

Estudio sobre los efectos del parámetro de suavizado en contrastes no paramétricos para k -muestras

Studying the Bandwidth Effects in Non Parametric k -Sample Tests

PABLO MARTÍNEZ-CAMBLOR^a

CIBER EPIDEMIOLOGÍA Y SALUD PÚBLICA, SUBDIRECCIÓN DE SALUD PÚBLICA DE GIPUZKOA, DONOSTIA, ESPAÑA

Resumen

Una de las principales limitaciones de las *técnicas de suavizamiento* es la necesidad de elegir un parámetro de suavizado o ventana. La influencia de este parámetro sobre los resultados obtenidos obliga a que el uso de estos métodos en inferencia sea delicado, ya que la decisión final puede verse determinada por la elección del parámetro. El objetivo principal de este trabajo es el estudio de algunos algoritmos para el cálculo automático del parámetro ventana en problemas de contrastes de hipótesis para la igualdad de k poblaciones independientes.

Palabras clave: tests no paramétricos, estimación núcleo, parámetro ventana.

Abstract

The election of the smoothing parameter or bandwidth is, probably, the most important concern in the statistical *smoothed techniques*. The relevance of this parameter, on the obtained results difficult, the use of these methods in statistical inference, because the final decision could be determined for the used bandwidth. The main goal of this paper is discussing and studying some algorithms for the automatic computation of the bandwidth in k -sample problems.

Key words: Nonparametric tests, Kernel estimation, Bandwidth.

^aInvestigador postdoctoral. E-mail: pmcamblor@hotmail.com

1. Introducción

La elección del parámetro de suavizado en el contexto de las técnicas de suavizamiento es un problema importante para el que, en muchas ocasiones, no se ha encontrado una respuesta óptima. A pesar de que hay innumerables métodos para estimar la ventana óptima en problemas de estimación de la función de densidad (por ejemplo Park & Marron 1990, Devroye 1997, entre otros) y de la función de distribución (Sarda 1993), cuando se trata de implementar estas técnicas en el contexto de la inferencia estadística la solución no parece tan clara.

Los estimadores de tipo núcleo, propuestos por Rosenblatt (1956) y Parzen (1962), son probablemente los más populares y más frecuentemente utilizados para la estimación no paramétrica de la función de densidad. Dada una muestra aleatoria simple $X = \{x_1, \dots, x_n\}$, la estimación núcleo para la función de densidad queda definida por

$$\hat{f}_h(X, t) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x_i - t}{h}\right) \quad (1)$$

donde K es una función núcleo, usualmente elegida para que sea una función de densidad simétrica de media cero y varianza finita, y $h = h_{(n)}$ es una sucesión de números reales positivos. Este estimador ha sido ampliamente estudiado y ha dado lugar a toda una serie de métodos conocidos como *técnicas de suavizamiento* (Bowman & Azzalini 2001).

Los usos de este estadístico en la inferencia son numerosos. Silverman (1981) propone un test para contrastar la multimodalidad de una distribución, posteriormente estudiado por Hall & York (2001). Ahmad & Li (1997) y Diks & Tong (1999) proponen tests de simetría para distribuciones univariantes y multivariantes respectivamente. Los tests de bondad de ajuste son, sin duda, los más extensamente tratados; autores como Ghosh & Huang (1991), Fan (1994), Liero et al. (1998) o Fan (1998) entre otros, han abordado este problema desde distintos puntos de vista. El problema de comparación de dos muestras independientes ha sido menos estudiado, aunque autores como Anderson et al. (1994), Li (1996, 1999) o Cao & Van Keilegom (2006) han propuesto tests basados en la estimación núcleo para la función de densidad. Martínez-Cambolor (2006) estudia un test para la comparación de k muestras independientes basado en el área común (\mathcal{AC}) entre las estimaciones núcleo para las respectivas densidades. Un estudio comparativo de la potencia de este test puede verse en Martínez-Cambolor et al. (2008). Una versión del mismo para muestras no independientes ha sido propuesta en Martínez-Cambolor (2008). En estos trabajos se observa como, para muestras pequeñas y de igual tamaño, el test \mathcal{AC} es más potente que los tests basados en la función de distribución empírica (FDE) cuando las distribuciones de origen no se diferencian únicamente en un parámetro de posición. En Martínez-Cambolor & De Uña-Álvarez (2008), se proponen y estudian diversos tests ómnibus basados en la estimación núcleo para la función de densidad (END), de los cuales el más potente globalmente es el basado en la medida L_1 que, para k -muestras independientes de tamaños n_i ($1 \leq i \leq k$)

y si $n = \sum_{i=1}^k n_i$ queda definido así:

$$L_{k,1} = \frac{1}{n} \sum_{i=1}^k n_i \int |\hat{f}_{h_i}(t) - \hat{f}_h(t)| dt$$

donde \hat{f}_{h_i} ($1 \leq i \leq k$) es la estimación núcleo referida a la i -ésima muestra, utilizando h_i como parámetro de suavizado, y \hat{f}_h es la estimación núcleo para la muestra conjunta, utilizando h como parámetro de suavizado.

La normalidad asintótica de la norma L_p entre la estimación núcleo y la función de densidad real, $L_p(\hat{f}_h, f)$, ha sido demostrada por Horvath (1991); las hipótesis necesarias para garantizar esta convergencia, fueron rebajadas posteriormente por Martínez-Cambor & Corral (2008). El caso $p = 1$ ha sido ampliamente estudiado por Devroye & Györfy (1985). La normalidad asintótica del estadístico $L_{k,1}$ ha sido tratada en Martínez-Cambor & De Uña-Álvarez (2008); si bien no se dan expresiones explícitas para sus parámetros de centralización, su distribución se aproxima mediante un procedimiento *bootstrap* suavizado (Hall et al. 1989) que, para un estadístico genérico \mathcal{F} , sigue este plan de remuestreo:

- A. Desde la muestra conjunta, X , se calcula la función de distribución empírica suavizada (FDES) (Nadaraya 1964) y se calcula el valor del estadístico $\mathcal{F}(X)$.
- B. Aleatoriamente, se generan muestras *bootstrap* X^b ($1 \leq b \leq B$) desde la distribución anterior, con los mismos tamaños muestrales de las muestras de origen. Para cada muestra *bootstrap* se calcula el valor del estadístico $\mathcal{F}(X^b)$ ($1 \leq b \leq B$).
- C. Se aproxima la distribución del estadístico $\mathcal{F}(X)$ a partir de los valores $\mathcal{F}(X^b)$ ($1 \leq b \leq B$).

Usualmente, el parámetro ventana utilizado para la FDES (función utilizada para generar las muestras *bootstrap* en el remuestreo), g , es diferente al utilizado en la estimación de la función de densidad. En Cao (1990) se propone usar g de la forma $g = Cn^{-1/9}$; este es el orden de convergencia que minimiza el error cuadrático medio integrado (MISE, por su sigla en inglés) para la estimación núcleo de la derivada segunda de la función de densidad. En todas las simulaciones realizadas en este trabajo, se usa g de forma que minimice el MISE de la FDES; este es de la forma $g = Cn^{-1/3}$ (Martínez-Cambor 2006) y, por simplicidad, se toma $C = 1$ en todos los casos. No obstante, simulaciones no presentadas en este trabajo, en concordancia con los resultados obtenidos por Cao & Van Keilegom (2006), sugieren que este parámetro, salvo elecciones anormalmente elevadas, tiene un efecto menor sobre el resultado final de los $\mathcal{F}(X^b)$ ($1 \leq b \leq B$).

El principal inconveniente que presenta este tipo de test es la selección del parámetro ventana, muy importante en la potencia final alcanzada (figura 1). Eggermont & LaRiccia (2003) establecen una propuesta para la elección del mismo en el contexto de bondad de ajuste. Cao & Van Keilegom (2006) estudian este problema en el ámbito de los contrastes para dos poblaciones, proponiendo un

procedimiento denotado por *doble bootstrap*. Posteriormente, Martínez-Camblor et al. (2008), generalizando a problemas con k -muestras, desarrollan un algoritmo denominado *doble mínimo* para la toma de decisiones.

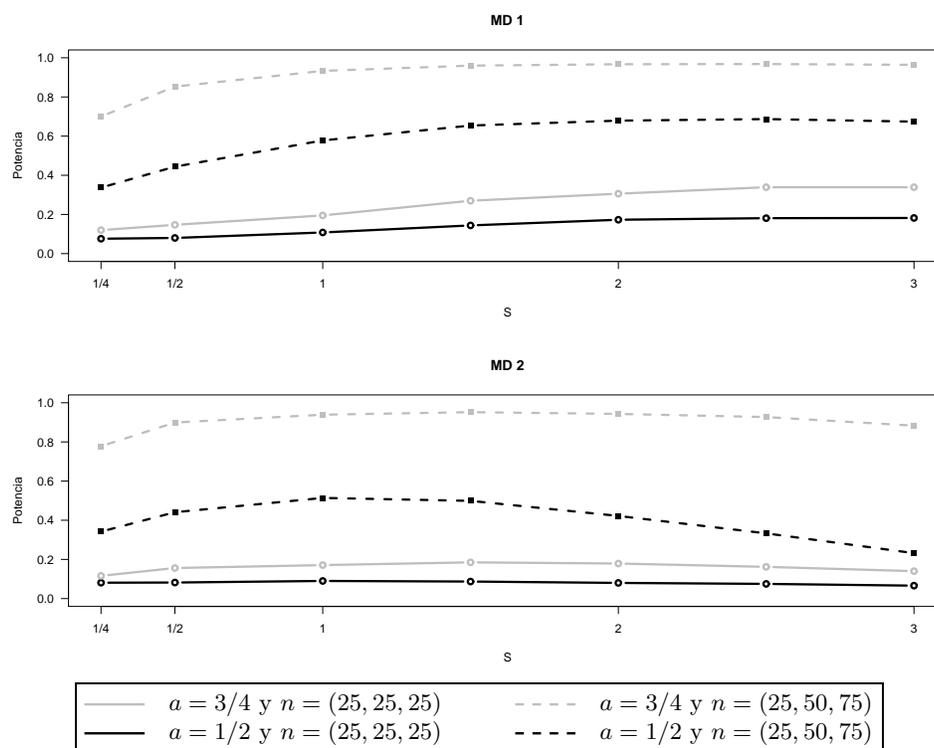


FIGURA 1: Potencias estimadas del estadístico $L_{k,1}$ para los modelos descritos en la sección 3. Modelo MD 1 (arriba) y modelo MD 2 (abajo) con dos valores distintos de a y distintos tamaños muestrales. El tamaño ventana es $S\sigma n^{-1/5}$, donde n es la suma de los tamaños muestrales y σ la desviación típica de cada muestra.

En este trabajo, en el contexto de contrastes para la comparación de k -muestras independientes, se estudia el efecto del parámetro ventana utilizado sobre el resultado final de la prueba. Se analizan los métodos *doble bootstrap* (DB) y *doble mínimo* (DM) para el cálculo automático de la ventana y se propone un nuevo método que, denotado por BM (*bootstrap mínimo*) es, de alguna manera, una mezcla de los dos métodos citados (sección 2). En la sección 3, y mediante un estudio de simulación, se examina el rendimiento de este procedimiento y se compara con el *doble bootstrap*. Finalmente, en la sección 4, se reflexiona sobre algunos problemas de este tema y se plantean algunas de sus fortalezas y sus debilidades.

2. Algoritmos *doble bootstrap*, *doble mínimo* y *BM*

En esta sección se describen y analizan dos de los algoritmos más recientemente propuestos para la selección del parámetro ventana en contrastes de igualdad para muestras independientes: el *doble bootstrap* (Cao & Van Keilegom 2006) y el *doble mínimo* (Martínez-Cambolor et al. 2008). Además, se propone un algoritmo que, denotado por BM, en el espíritu de Cao & Van Keilegom (2006), calcula el tamaño ventana basándose en las ideas de Martínez-Cambolor et al. (2008).

Los tres métodos considerados están basados en la idea de que a partir de una malla de posibles valores para el parámetro ventana, se encuentre el que mejor se adapte a cada problema concreto.

Cao & Van Keilegom (2006) observan que el porcentaje de rechazos obtenidos por el estadístico estudiado en su trabajo, cuando la hipótesis nula es cierta, es el adecuado con independencia del parámetro de suavizado utilizado. Apoyándose en este hecho, los autores desarrollan el método DB que, estimando la distribución del estadístico bajo la hipótesis alternativa, se queda con el valor de h que maximice la potencia del test para, finalmente, elegir el h como el promedio de una serie de h calculados mediante el siguiente algoritmo (se ha procurado respetar la redacción del algoritmo hecha en Cao & Van Keilegom 2006):

D_1 Sea $\mathcal{H} = \{h_1, \dots, h_T\}$ una malla de valores de h entre los cuales se quiere seleccionar el óptimo.

D_2 Para cada $b \in \{1, \dots, B\}$, se generan muestras $X_{b1}^*, \dots, X_{bn_j}^*$ desde \hat{f}_{h_j} con $j \in \{1, \dots, k\}$. Note que el objetivo es maximizar la potencia (y no la significación); por tanto, se generan muestras desde las distribuciones separadas.

D_3 Para cada $b \in \{1, \dots, B\}$ y cada $t \in \{1, \dots, T\}$:

- a) Se calcula el valor del estadístico, $\mathcal{F}_{b,t}^B(X)$ para cada muestra ($1 \leq b \leq B$) y para cada valor de h ($1 \leq t \leq T$).
- b) A partir de un nuevo *bootstrap* (segundo nivel), asumiendo H_0 cierta, se calculan valores críticos, $c(b, t)$, para cada muestra y para cada valor de h . Se utilizan los pasos A, B y C del algoritmo descrito en la sección 1.

D_4 Para cada $t \in \{1, \dots, T\}$ se calcula

$$\widehat{\text{power}}(h_t) = \frac{1}{B} \sum_{i=1}^B I\{\mathcal{F}_{b,t}^B > c(b, t)\}$$

D_5 El valor de la ventana óptima será

$$\hat{h}_{opt} = \operatorname{argm\acute{a}x}_{\{h \in \mathcal{H}\}} \{\widehat{\text{power}}(h)\}$$

Note que si se generan B_0 muestras para el segundo nivel del *bootstrap*, el número de veces que debe calcularse el estadístico en cada iteración es $T(B + B_0)$. Luego

si el parámetro ventana final se obtiene a partir del promedio de B_1 h y, la distribución final se aproxima a partir de B_2 muestras *bootstrap*, el coste computacional final (número de veces que se debe calcular el estadístico) es $TB_1(B + B_0) + B_2$. Este es uno de los principales problemas del algoritmo.

El método DM parte de la idea de que el valor de h que más separa las hipótesis nula y alternativa es el que da lugar a una significación menor. Desde este supuesto y a partir de una malla de posibles valores para h , se desarrolla un algoritmo basado en la corrección de la significación inicial mediante un nuevo *bootstrap*. Su algoritmo es el siguiente (Martínez-Cambolor et al. 2008):

- M_1 Se elige una rejilla de posibles valores de $\mathcal{H} = \{h_1, \dots, h_T\}$, entre los cuales se va a buscar la potencia óptima.
- M_2 Desde el plan de remuestreo descrito anteriormente (pasos A, B y C), se calcula la significación del test para cada $h_t : p_t$, con $1 \leq t \leq T$.
- M_3 El h_t elegido será el que minimice las significaciones anteriores, esto es, la significación obtenida será $p_M = \min\{p_1, \dots, p_T\}$.
- M_4 Se obtiene una muestra *bootstrap* bajo la hipótesis alternativa y, sobre ella se repiten los pasos M_1 , M_2 y M_3 . Se repite este proceso un número determinado de veces B y se obtienen los valores $D_B = \{p_M^1, \dots, p_M^B\}$.
- M_5 Se *corrige* el valor de p_M . La significación final obtenida será

$$p_F = \frac{1}{B} \sum_{i=1}^B I\{p_M > p_M^i\}$$

Realmente, este método no proporciona una elección del parámetro ventana sino que da una significación final para el contraste de hipótesis. Su coste computacional (número de veces que hay que calcular el estadístico), asumiendo que se realiza el mismo número de repeticiones que en el proceso anterior, sería $TB_0(B+1)$ aproximadamente, la mitad del coste computacional del algoritmo DB. Tiene como principal inconveniente que, al estimar la significación mediante un número finito de remuestras *bootstrap*, se produce cierto número de empates.

La filosofía del método DM puede adaptarse fácilmente a la elección de un parámetro ventana siguiendo el algoritmo descrito por el procedimiento DB. Para ello, basta elegir el valor de h como la media de aquellos que hacen mínima la significación. Con los supuestos anteriores, el coste computacional del cálculo del h mediante este método que, por ser una mezcla de los dos algoritmos anteriores, se denotará por BM, es $TBB_0 + B_2$. Su algoritmo es el siguiente:

- I_1 Se elige una rejilla de posibles valores para el parámetro ventana, $\mathcal{H} = \{h_1, \dots, h_T\}$, entre los que se va a buscar la potencia óptima.
- I_2 Desde el plan de remuestreo descrito anteriormente (pasos A, B y C), se calcula la significación del test para cada $h_t : p_t$, con $1 \leq t \leq T$.

- I_3 El h_t elegido será el que minimice las significaciones anteriores, esto es, $h_B = \text{argmín}\{p_1, \dots, p_T\}$.
- I_4 Se repiten los pasos I_1, I_2 e I_3 un número determinado de veces B , obteniéndose los valores $H_B = \{h_B^1, \dots, h_B^B\}$.
- I_5 El valor de h que se usará es

$$h_{BM} = \frac{1}{B} \sum_{b=1}^B h_B^b$$

Note que, en los desarrollos teóricos de cualquier resultado relacionado con la END, se exige que el valor del parámetro ventana converja a cero, esto es, $h_n \rightarrow_n 0$. Esta condición no se verifica si se debe elegir entre un malla fija de posibles valores de h ; por este motivo, al igual que en Martínez-Cambolor et al. (2008), se utilizan parámetros ventana de la forma $S\hat{\sigma}n^{-1/5}$ ($\hat{\sigma}$ es la raíz cuadrada de la varianza muestral y n el tamaño de cada muestra) y se elige entre una malla de posibles valores de S .

3. Estudio de simulación

En esta sección, a través de un estudio de simulación de Monte Carlo, se analizan las potencias obtenidas (para $\alpha = 0.05$) por el método DB y por el BM en dos modelos diferentes. Además, con la intención de observar el efecto que sobre los resultados finales tiene la elección de la malla considerada, se estudian dos situaciones. Cabe recordar que, en todas las ocasiones, el parámetro utilizado para el remuestreo (parámetro g) es de la forma $n^{-1/3}$, siendo $n = n_1 + n_2 + n_3$.

La potencia de los métodos anteriormente descritos se estudia en dos modelos simétricos (modelos similares han sido considerados por Cao & Van Keilegom 2006, Martínez-Cambolor et al. 2008). En ellos, se generan dos muestras aleatorias de tamaños n_1 y n_2 desde una distribución normal estandarizada y una tercera muestra de tamaño n_3 desde cada una de las distribuciones siguientes:

MD 0 : $Z \equiv N(0, 1)$ (hipótesis nula)

MD 1 : $Z \equiv (1 - a)N(0, 1) + aN(0, 2)$ para $a = 1/2$ y $a = 3/4$.

MD 2 : $Z \equiv (1 - a)N(0, 1) + aN(1, 1)$ para $a = 1/2$ y $a = 3/4$.

donde $N(\mu, \sigma)$ representa una distribución normal de media μ y varianza σ^2 . En la figura 2 puede verse una representación gráfica de las densidades de las variables aleatorias consideradas.

En la tabla 1, se muestran los resultados obtenidos cuando las mallas consideradas son $\mathcal{H}_1 = \{1/4, 1/2, 1\}$ y $\mathcal{H}_2 = \{1, 2, 3\}$. Las distribuciones bajo la hipótesis nula y alternativa se aproximan a partir de 100 réplicas. Se repite el proceso 100 veces (el h a usar será el promedio de esos 100 valores); la distribución final para el h elegido se aproxima mediante 199 simulaciones *bootstrap*. Se generan 500 muestras de cada problema y se da la proporción de rechazos.

Reseñe que los valores de S óptimos entre los casos estudiados son el uno y el dos para el problema uno y $n = (25, 25, 25)$; y entre dos y tres para $n = (25, 50, 75)$. En el problema dos, son $1/2$ y 1 para $n = (25, 25, 25)$ pasando a 1 y 2 para $n = (25, 50, 75)$. También destaca el bajo porcentaje de rechazos observados para $n = (25, 50, 75)$ para MD 0 (hipótesis nula cierta) en todos los valores considerados para h . El DB obtiene mejores resultados que el BM en todos los casos, si bien estas diferencias son pequeñas (en media del 5.05%, 4.75% en los modelos MD 1 y 5.35% en los modelos MD 2). Sin embargo, un análisis más detallado de los resultados muestra que, mientras que el algoritmo BM obtiene resultados muy próximos al óptimo de los de su malla, el DB obtiene resultados sensiblemente mejores en la mayoría de los casos. Este hecho, que puede ser visto de forma positiva, tiene la contrapartida de que, cuando la hipótesis nula es cierta, también se obtiene un porcentaje de rechazo mayor del esperado. En concreto, el tamaño estimado de las pruebas es siempre mayor que el mayor de estos tamaños en las mallas, llegando a estar 1.8% por encima para $n = (25, 25, 25)$ y la malla \mathcal{H}_1 .

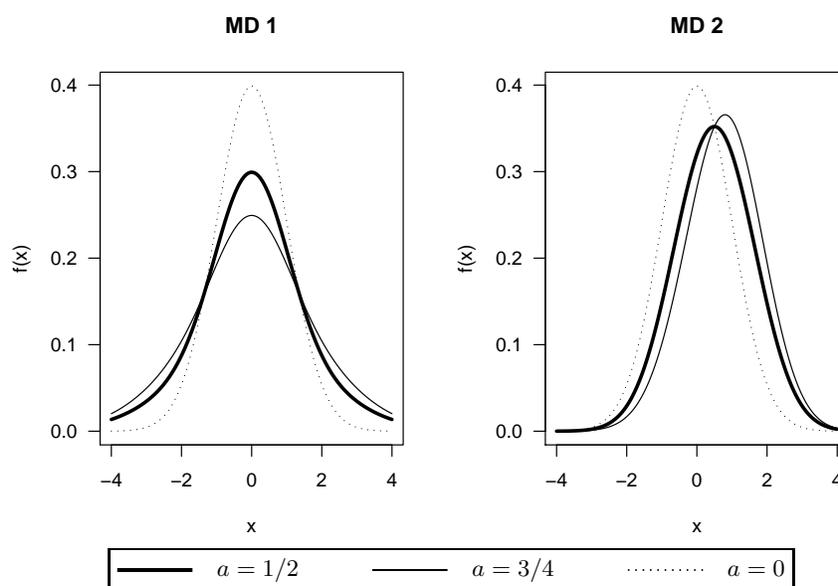


FIGURA 2: Representación gráfica de los modelos para los valores de $a = 1/2$, $a = 3/4$ y $a = 0$.

4. Discusión y conclusiones

El problema de elegir el parámetro ventana en la comparación de k -poblaciones independientes pasa por elegir el mejor entre una malla de posibles valores. La teoría plantea que el valor de h también viene determinado por la variabilidad de las

TABLA 1: Proporción de rechazos obtenidos por el estadístico $L_{k,1}$ en los modelos propuestos para para $n = (n_1, n_2, n_3)$, $\mathcal{H}_1 = \{1/4, 1/2, 1\}$ y $\mathcal{H}_2 = \{1, 2, 3\}$

		$n = (25, 25, 25)$									
		S					\mathcal{H}_1		\mathcal{H}_2		
		1/4	1/2	1	2	3	BM	DB	BM	DB	
MD 0		0.052	0.060	0.056	0.054	0.052	0.064	0.078	0.052	0.074	
MD 1	$a = 1/2$	0.186	0.220	0.310	0.334	0.282	0.286	0.292	0.338	0.372	
	$a = 3/4$	0.326	0.426	0.524	0.556	0.494	0.450	0.500	0.544	0.596	
MD 2	$a = 1/2$	0.186	0.246	0.258	0.214	0.206	0.254	0.284	0.222	0.288	
	$a = 3/4$	0.352	0.420	0.499	0.416	0.258	0.468	0.582	0.402	0.542	

		$n = (25, 50, 75)$									
		S					\mathcal{H}_1		\mathcal{H}_2		
		1/4	1/2	1	2	3	BM	DB	BM	DB	
MD 0		0.046	0.026	0.029	0.030	0.028	0.028	0.049	0.032	0.042	
MD 1	$a = 1/2$	0.338	0.436	0.580	0.672	0.622	0.488	0.544	0.664	0.692	
	$a = 3/4$	0.652	0.782	0.929	0.945	0.932	0.792	0.934	0.948	0.960	
MD 2	$a = 1/2$	0.402	0.528	0.610	0.612	0.466	0.564	0.572	0.606	0.644	
	$a = 3/4$	0.768	0.874	0.942	0.942	0.886	0.888	0.888	0.936	0.968	

variables consideradas y el tamaño muestral disponible; por tanto es aconsejable elegir la malla de posibles valores teniendo en cuenta estas consideraciones. Desde este punto de partida, se debe fijar un criterio para decidir cuál de los valores de la malla seleccionar. Dado que el objetivo es encontrar el test estadísticamente más potente, parece lógico elegir el que, para un tamaño muestral y una significación dadas, maximice la potencia; esto es, dado un estadístico (dependiente del parámetro de suavizado) y una malla de posibles valores para h , \mathcal{H} , para un nivel de significación α se elegirá: $h = \operatorname{argm\acute{a}x}_{h \in \mathcal{H}} \mathcal{P}_{\mathcal{H}_1} \{ \mathcal{F}_h > \lambda_h \} = \beta_h$, donde λ_h es el valor que verifica que $\mathcal{P}_{\mathcal{H}_0} \{ \mathcal{F}_h > \lambda_h \} = 1 - \alpha$.

En el DB, se utiliza un método *bootstrap* para calcular los valores λ_h (note que λ_h coincide con $c(b, t)$) y un nuevo *bootstrap* para calcular los β_h (note que β_h coincide con $\widehat{\text{power}}(h_t)$) así se obtiene un h de cada iteración. Este proceso se realiza determinado número de veces para, finalmente, usar el promedio de los valores obtenidos.

El método BM, partiendo de la misma idea, aplica una única realización del estadístico bajo la hipótesis alternativa, comprueba cómo es de *creíble* bajo la hipótesis nula (P -valor) y se queda con el valor de h que hace que los valores del estadístico bajo la hipótesis alternativa sean poco creíbles cuando la hipótesis nula es cierta. Este método tiene la ventaja de, por un lado, ahorrarse la estimación de la distribución bajo la hipótesis alternativa y, por otro, no estar sujeto a un tamaño del test prefijado. Su principal inconveniente es que está *menos pegado* a los datos, por lo que consigue potencias más pequeñas. Este inconveniente se convierte en ventaja cuando la hipótesis nula es cierta, ya que su tamaño tiende a coincidir con el óptimo de los tamaños de la malla (α), mientras que los resultados observados sugieren que, en el procedimiento DB, el tamaño del test es sensiblemente superior al óptimo de la malla.

En general, los resultados obtenidos por ambos métodos son buenos y hacen pensar que se está cerca de una solución que permita generalizar el uso de este tipo de tests que han demostrado ser mucho más potentes que los clásicos cuando las diferencias entre las distribuciones de origen se centran en la forma y no en un parámetro de localización. Sin embargo, y como aspecto negativo, parece que la elección de la malla representa un papel relativamente importante y fruto de estudio. Queda por tanto, como objetivo de estudio, el comportamiento de estos algoritmos ante distintos tipos de mallas (más largas, más cortas, ...) así como, por supuesto, la reducción de su coste computacional.

Agradecimientos

El autor desea mostrar su agradecimiento a los tres árbitros anónimos por su exhaustiva revisión de este documento, así como por sus valiosos comentarios, los cuales han servido para mejorar este trabajo.

[Recibido: febrero de 2008 — Aceptado: octubre de 2008]

Referencias

- Ahmad, A. I. & Li, Q. (1997), ‘Testing Symmetry of an Unknown Density Function by Kernel Method’, *Journal of Nonparametric Statistics* **7**, 279–293.
- Anderson, N. H., Hall, P. & Titterton, D. M. (1994), ‘Two-Sample Test Statistics for Measuring Discrepancies between two Multivariate Probability Density Functions using Kernel-Based Density Estimates’, *Journal of Multivariate Analysis* **50**, 41–54.
- Bowman, A. & Azzalini, A. (2001), *Applied Smoothing Techniques for Data Analysis*, Oxford University Press, Oxford, United Kingdom.
- Cao, R. (1990), Aplicaciones y nuevos resultados del método *Bootstrap* en la estimación no paramétrica de curvas, Tesis doctoral, Universidad de Santiago de Compostela.
- Cao, R. & Van Keilegom, I. (2006), ‘Empirical Likelihood Tests for Two-Sample Problems via Nonparametric Density Estimation’, *Canad. J. Statist.* **34**, 61–77.
- Devroye, L. (1997), ‘Universal Smoothing Factor Selection in Density Estimation: Theory and Practice’, *Test* **6**(2), 223–320.
- Devroye, L. & Györfy, L. (1985), *Nonparametric Density Estimation: The L_1 View*, John Wiley & Son, New York, United States.
- Diks, D. & Tong, H. (1999), ‘A Test for Symmetries of Multivariate Probability Distributions’, *Biometrika* **86**(3), 605–614.

- Eggermont, P. P. B. & LaRiccia, V. N. (2003), *Selecting the Smoothing Parameter in Goodness of Fit Testing*. Consultado el 04/06/08 en la web:
*www.udel.edu/FREC/eggermont/Preprints/smoselnew.pdf
- Fan, Y. (1994), 'Testing the Goodness of Fit of a Parametric Density Function by Kernel Method', *Econometric Theory* **10**, 316–356.
- Fan, Y. (1998), 'Goodness-of-fit Tests Based on Kernel Density Estimators with Fixed Smoothing Parameters', *Econometric Theory* **14**, 604–621.
- Ghosh, B. K. & Huang, W. M. (1991), 'The Power and Optimal Kernel of the Bickel-Rosenblatt Test for Goodness of Fit', *Annals of Statistics* **19**(2), 999–1008.
- Hall, P., DiCiccio, J. T. & Romano, J. P. (1989), 'On Smoothing and the Bootstrap', *Annals of Statistics* **17**(2), 692–704.
- Hall, P. & York, M. (2001), 'On the Calibration or Silverman's Test for Multimodality', *Statistica Sinica* **11**, 516–536.
- Horvath, L. (1991), 'On L_p -norms of Multivariate Density Estimations', *Annals of Statistics* **19**(4), 1933–1949.
- Li, Q. (1996), 'Nonparametric Testing of Closeness Between two Unknown Distributions Functions', *Econometric Review* **15**(3), 216–274.
- Li, Q. (1999), 'Nonparametric Testing the Similarity of two Unknown Density Functions: Local Power and Bootstrap Analysis', *Journal of Nonparametric Statistics* **11**, 189–213.
- Liero, H., Läuter, H. & Konakov, V. (1998), 'Nonparametric versus Parametric Goodness of Fit', *Statistics* **31**, 115–149.
- Martínez-Camblor, P. (2006), *Tests no paramétricos basados en una distancia entre funciones de densidad*, Servicio de Publicaciones de la Universidad de Oviedo, Oviedo, España.
- Martínez-Camblor, P. (2008), 'Test de hipótesis para contrastar la igualdad entre k -poblaciones', *Revista Colombiana de Estadística* **31**(1), 1–18.
- Martínez-Camblor, P. & Corral, N. (2008), 'Weaker Conditions for Asymptotic Approximation to L_p -norms of the Kernel Estimators', *InterSTAT Journal* **June**, 1–18.
- Martínez-Camblor, P., De Uña, J. & Corral, N. (2008), ' k -Sample Test Based on the Common Area of Kernel Density Estimator', *Journal of Statistical Planning and Inference* **138**(12), 4006–4020.
- Martínez-Camblor, P. & De Uña-Álvarez, J. (2008), Nonparametric k -sample Tests: Density Function vs. Distribution Function, Discussion Papers in Statistics and Operation Research Report 08/09, Universidade de Vigo, Dpto. de Estadística e Investigación Operativa.

- Nadaraya, E. A. (1964), 'Some New Estimates for Distribution Functions', *Theory Prob. Applic.* **9**, 497–500.
- Park, B. U. & Marron, J. S. (1990), 'Comparison of Data-Dirven Bandwidth Selectors', *Journal of American Statistics Association* **85**(409), 66–72.
- Parzen, E. (1962), 'On Estimation of a Probability Density Function and Mode', *Annals of Mathematical Statistics* **33**, 832–837.
- Rosenblatt, M. (1956), 'Remarks on Some Nonparametric Estimates of a Density Functions', *Ann. Math. Statistics* **27**, 832–837.
- Sarda, P. (1993), 'Smoothing Parameter Selection for Smooth Distribution Function', *Journal of Statistical Planning and Inference* **35**, 65–75.
- Silverman, B. W. (1981), 'Using Kernel Density Estimation to Investigate Multimodality', *Journal of the Royal Statistics Society* **B**(43), 97–99.