	0.020	0.077	0.046	0.127	0.148	0.037	0.025
2	0.122	0.032	0.182	0.043	0.072	0.018	0.050	0.026	0.074	0.044	0.031	0.019
3	0.220	0.088	0.259	0.072	0.090	0.024	0.062	0.075	0.091	0.160	0.027	0.017
4	0.386	0.120	0.478	0.099	0.113	0.024	0.086	0.089	0.095	0.114	0.022	0.014
5	0.070	0.006	0.132	0.011	0.267	0.021	0.010	0.006	0.015	0.009	0.040	0.026
6	0.195	0.081	0.391	0.158	0.095	0.039	0.167	0.085	0.375	0.200	0.095	0.052
7	0.087	0.028	0.196	0.060	0.059	0.019	0.072	0.031	0.162	0.070	0.051	0.024
8	0.212	0.094	0.494	0.306	0.132	0.074	0.235	0.145	0.552	0.350	0.151	0.091
9	0.419	0.245	0.510	0.163	0.114	0.034	0.624	0.180	0.496	0.214	0.114	0.049
10	0.373	0.132	0.860	0.454	0.227	0.061	0.187	0.175	0.385	0.194	0.141	0.051
11	0.368	0.079	1.816	0.837	0.128	0.031	0.426	0.077	1.035	0.761	0.116	0.034
12	0.908	0.097	0.571	0.301	0.142	0.045	0.955	0.180	0.542	0.277	0.147	0.064
13	0.237	0.083	2.008	0.767	0.148	0.038	0.384	0.286	1.313	0.853	0.166	0.048
14	1.217	0.458	0.713	0.158	0.219	0.048	1.419	0.499	0.600	0.113	0.194	0.039
15	1.344	0.036	0.596	0.059	0.185	0.018	1.355	0.036	0.535	0.050	0.170	0.018
Exp(1)	0.231	0.073	0.315	0.068	0.102	0.025	0.083	0.074	0.097	0.160	0.038	0.024

Tabla 14: Resumen descriptivo errores para $\tau = 0.5$, $M = 250$, $n = 1000$ y $r_n = 0.1$.

Exponemos de forma gráfica ciertas comparaciones de interés para los errores obtenidos. La figura siguiente debe ser observada con criterios idénticos a los establecidos en las secciones previas.

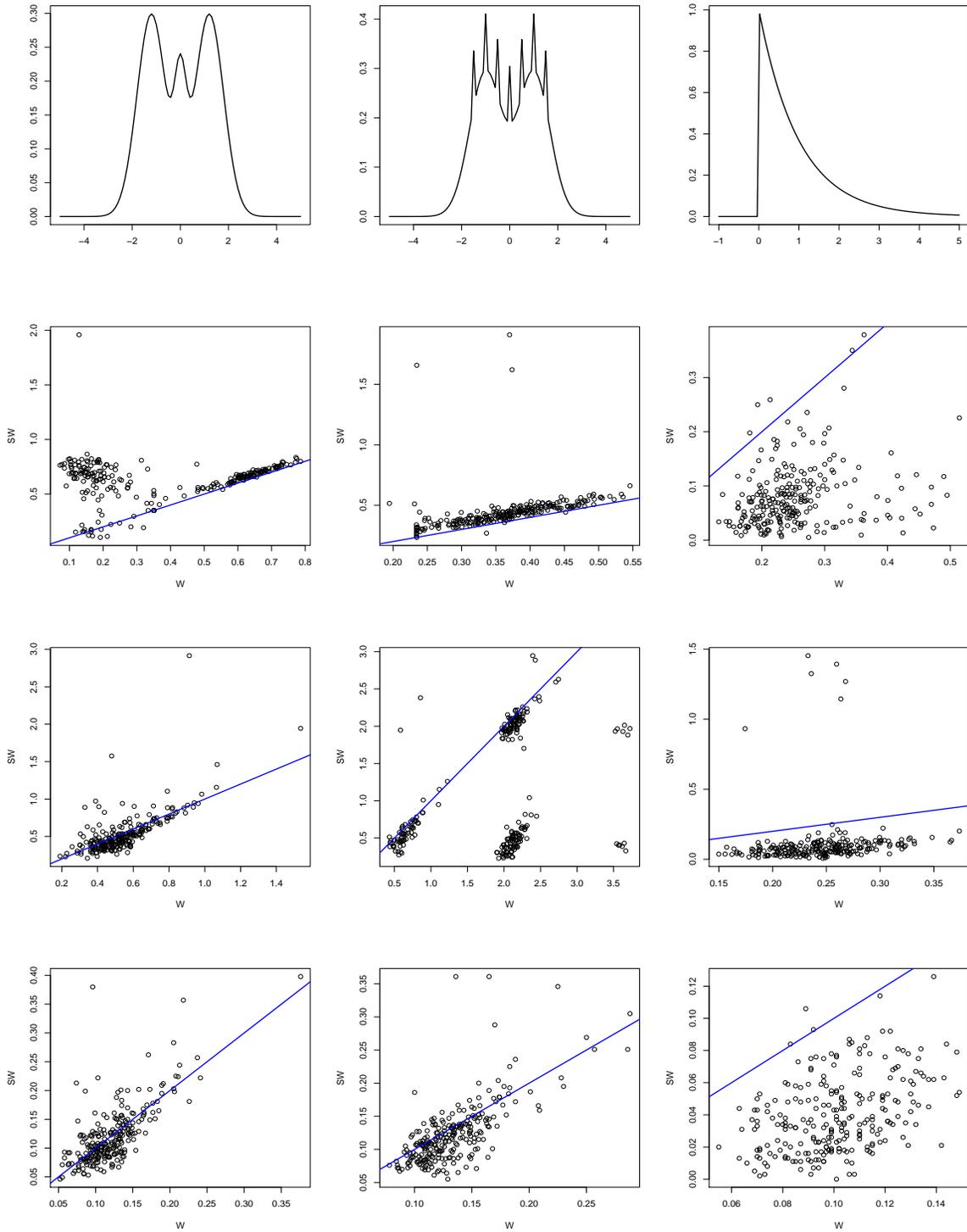


Figura 14: Comparación errores métodos plug-in y Walther, $\tau = 0.5$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

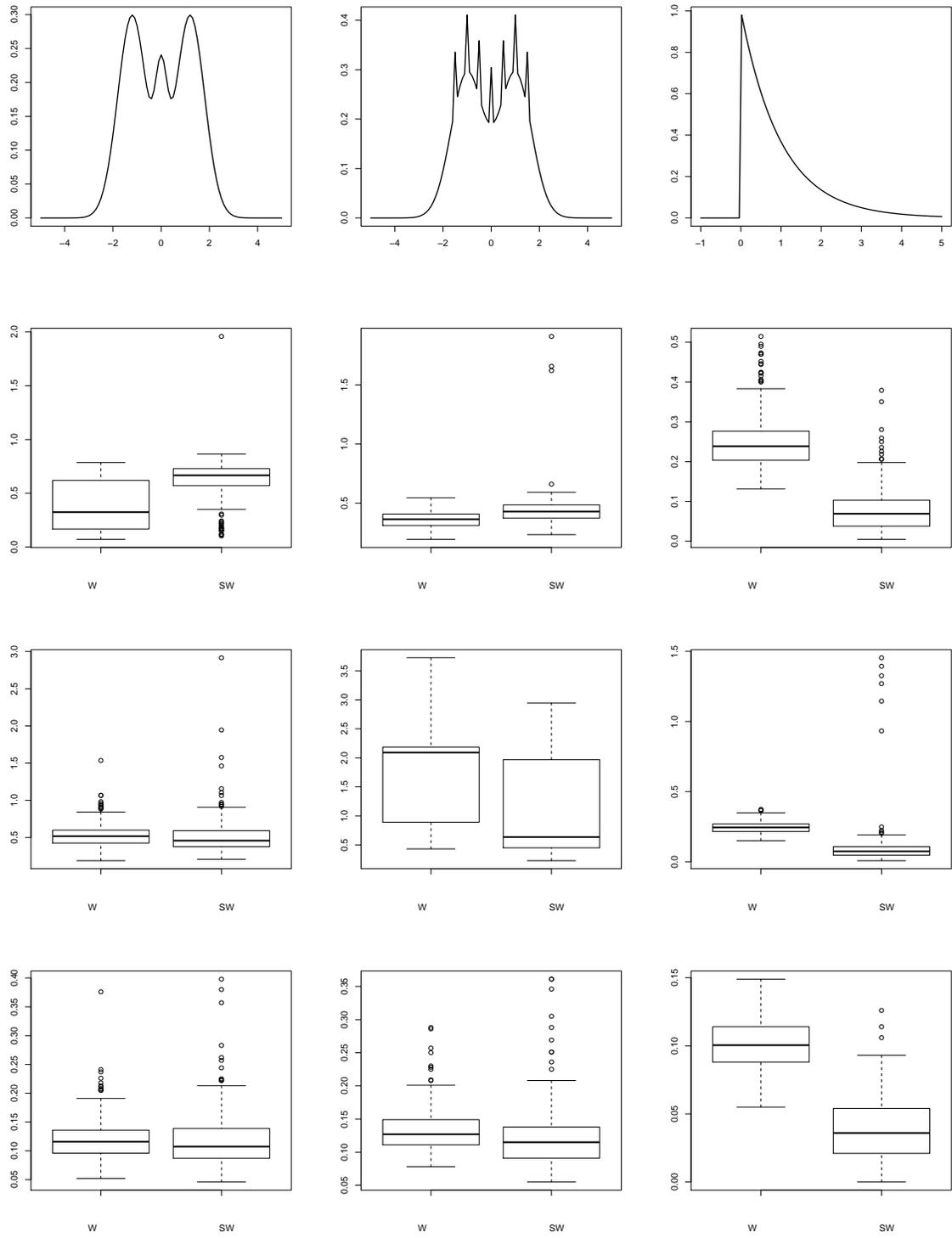


Figura 15: Boxplots errores para $\tau = 0.5$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

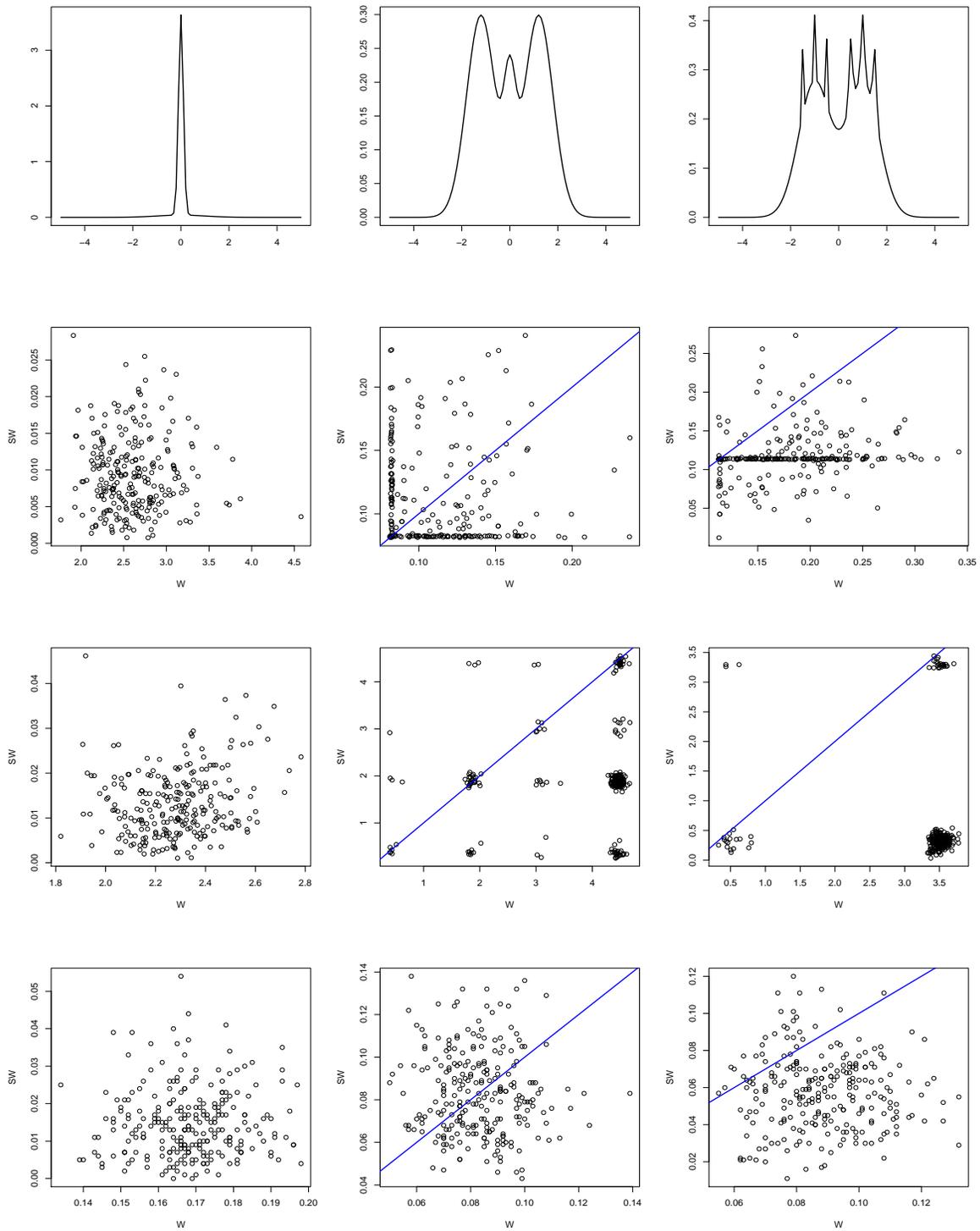


Figura 16: Comparación errores métodos plug-in y Walther, $\tau = 0.2$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

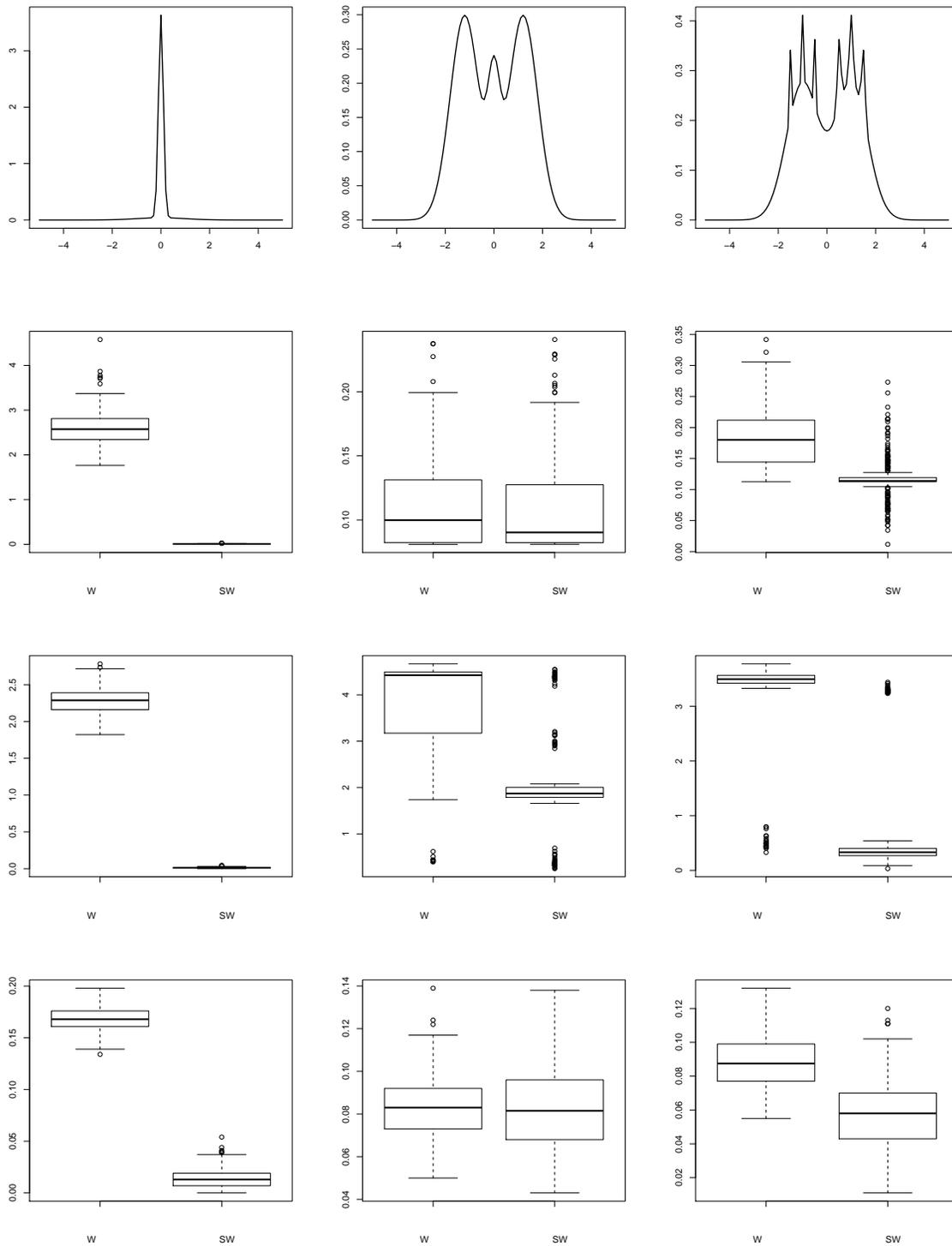


Figura 17: Boxplots errores para $\tau = 0.2$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

En general, el método plug-in proporciona mejores resultados. Un caso especial es el de la densidad exponencial de parámetro 1 para la que el método plug-in proporciona mejores resultados ya que el método propuesto por Walther asume suavidad en el borde y esta densidad es discontinua.

Una vez más, puede analizarse las diferencias existentes entre los tres criterios de error. A partir de los resultados obtenidos queda patente que la elección del parámetro r_n no resulta relevante. Por ejemplo, para $\tau = 0.5$ y $r_n = 0.02$ el método de Walther mejora (en media) al método plug-in considerado para los modelos 8, 9, 11, 12, 13 y 14. Al aumentar el valor de r_n las conclusiones son idénticas. La misma reflexión resulta si consideramos el criterio de error μ_f .

5.3.1. Análisis estadístico longitudes estimadores

En esta sección, estudiaremos brevemente el comportamiento de las longitudes de los estimadores $\hat{R}_{\tau,n}$ obtenidos al considerar los métodos de Walther y Samworth y Wand (2010), después de generar 250 muestras de tamaño $n = 1000$ para los modelos considerados previamente.

Las tablas siguientes resumen la estadística descriptiva a través de la media (M), la desviación típica (DT) y el sesgo para la longitud del estimador del conjunto de nivel, $\hat{R}_{\tau,n}$, para $\tau = 0.2$ o $\tau = 0.5$ y $r_n = 0.02$. Además, se incluye el p-valor procedente de la aplicación del test de normalidad Shapiro-Wilk.

	Método W				Método SW			
	M	DT	Sesgo	p-valor	M	DT	Sesgo	p-valor
1	3.305	0.139	0.741	$4 \cdot 10^{-4}$	2.605	0.069	0.041	0.324
2	2.669	0.129	0.720	0.622	1.984	0.063	0.035	0.126
3	2.782	0.145	0.961	0.079	1.897	0.119	0.076	0.001
4	3.393	0.206	1.319	0.001	2.068	0.113	-0.005	0.776
5	3.394	0.206	3.086	0.001	2.068	0.113	1.760	0.777
6	3.622	0.099	0.503	0.216	3.194	0.080	0.075	0.081
7	3.045	0.121	0.481	0.006	2.667	0.071	0.103	0.010
8	3.277	0.102	0.534	0.220	2.783	0.060	0.040	0.986
9	3.735	0.080	0.457	0.340	3.458	0.108	0.068	0.283
10	2.779	0.154	0.782	0.030	2.201	0.091	0.204	10^{-4}
11	3.599	0.094	0.501	0.686	3.192	0.087	0.094	0.717
12	3.368	0.120	0.690	0.343	2.775	0.071	0.097	0.650
13	3.559	0.094	0.526	0.043	3.176	0.094	0.143	0.204
14	3.350	0.134	0.787	0.257	3.040	0.081	0.477	0.011
15	3.252	0.102	0.690	0.169	3.016	0.094	0.454	0.404
Exp(1)	2.594	0.170	0.984	0.009	1.713	0.090	0.103	0.358

Tabla 15: Resumen descriptivo longitudes para $\tau = 0.2$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

	Método W				Método SW			
	M	DT	Sesgo	p-valor	M	DT	Sesgo	p-valor
1	1.583	0.059	0.233	0.274	1.357	0.060	0.007	10^{-6}
2	1.162	0.046	0.190	0.587	0.982	0.044	0.010	0.032
3	0.868	0.048	0.243	0.035	0.667	0.055	0.042	$3 \cdot 10^{-8}$
4	1.065	0.061	0.425	0.475	0.647	0.058	0.007	$5 \cdot 10^{-7}$
5	0.311	0.011	0.160	$3 \cdot 10^{-5}$	0.149	0.007	-0.001	0.079
6	1.994	0.072	0.238	0.550	1.753	0.102	-0.003	$\cdot 10^{-4}$
7	1.476	0.052	0.126	0.518	1.353	0.044	0.003	$8 \cdot 10^{-4}$
8	1.708	0.061	0.154	0.230	1.599	0.158	0.045	10^{-4}
9	2.020	0.070	0.193	0.872	1.815	0.146	-0.011	10^{-6}
10	1.266	0.076	0.270	0.028	1.107	0.061	0.111	10^{-9}
11	1.997	0.070	0.260	0.289	1.736	0.128	-0.001	$6 \cdot 10^{-14}$
12	1.665	0.066	0.238	0.975	1.528	0.095	0.101	$4 \cdot 10^{-4}$
13	1.929	0.068	0.265	0.622	1.592	0.150	-0.071	$3 \cdot 10^{-11}$
14	1.590	0.062	0.240	0.711	1.414	0.072	0.064	$8 \cdot 10^{-9}$
15	1.576	0.061	0.227	0.780	1.484	0.045	0.135	0.309
Exp(1)	0.936	0.042	0.243	0.186	0.753	0.043	0.060	0.242

Tabla 16: Resumen descriptivo longitudes para $\tau = 0.5$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

A partir de la información reflejada en la tablas anteriores, es posible comparar los sesgos para cada método y modelo, resultando considerablemente mayores los obtenidos a través del método de suavizado granulométrico. En cambio, las desviaciones típicas no presentan diferencias tan extremas.

A continuación, representamos los qq-plots y estimaciones tipo núcleo de la densidad correspondientes a las longitudes de los estimadores para ciertos modelos concretos.

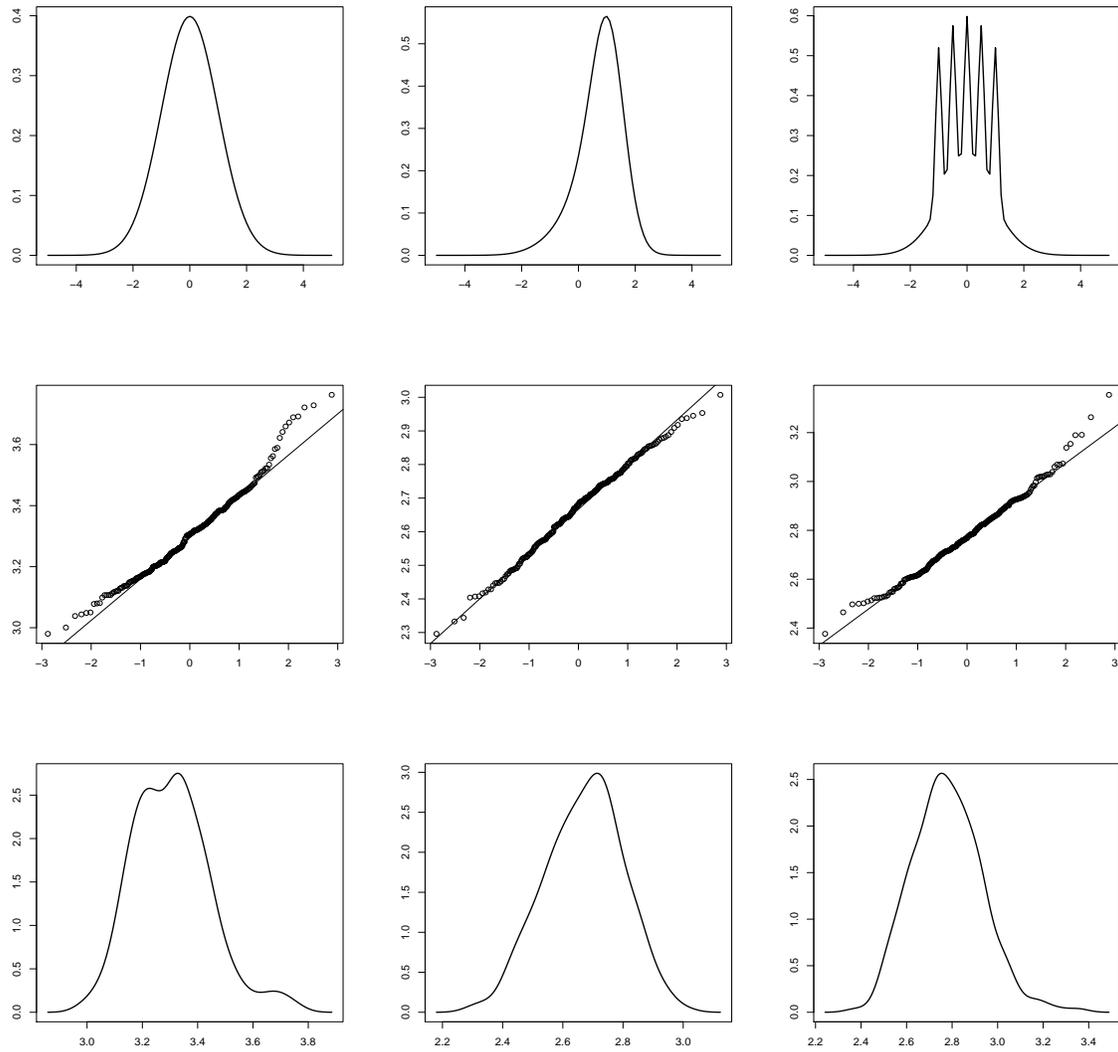


Figura 18: Qq-plots y estimación NP densidad longitud Walther $\tau = 0.2$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

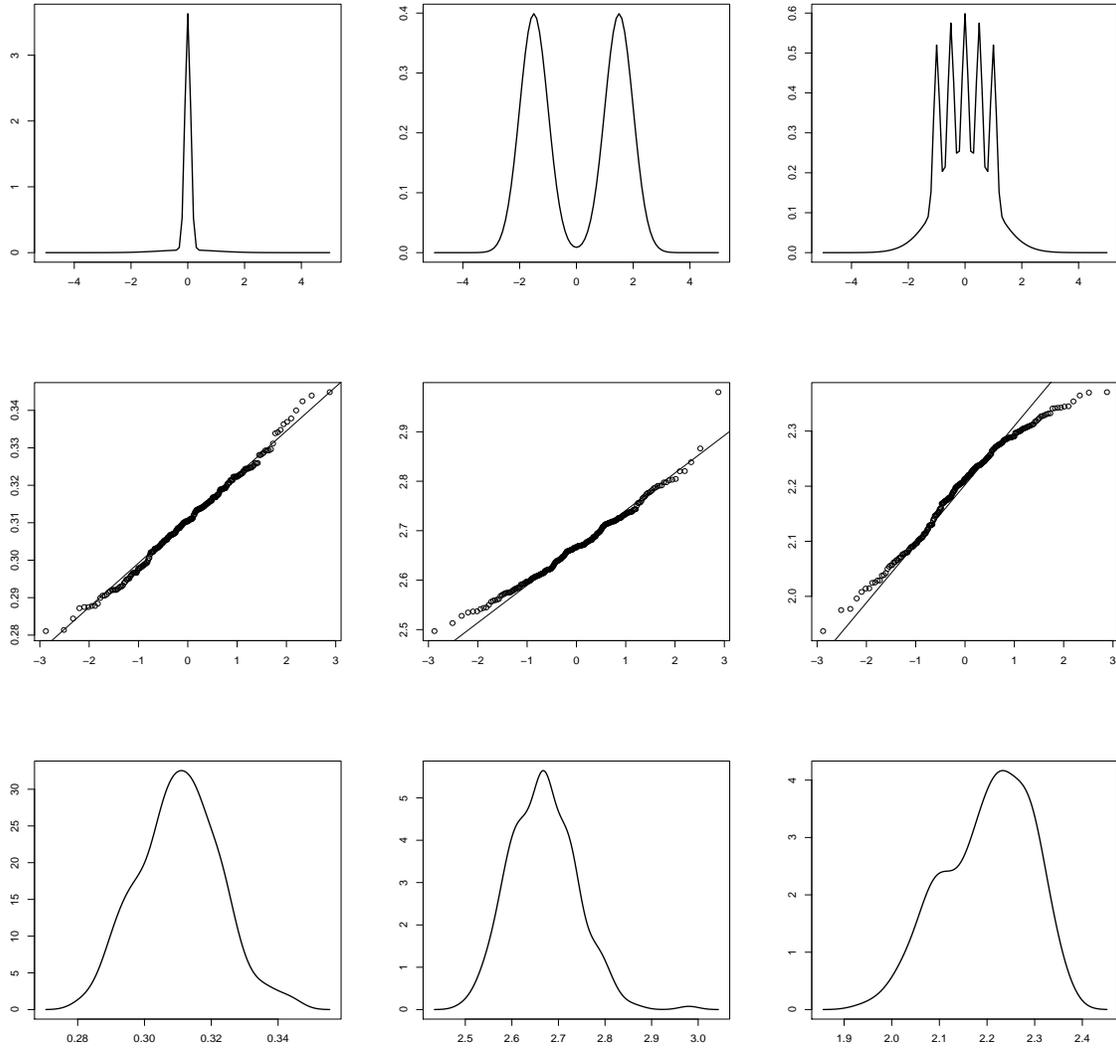


Figura 19: Qq-plots y estimación NP densidad longitud SW $\tau = 0.2$, $M = 250$, $n = 1000$ y $r_n = 0.02$.

6. Apéndice

6.1. Estimación tipo núcleo función densidad

El estimador no paramétrico tipo núcleo de una densidad f unidimensional de una variable aleatoria X se define como

$$f_n(t) = \frac{1}{n} \sum_{i=1}^n K_h(t - X_i),$$

donde $K_h(z) = (1/h)K(z/h)$, $K : \mathbb{R} \rightarrow \mathbb{R}$ función de densidad simétrica fija, llamada núcleo y h , parámetro ventana o de suavizado. A continuación se muestran algunos núcleos utilizados en la práctica:

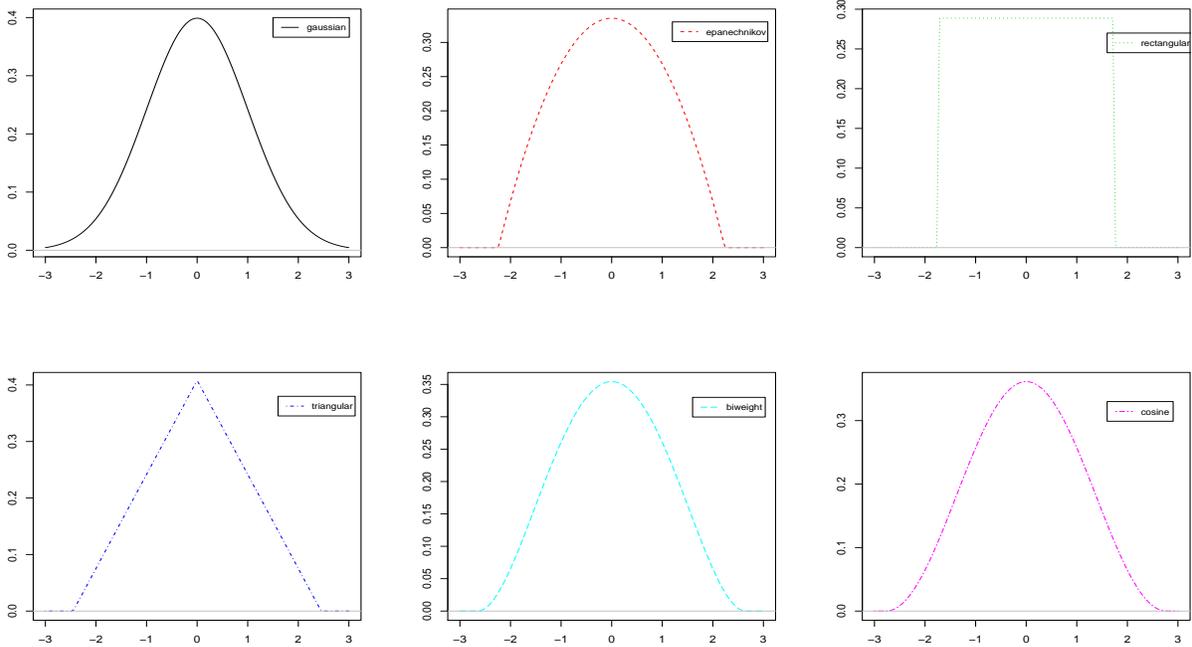


Figura 20: Funciones núcleo.

En general, el estimador tipo núcleo hereda las propiedades de suavidad del núcleo. La forma del núcleo no tiene un papel tan determinante como la ventana h . Sin embargo, el efecto del parámetro de suavizado resulta fundamental para abordar la estimación de la densidad.

Para cada punto t fijo, $\hat{f}_n(t)$ es una variable aleatoria a la que podemos medir su calidad como estimador de $f(t)$ puntualmente utilizando el Error Cuadrático Medio (MSE):

$$MSE(t) = \mathbb{E}(f_n(t) - f(t))^2,$$

que podemos descomponer en sesgo al cuadrado y varianza:

$$MSE(t) = (\mathbb{E}(\hat{f}_{n,K}(t)) - f(t))^2 + Var(\hat{f}_n(t)).$$

A continuación, exponemos los principales métodos de selección del parámetro ventana, h :

Minimización Error Cuadrático Medio Integrado Asintótico. Para medir el error de forma global, es posible considerar el Error Cuadrático Medio Integrado (MISE), integrando el Error Cuadrático Medio en t ,

$$MISE(h) = \frac{1}{nh}R(K) + \frac{1}{4}h^2\mu_2(K)^2R(f'') + o((nh)^{-1} + h^4),$$

donde $\mu_2(K) = \int u^2K(u) du < \infty$, $R(g) = \int g^2(t) dt$ y $R(f'') < \infty$ o su versión asintótica,

$$AMISE(h) = \frac{1}{nh}R(K) + \frac{1}{4}h^2\mu_2(K)^2R(f'').$$

Así, una posible elección del parámetro ventana pasa por considerar

$$h_{AMISE} = \left(\frac{R(K)}{n\mu_2(k)^2R(f'')} \right)^{1/5}$$

resultado de considerar la minimización del AMISE.

Validación cruzada. El método de validación cruzada selecciona como parámetro de suavizado el valor h que minimiza la función siguiente:

$$CV(h) = \frac{1}{n^2h} \sum_{i,j} K * K \left(\frac{X_i - X_j}{h} \right) - \frac{4}{n(n-1)} \sum_{i=1}^n \sum_{j>i} K_h(X_i - X_j)$$

donde $K * K(t) = \int K(y)K(t-y) dy$ denota la convolución de K con K .

Seather-Jones. La ventana obtenida al minimizar el Error Cuadrático Medio Integrado,

$$h_{AMISE} = \left(\frac{R(K)}{n\mu_2(k)^2R(f'')} \right)^{1/5},$$

depende de $R(f'')$. Bajo la suposición $R(f'') < \infty$, condiciones de regularidad sobre f y aplicando integración por partes, se puede demostrar que:

$$\int (f''(t))^2 dt = \int f^4(t)f(t) dt = \psi_4.$$

Así, Jones y Seather (1991) propusieron estimar $\psi_r = \mathbb{E}(f^r)^{(X)}$ como:

$$\hat{\psi}_r(g) = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n L_g^r(X_i - X_j)$$

donde L es una función núcleo y g , una ventana que podría ser seleccionada (Wand y Jones, 1995) según:

$$g = \left(\frac{-2K^r(0)}{\mu_2(K)n\psi_{r+2}} \right)^{1/(r+3)}.$$

Así, volviendo al problema inicial se reemplaza ψ_4 por su estimador y utilizando la ventana g propuesta se obtendría

$$\hat{h} = \left(\frac{R(K)}{n\mu_2(k)^2g} \right)^{1/5}$$

donde

$$g = \left(-\frac{2K^4(0)}{\mu_2(k)n\psi_6} \right)^{1/7}.$$

La expresión anterior vuelve a depender de ψ_6 , y repitiendo el proceso, obtenemos ψ_8 ... Normalmente, se para después de dos etapas. En la última, se reemplaza el valor de ψ_r por lo que valdría para la densidad Normal. Así, en el caso de una Normal con desviación σ , para r par se tiene:

$$\psi_r = \frac{(-1)^{r/2}r!}{(2\sigma)^{r+1}(r/2)!\sqrt{\pi}}.$$

6.2. Densidades de Marron y Wand

Las quince densidades descritas en Marron y Wand (1992) poseen la peculiaridad de poder ser escritas como *normal mixtures*. Estas densidades se han convertido en un estándar para validar cualquier método de estimación de la densidad. Enumeremos el conjunto de las 15 densidades contenidas en la librería *normix* de R:

Densidad	$\omega_1 N(\mu_1, \sigma_1^2) + \dots + \omega_k N(\mu_k, \sigma_k^2)$
1 Gaussian	$N(0, 1)$
2 Skewed unimodal	$\frac{1}{5}N(0, 1) + \frac{1}{5}N\left(\frac{1}{2}, \left(\frac{2}{3}\right)^2\right) + \frac{3}{5}N\left(\frac{13}{12}, \left(\frac{5}{9}\right)^2\right)$
3 Strongly skewed	$\sum_{i=0}^7 \frac{1}{8}N\left(3\left\{\left(\frac{2}{3}\right)^i - 1\right\}, \left(\frac{2}{3}\right)^{2i}\right)$
4 Kurtotic unimodal	$\frac{2}{3}N(0, 1) + \frac{1}{3}N\left(0, \left(\frac{1}{10}\right)^2\right)$
5 Outlier	$\frac{1}{10}N(0, 1) + \frac{9}{10}N\left(0, \left(\frac{1}{10}\right)^2\right)$
6 Bimodal	$\frac{1}{2}N\left(-1, \left(\frac{2}{3}\right)^2\right) + \frac{1}{2}N\left(1, \left(\frac{2}{3}\right)^2\right)$
7 Separated Bimodal	$\frac{1}{2}N\left(-\frac{3}{2}, \left(\frac{1}{2}\right)^2\right) + \frac{1}{2}N\left(\frac{3}{2}, \left(\frac{1}{2}\right)^2\right)$
8 Skewed Bimodal	$\frac{3}{4}N(0, 1) + \frac{1}{4}N\left(\frac{3}{2}, \left(\frac{1}{2}\right)^2\right)$
9 Trimodal	$\frac{9}{20}N\left(-\frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{9}{20}N\left(\frac{6}{5}, \left(\frac{3}{5}\right)^2\right) + \frac{1}{10}N\left(0, \left(\frac{1}{4}\right)^2\right)$
10 Claw	$\frac{1}{2}N(0, 1) + \sum_{l=0}^4 \frac{1}{10}N\left(l/2 - 1, \left(\frac{1}{10}\right)^2\right)$
11 Double claw	$\frac{49}{100}N\left(-1, \left(\frac{2}{3}\right)^2\right) + \frac{49}{100}N\left(1, \left(\frac{2}{3}\right)^2\right) + \sum_{l=0}^6 \frac{1}{350}N\left((l-3)/2, \left(\frac{1}{100}\right)^2\right)$
12 Asymmetric claw	$\frac{1}{2}N(0, 1) + \sum_{l=-2}^2 (2^{1-l}/31)N\left(l + \frac{1}{2}, (2^{-l}/10)^2\right)$
13 Asymmetric double claw	$\sum_{l=0}^1 \frac{46}{100}N\left(2l - 1, \left(\frac{2}{3}\right)^2\right) + \sum_{l=1}^3 \frac{1}{300}N\left(-l/2, \left(\frac{1}{100}\right)^2\right) + \sum_{l=1}^3 \frac{7}{100}N\left(l/2, \left(\frac{7}{100}\right)^2\right)$
14 Smooth Comb	$\sum_{l=0}^5 (2^{5l}/63)N\left(\{65 - 96\left(\frac{1}{2}\right)^l\}/21, \left(\frac{32}{63}\right)^2/2^{2l}\right)$
15 Discrete Comb	$\sum_{l=0}^2 \frac{2}{7}N\left((12l - 15)/7, \left(\frac{2}{7}\right)^2\right) + \sum_{l=8}^{10} \frac{1}{21}N\left(2l/7, \left(\frac{1}{21}\right)^2\right)$

Tabla 17: Parámetros 15 densidades de Marron y Wand.

Como primera aproximación, consideremos sus representaciones gráficas comparándolas con la distribución normal:

Densidades Marron–Wand

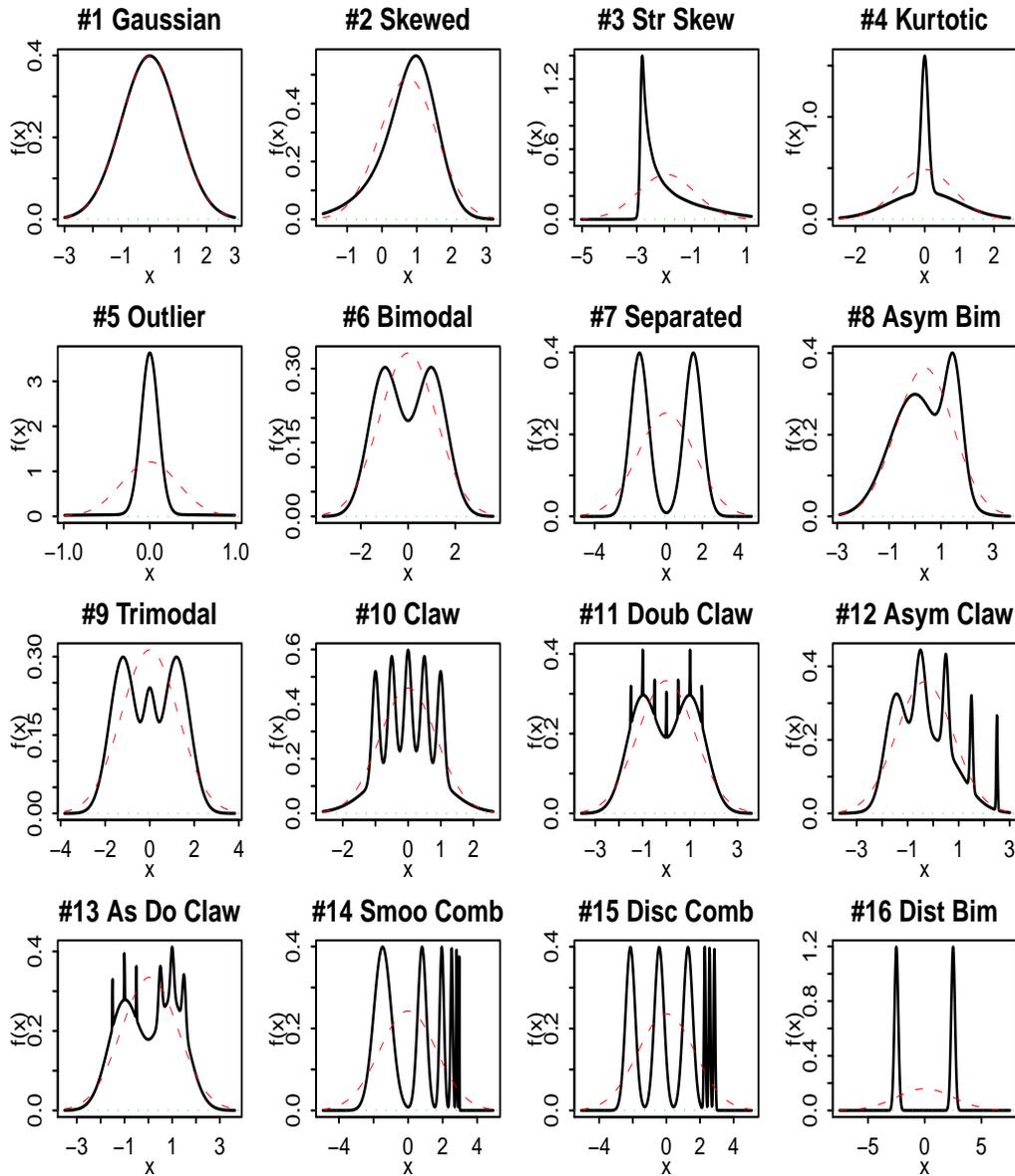


Figura 21: Densidades de Marrow y Wand

En particular, si X es una variable aleatoria distribuida según una *mixtura normal* con densidad f ,

$$f(x) = \omega_1 \phi_{\sigma_1}(x - \mu_1) + \dots + \omega_n \phi_{\sigma_n}(x - \mu_n),$$

donde

$$\mu_i, \omega_i \in \mathbb{R}, \quad i = 1, \dots, n,$$

$$\sigma_i \in \mathbb{R}^+, \quad i = 1, \dots, n,$$

$$\omega_1 + \dots + \omega_n = 1,$$

$$\phi_\sigma(x) = \frac{1}{\sigma} \phi\left(\frac{x}{\sigma}\right)$$

con ϕ denotando la densidad normal estándar. Según las consideraciones previas,

$$f(x) = \sum_{i=1}^n \omega_i \phi_{\sigma_i}(x - \mu_i).$$

Por tanto,

$$\begin{aligned} f'(x) &= \left(\sum_{i=1}^n \omega_i \phi_{\sigma_i}(x - \mu_i) \right)' = \sum_{i=1}^n \omega_i (\phi_{\sigma_i}(x - \mu_i))' \\ &= \sum_{i=1}^n \omega_i (-1) H_1 \left(\frac{x - \mu_i}{\sigma_i} \right) \phi \left(\frac{x - \mu_i}{\sigma_i} \right) \frac{1}{\sigma_i^2} = - \sum_{i=1}^n \omega_i \left(\frac{x - \mu_i}{\sigma_i^3} \right) \phi \left(\frac{x - \mu_i}{\sigma_i} \right) \end{aligned}$$

y

$$\begin{aligned} f''(x) &= \left(\sum_{i=1}^n \omega_i \phi_{\sigma_i}(x - \mu_i) \right)'' = \sum_{i=1}^n \omega_i (\phi_{\sigma_i}(x - \mu_i))'' \\ &= \sum_{i=1}^n \omega_i H_2 \left(\frac{x - \mu_i}{\sigma_i} \right) \phi \left(\frac{x - \mu_i}{\sigma_i} \right) \frac{1}{\sigma_i^3} = \sum_{i=1}^n \frac{\omega_i}{\sigma_i^3} \left[\left(\frac{x - \mu_i}{\sigma_i} \right)^2 - 1 \right] \phi \left(\frac{x - \mu_i}{\sigma_i} \right). \end{aligned}$$

Nota 6.1 Para realizar los cálculos anteriores se ha tenido en cuenta:

$$\phi_\sigma^{(r)}(x) = \frac{1}{\sigma^{r+1}} \phi^{(r)}\left(\frac{x}{\sigma}\right), \quad r = 0, 1, \dots$$

y

$$\phi^{(r)}(x) = (-1)^r H_r(x) \phi(x), \quad r = 0, 1, \dots$$

donde $H_r(x)$ denota el polinomio de Hermite de grado r .

Bibliografía

- Baddeley, A. y Molchanov, I. (1998). Averaging of random sets based on their distance functions. *Journal of Mathematical Imaging and Vision* 8, 79-92.
- Cadre, B. (2006). Kernel estimation of density level sets. *Journal of Multivariate Analysis* 97, 999-1023.
- Cadre, B., Pelletier B. y Pudlo, P. (2009). Clustering by estimation of density level sets at a fixed probability.
- Cuevas, A., Febrero, M. y Fraiman, R. (2000). Estimating the number of clusters. *Canadian Journal of Statistics* 28, 367-382.
- Cuevas, A., Fraiman, R. y Rodríguez-Casal, A. (2007). A nonparametric approach to the estimation of lengths and surface areas. *Annals of Statistics* 35, 1031-1051.
- Cuevas, A., González-Manteiga, W. y Rodríguez-Casal, A. (2006). Plug-in estimation of general level sets. *Australian & New Zealand Journal of Statistics* 48, 7-19.
- Hartigan, J.A. (1987). Estimation of a convex density contour in two dimensions. *Journal of the American Statistical Association* 82, 267-270.
- Jang, W. y Hendry, M. (2007). Cluster analysis of massive datasets in astronomy. *Statistics and Computing* 17, 253-262.
- Korostelev, A.P. y Tsybakov, A.B. (1993). *Minimax Theory of Image Reconstruction*, Lecture Notes in Statistics 82, Springer-Verlag, New York.
- Mason, D. y Polonik, W. (2009). Asymptotic normality of plug-in level set estimates. *The Annals of Applied Probability*, 19, 1108-1142.
- Molchanov, I. (1998). A limit theorem for solutions of inequalities. *Scandinavian Journal of Statistics* 25, 235-242.
- Müller, D.W. y Sawitzki, G. (1991). Excess mass estimates and tests of multimodality. *Journal of the American Statistical Association* 86, 738- 746.
- Nolan, D. (1991). The excess-mass ellipsoid. *Journal of Multivariate Analysis* 39 348-371.
- Polonik (1995). Measuring mass concentration and estimating density contour clusters-an excess mass approach. *Annals of Statistics* 23, 855- 881.
- Polonik, W. y Wang, Z. (2005). Estimation of regression contour clusters. An application of the excess mass approach to regression. *Journal of Multivariate Analysis* 94, 227-249.
- Samworth, R.J. y Wand , M. P. (2010). Asymptotics and optimal bandwidth selection for highest density region estimation. *The Annals of Statistics* 38, 1767-1792.

Simonoff, J. S. (1996). Smoothing Methods in Statistics. Springer- Verlag, New York.

Tsybakov, A.B. (1997). On nonparametric estimation of density level sets. *Annals of Statistics* 25, 948-969.

Walther, G. (1997). Granulometric smoothing. *Annals of Statistics* 25, 2273-2299.