TRABAJOS DE ESTADISTICA
Vol. 3. Núm. 2, 1988 pp. 177 a 194

USO DEL ESTADISTICO D_n DE KOLMOGOROV-SMIRNOV EN INFERENCIA PARAMETRICA

FELIPE ORTEGA, ANGEL
Dpto. Estadística e Investigación Operativa
Facultad de Ciencias Matemáticas
Universidad Complutense de Madrid

RESUMEN

Se estudia un método de estimación paramétrica basado en la minimización del estadístico D_n de Kolmogorov-Smirnov. Se prueba la existencia y unicidad de este estimador en familias de distribuciones monótonas en alguno de sus parámetros y se compara computacionalmente con el método de máxima verosimilitud.

Palabras clave: Estadístico D_n. Estimación paramétrica. Minimización univariable.
Clasificación AMS: 62F10, 65U05, 90C30.

SUMMARY

We study a method of parametric estimation which is based in the minimization of D_n Komogorov-Smirnov's statistic. It is shown the existence and the unicity of that estimator on families of monotone distributions on some of its parameters. We compare computationally this method with the maximum likelihood one.

Key words: D_n Statistic. Parametric Estimation. Univariate Minimization.
AMS classification: 62F10, 65U05, 90C30.

Revisado, septiembre 1988.

1. INTRODUCCION

Dada una muestra $(x_1, ..., x_n)$ de una variable aleatoria X con función de distribución $F(x, \vartheta)$ donde $\vartheta \in \Theta \subset \mathbb{R}^n$ es el vector de parámetros, el problema de estimación paramétrica consta de dos fases: en la primera se estima mediante $\hat{\theta}$ el parámetro ϑ, en la segunda se contrasta la hipótesis nula H_0 de que X tiene distribución $F(x, \hat{\theta})$.

El proceso de obtención de $\hat{\theta}$ consiste en encontrar el vector $\hat{\vartheta} \in \Theta$ que maximiza la función de verosimilitud, o resuelve un sistema de ecuaciones o minimiza una suma de cuadrados, etc., según el método de estimación utilizado.

En la fase de contraste además de los tests de hipótesis paramétricos también pueden utilizarse técnicas no paramétricas como el test de ajuste de la χ-cuadrado o el estadístico D_n de Kolmogorov-Smirnov.

El estadístico D_n no se considera apto para estimar los parámetros de una distribución («The chi-square test, on the other hand, can be easily modified to allow estimation of parameters from the data, but the Kolmogorov-Smirnov test does not have this flexibility», Rohatgi (1976), pág. 543). Sin embargo, dado que el valor de D_n no sólo indica si se debe aceptar o rechazar la hipótesis H_0 a nivel α sino que además es una medida de la calidad del estimador $\hat{\theta}$, su minimización conduce a una mejora del ajuste. Por tanto, cabe considerar la minimización de D_n como un método válido de estimación paramétrica análogo al de máxima verosimilitud.

Este procedimiento es un caso particular de los métodos de distancia mínima introducidos por Wolfowitz (1957). En Parr (1981) se recoge una extensa bibliografía sobre estos métodos.

El propósito de este trabajo es comparar computacionalmente esta metodología con el método clásico de máxima verosimilitud. En la sección 2 se dan las definiciones básicas y se estudia el problema de estimación en distribuciones monótonas. En la sección 3 se describen las distribuciones consideradas en el estudio computacional que se expone en la sección 4.

2. MINIMIZACION DE D_n

Sea $(x_1, ..., x_n)$ una m.a.s. de una v.a. X con función de distribución $F(x, \vartheta)$, continua en $x \in R$, $\forall\, \vartheta \in \Theta \subset \mathbb{R}^m$ fijo y continua en $\vartheta \in \Theta$, $\forall\, x \in \mathbb{R}$ fijo. Utilizamos la siguiente notación:

$$D_n^+(\vartheta) = \sup_{x \in \mathbb{R}} \{F_n^*(x) - F(x, \vartheta)\}$$

$$D_n^-(\vartheta) = \sup_{x \in \mathbb{R}} \{F(x, \vartheta) - F_n^*(x)\}$$

$$D_n(\vartheta) = \sup_{x \in \mathbb{R}} \{|F_n^*(x) - F(x, \vartheta)|\} = \text{máx} \{D_n^+(\vartheta), D_n^-(\vartheta)\}$$

Se verifica:

$$D_n^+(\vartheta) = \text{máx} \left\{0, \text{máx} \left\{\frac{i}{n} - F(x_{(i)}, \vartheta) \middle/ i = 1, ..., n\right\}\right\}$$

$$D_n^-(\vartheta) = \text{máx} \left\{0, \text{máx} \left\{F(x_{(i)}, \vartheta) - \frac{i-1}{n} \middle/ i = 1, ..., n\right\}\right\}$$

donde $x_{(i)}$ son los estadísticos de orden y $F_n^*(x)$ la función de distribución muestral.

Definición 1

Llamamos estimador de mínimo D_n (ED_n) al parámetro $\hat{\vartheta} \in \Theta$ tal que $D_n(\hat{\vartheta}) = \text{mín} \{D_n(\vartheta)/\vartheta \in \Theta\}$.

Por continuidad de $F(x, \vartheta)$ las funciones $D_n^+(\vartheta)$, $D_n^-(\vartheta)$ y $D_n(\vartheta)$ son continuas, aunque no diferenciables. Debido a la no diferenciabilidad de D_n, los estimadores ED_n se obtendrán mediante algoritmos de Programación No Lineal que no empleen derivadas. Obviamente, este hecho sólo afecta al cálculo numérico de $\hat{\vartheta}$, no a su fundamento.

Un caso particular interesante, que se analiza en este trabajo, es aquél en que Θ es un intervalo de $\mathbb{R}$ y la función $F(x, \vartheta)$ tiene propiedades de monotonía.

Teorema 1

Sean f, g: $[a, b] \to \mathbb{R}$, f estrictamente creciente, g estrictamente decreciente en $[a, b]$; sea $h = \text{máx} \{f, g\}$. La función h es fuertemente cuasiconvexa en $[a, b]$.

DEMOSTRACIÓN

Sean x_1, x_2 tales que $a \leqslant x_1 < x_2 \leqslant b$, $\lambda \in (0,1)$ y $x = \lambda x_1 + (1-\lambda)x_2$. Sea $h(x) = \text{máx}\{f(x), g(x)\}$. Si $h(x) = f(x)$ entonces $h(x) = f(x) < f(x_2) \leqslant \leqslant \text{máx}\{f(x_2), h(x_2)\} = h(x_2) \leqslant \text{máx}\{h(x_1), h(x_2)\}$. Si $h(x) = g(x)$ entonces $h(x) = g(x) < g(x_1) \leqslant \text{máx}\{h(x_1), h(x_2)\}$. ■

Teorema 2

Sea $F(x, \hat{\theta})$ continua y estrictamente creciente (o decreciente) en $\vartheta \in [\vartheta_L, \vartheta_U]$, para cada $x \in [x_{(1)}, x_{(n)}]$ fijo. Entonces $D_n(\vartheta)$ es fuertemente cuaxiconvexa en $[\vartheta_L, \vartheta_U]$.

DEMOSTRACIÓN

Supongamos $F(x, \vartheta)$ creciente en ϑ. Entonces $D_n^+(\vartheta)$ es continua y estrictamente decreciente en $[\vartheta_L, \vartheta_U]$. En efecto, sean $\vartheta_1 < \vartheta_2$ y $D_n^+(\vartheta_1) = F_n^*(y) - F(y, \vartheta_1)$, $D_n^+(\vartheta_2) = F_n^*(z) - F(z, \vartheta_2)$. Se verifica $D_n^+(\vartheta_1) \geqslant F_n^*(z) - F(z, \theta_1) > F_n^*(z) - F(z, \vartheta_2) = D_n^+(\vartheta_2)$. Análogamente se prueba que $D_n^-(\vartheta)$ es continua y estrictamente creciente en $[\vartheta_L, \vartheta_U]$ y como consecuencia del teorema 1 se deduce que $D_n(\vartheta)$ es fuertemente cuasiconvexa en $[\vartheta_L, \vartheta_U]$. ■

Por las propiedades de las funciones fuertemente cuasiconvexas —que pueden verse en numerosos textos de Programación No Lineal, por ejemplo, en Bazaraa, Shetty (1979)— se concluye que la función $D_n(\vartheta)$ tiene un único mínimo global en $[\vartheta_L, \vartheta_U]$, esto es, existe el estimador de mínimo D_n y es único.

Los resultados de los teoremas 1 y 2 son válidos en cualquier tipo de intervalo, acotado o no, de la recta real; sin embargo, no pueden generalizarse a espacios de mayor dimensión.

Cuando no se cumplen estas propiedades de monotonía las dificultades computacionales aumentan y puede suceder que haya varios estimadores ED_n mínimos locales, lo cual es frecuente en problemas de optimización.

3. EJEMPLOS DE DISTRIBUCIONES UNIPARAMETRICAS MONOTONAS

Ciertas distribuciones presentan monotonía en alguno de sus parámetros, lo que garantiza la unicidad del estimador ED_n de tal parámetro cuando los demás son conocidos. Además hay algoritmos eficientes para el cálculo de ED_n. Algunas distribuciones con esta propiedad son la Normal $N(\mu, \sigma)$ en μ, la Exponencial $E(\lambda)$, la Cauchy $C(\mu, \vartheta)$ en ϑ, la Weibull $W(a, b, c)$ en a y en b, entre otras.

Otra distribución con monotonía es la que llamamos semidiscreta por la derecha. Utilizaremos esta distribución junto con la Weibull en el estudio computacional, por lo que a continuación describimos ambas brevemente.

3.1. Distribución Weibull

La función de distribución de la variable aleatoria Weibull de parámetro de localización $a \in \mathbb{R}$, de escala $b > 0$ y de forma $c > 0 (X \equiv \equiv W(a, b, c))$ es:

$$F(x, \vartheta) = 1 - \exp\left[-\left(\frac{x-a}{b}\right)^c\right] \quad \text{si } x \geqslant a \ (0, \text{ si } x < a)$$

donde $\vartheta = (a, b, c)$ es el vector de parámetros.

Para $b_0 > 0$, $c_0 > 0$ y $x \in (x_{(1)}, x_{(n)})$ fijos, la función $F(x, a, b_0, c_0)$ es estrictamente decreciente en a en el intervalo $(-\infty, x_{(1)})$.

Para $a_0 \in \mathbb{R}$, $c_0 > 0$ y $x \in (x_{(1)}, x_{(n)})$ fijos, la función $F(x, a_0, b, c_0)$ es estrictamente decreciente en b en el intervalo $(0, \infty)$.

Esta propiedad de monotonía no se cumple para el parámetro c.

En virtud de los resultados de la sección 2 el estimador ED_n para a con b_0, c_0 fijos y para b con a_0, c_0 fijos existe y es único.

3.2. Distribución semidiscreta por la derecha

Vamos a construir una familia de funciones de distribución dependientes de un parámetro ϑ con las propiedades

$$\lim_{x \to n^+} f(x, \vartheta) = \infty, \quad \lim_{x \to n^-} f(x, \vartheta) = 0, \quad \forall\, n \in \mathbb{N} \cup \{0\}$$

esto es, la distribución de X está mayoritariamente concentrada en intervalos $[n, n + \varepsilon_n]$, con $\varepsilon_n \simeq 0$, de ahí el nombre que le asignamos. La función básica que nos va a permitir construir mediante traslaciones y homotecias la familia $F(x, \vartheta)$ es

$$g(x) = [1 - (1 - x)^{1+\theta}]^{\frac{1}{1+\theta}}$$

para la que

$$g'(x) = [1 - (1 - x)^{1+\theta}]^{\frac{-\theta}{1+\theta}}(1 - x)^{\theta}; \quad g'_{+}(0) = \infty, \quad g'_{-}(1) = 0$$

y la función utilizada para las homotecias y las traslaciones es

$$b(n, \vartheta) = 1 - \frac{1}{2^n} + \frac{1}{2^n}\frac{\vartheta}{\vartheta + 1}\frac{n}{n + 1}$$

que verifica

$$b(n, \vartheta) \in \left[1 - \frac{1}{2^n}, 1 - \frac{1}{2^n}\frac{1}{n + 1}\right]$$

$b(n, \vartheta)$ es creciente en ϑ, $\forall n > 0$ fijo y creciente en n, $\forall \vartheta > 0$ fijo; $b(0, \vartheta) = 0$, $\forall \vartheta \geqslant 0$.

Definición 2

Llamamos v.a. semidiscreta por la derecha de parámetro ϑ ($X \equiv SDD(\vartheta)$) a la v.a. X con función de distribución

$$F(x, \vartheta) = b(n, \vartheta) + (b(n + 1, \vartheta) - b(n, \vartheta))[1 - (n + 1 - x)^{1+\vartheta}]^{\frac{1}{1+\vartheta}}$$

para $n \leqslant x < n + 1$, $x \geqslant 0$, siendo $n = [x]$ (parte entera de x).

Es inmediato comprobar que $\forall \vartheta \geqslant 0$ $F(x, \vartheta)$ es función de distribución con derivada lateral por la izquierda nula y por la derecha infinita en los puntos enteros, $\forall \vartheta > 0$.

Las gráficas de esta familia están comprendidas, por monotonía, entre las correspondientes a $\vartheta = 0$ y a $\vartheta = \infty$.

Se verifica

$$F(x,0) = 1 - \frac{1}{2^n} + \frac{1}{2^{n+1}}(x - n), \quad \text{si } n \leqslant x < n + 1, \quad n = [x]$$

y es una función de distribución lineal a trozos.

$$F(x, \infty) = \lim_{\theta \to \infty} F(x, \vartheta) = \begin{cases} 1 - \dfrac{1}{(n+1)2^n} \;, & \text{si } x = n \text{ es entero} \\ 1 - \dfrac{1}{(n+2)2^{n+1}}, & \text{si } n < x < n+1 \;,\; n = [x] \end{cases}$$

$F(x, \infty)$ no es función de distribución ya que no es continua por la derecha (sí lo es por la izquierda). Sin embargo, podría considerarse como «límite natural» cuando $\vartheta \to \infty$ a la función de distribución

$$\bar{F}(x, \vartheta) = 1 - \frac{1}{(n + 2)2^{n+1}} \text{ si } n \leqslant x < n + 1 \quad , \quad n = [x]$$

Se puede demostrar que

$$E_\theta(X) = 2 - \frac{21}{1 + \vartheta}(1 - \lg 2) - \frac{1}{2(1 + \vartheta)} \frac{\Gamma^2\left(\dfrac{1}{1 + \vartheta}\right)}{\Gamma\left(\dfrac{2}{1 + \vartheta}\right)}$$

La familia $F(x, \vartheta)$ puede generalizarse sustituyendo los puntos enteros por una sucesión creciente $\{a_n\}$, $n \in \mathbb{N}$ y cambiando los $b(n, \vartheta)$. Se puede definir análogamente la v.a. semidiscreta por la izquierda a partir de la función

$$h(x) = 1 - [1 - x^{1+\theta}]^{\frac{1}{1+\theta}}$$

Dadas las peculiares características de la familia $F(x, \vartheta)$ cabe esperar dificultades computacionales en su tratamiento, mayores cuanto mayor sea ϑ.

4. RESULTADOS COMPUTACIONALES

Para comprobar la calidad de los estimadores de mínimo D_n realizamos un experimento de simulación con las v.a. Weibull y SDD, y comparamos esta metodología con la clásica de máxima verosimilitud (EMV).

4.1. Distribución Weibull

La simulación de valores aleatorios de la v.a. Weibull $W(a, b, c)$ se realiza mediante la fórmula $x = a + b(-lg\, r)^{1/c}$, donde r es un número aleatorio uniforme en $[0, 1]$.

Realizamos estimaciones uniparamétricas en cada parámetro a, b, c en dos modalidades, M_1: resto de parámetros verdaderos, M_2: resto de parámetros estimados por el método de Dannenbring (1977)

$$\tilde{a} = 2x_{(1)} - x_{(2)} \quad , \quad \tilde{b} = x_{([0.63n]+1)} - \tilde{a} \quad , \quad \tilde{c} = \frac{\lg \lg 2}{\lg \dfrac{x_M - \tilde{a}}{\tilde{b}}}$$

donde x_M es la mediana muestral.

Consideramos 10 valores de c, $c_i = 0.5i$, para $i = 1, ..., 10$ y $a = 0$, $b = 1$ como verdaderos parámetros de la v.a. Weibull. Generamos 50 muestras para cada tamaño muestral $n_1 = 10$ y $n_2 = 50$, cada modalidad M_1, M_2, cada parámetro a, b, c y cada valor c_i. Para comprobar la robustez de los estimadores realizamos dos tipos de contaminación en las muestras:

inferior: $x_{(i)} = x_{(i)} - \dfrac{1}{n} x_{(n)}$, para $i = 1, 2$ (si $n = 10$) o para
$i = 1, ..., 5$ (si $n = 50$)

superior: $x_{(i)} = 2x_{(i)}$, para $i = 8, ..., 10$ (si $n = 10$ o para
$i = 45, ..., 50$ (si $n = 50$)

El algoritmo de minimización univariable utilizado es el método de interpolación cuadrática descrito en Wolfe (1978), con algunas modificaciones en el criterio de convergencia.

En cada situación obtuvimos los estimadores ED_n y EMV. Para cada estimador evaluamos D_n y el logaritmo de la función de verosimilitud y calculamos la media y desviación típica de ambos valores y del estimador en las 50 muestras (cuando alguna muestra no permitió obtener los estimadores por desbordamiento numérico fue eliminada al calcular los valores medios). No fue posible obtener los estimadores para b y c en las muestras contaminadas inferiormente.

Ante la imposibilidad de mostrar todos los resultados obtenidos únicamente presentamos 12 tablas con los valores medios de los estimadores. En estas tablas la primera columna contiene el verdadero valor de c y las columnas ED_n y EMV los valores medios de estos estimadores. Se indica con un asterisco «*» el mejor de ellos.

Las tablas 1 a 12 son bastante explicativas. Cabe destacar, grosso modo, las siguientes observaciones generales:

- Muestras no contaminadas. En el tamaño muestral 10 los estimadores ED_n son mejores que los EMV para los parámetros a y c, mientras que son algo peores para el parámetro b. En muestras de tamaño 50 los estimadores ED_n sólo superan a los EMV para el parámetro c en la modalidad M_2.
- Muestras contaminadas inferiormente (sólo para el parámetro a). Los ED_n son claramente mejores en la modalidad M_1 y similares en la M_2.
- Muestras contaminadas superiormente. Los ED_n son claramente mejores en los dos tamaños muestrales 10 y 50, en los tres parámetros a, b, c en las dos modalidades M_1, M_2 y en casi todos los valores c_i.

TABLA 1

Parámetro a. Modalidad M_1. Tamaño muestral 10.

c_i	NO CONTAMINADAS		CONTAM. INFERIOR		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV	EDN	EMV
0,5	-0,0206	-0,0125*	*-1,1506	-1,1682	-0,0160	-0,0125*
1,0	*-0,0245	0,0502	*-0,2446	-0,2528	*-0,0025	0,0502
1,5	* 0,0254	0,0735	*-0,1043	-0,1189	* 0,0600	0,0919
2,0	* 0,0138	0,0508	*-0,0313	-0,0481	* 0,0642	0,1341
2,5	* 0,0188	0,0296	*-0,0041	-0,0404	* 0,0478	0,1849
3,0	* 0,0082	0,0145	*-0,0099	-0,0342	* 0,0353	0,2456
3,5	* 0,0213	0,0287	0,0058	-0,0056*	* 0,0333	0,3421
4,0	-0,0077	-0,0058*	*-0,0178	-0,0331	* 0,0028	0,3608
4,5	0,0058	0,0054*	*-0,0021	-0,0195	* 0,0146	0,3943
5,0	*-0,0039	-0,0044	*-0,0133	-0,0229	* 0,0014	0,4584

TABLA 2

Parámetro a. Modalidad M_2. Tamaño muestral 10.

c_i	NO CONTAMINADAS		CONTAM. INFERIOR		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV	EDN	EMV
0,5	-0,0641	-0,0170*	-0,9820	-0,9196*	-0,0515	-0,0137*
1,0	-0,0371	0,0293*	-0,3436	-0,2825*	* 0,0006	0,0430
1,5	* 0,0399	0,1100	-0,1773	-0,1219*	* 0,1024	0,1448
2,0	* 0,1351	0,1686	-0,0536	-0,0392*	* 0,1872	0,2249
2,5	* 0,1329	0,1916	-0,0279	0,0128*	* 0,1851	0,2620
3,0	* 0,2638	0,3121	* 0,1173	0,1525	* 0,3026	0,3774
3,5	* 0,3257	0,3448	* 0,1865	0,1909	* 0,3547	0,4169
4,0	* 0,3602	0,4113	* 0,2252	0,2617	* 0,4037	0,4604
4,5	* 0,3408	0,3683	* 0,2090	0,2161	* 0,3817	0,4566
5,0	* 0,4280	0,4586	* 0,2992	0,3145	* 0,4586	0,5243

TABLA 3

Parámetro b. Modalidad M_1. Tamaño muestral 10.

c_i	NO CONTAMINADAS		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV
0,5	1,2358	1,1514*	* 1,3316	1,6857
1,0	1,1016	1,0858*	* 1,1771	1,6243
1,5	* 1,0183	0,9531	* 1,1044	1,4263
2,0	1,0185	0,9932*	* 1,1213	1,5482
2,5	0,9570	0,9628*	* 1,0038	1,5529
3,0	1,0148	1,0132*	* 1,0557	1,6592
3,5	* 0,9964	0,9929	* 1,0182	1,6628
4,0	* 0,9988	0,9890	* 1,0195	1,6780
4,5	0,9779	0,9784*	* 0,9946	1,6843
5,0	0,9754	0,9763*	* 0,9886	1,7058

TABLA 4

Parámetro b. Modalidad M_2. Tamaño muestral 10.

c_i	NO CONTAMINADAS EDN	NO CONTAMINADAS EMV	CONTAM. SUPERIOR EDN	CONTAM. SUPERIOR EMV
0,5	* 1,2432	1,3283	* 1,4731	2,1085
1,0	1,1916	1,1377*	* 1,4050	1,7944
1,5	* 1,0072	1,0780	* 1,1287	1,7992
2,0	0,8264	0,8764*	* 0,8926	1,5329
2,5	* 0,7859	0,7648	* 0,8754	1,2918
3,0	0,6946	0,7102*	* 0,7518	1,2918
3,5	* 0,6858	0,6736	0,7544	1,1980*
4,0	0,6441	0,6579*	0,6813	1,2615*
4,5	0,5620	0,5705*	0,6017	1,1322*
5,0	* 0,5180	0,5090	0,5701	1,0351*

TABLA 5

Parámetro c. Modalidad M_1. Tamaño muestral 10.

c_i	NO CONTAMINADAS EDN	NO CONTAMINADAS EMV	CONTAM. SUPERIOR EDN	CONTAM. SUPERIOR EMV
0,5	* 0,4988	0,5468	0,4662	0,4739*
1,0	* 1,0853	1,1694	* 0,9514	0,8401
1,5	* 1,5859	1,7496	* 1,3293	1,1427
2,0	2,3460	2,2210*	* 2,1434	1,2215
2,5	* 2,5449	3,0332	* 2,2411	1,6623
3,0	* 2,8574	3,4299	* 2,5468	1,6785
3,5	* 3,4261	4,4676	* 3,0644	2,1576
4,0	* 4,0660	4,3154	* 3,7399	1,6777
4,5	6,0277	4,9913*	* 5,3681	1,9216
5,0	* 4,6636	5,6934	* 3,9374	2,3145

TABLA 6

Parámetro c. Modalidad M_2. Tamaño muestral 10.

c_i	NO CONTAMINADAS		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV
0,5	0,6653	0,6075*	0,6020	0,5015*
1,0	1,3963	1,1554*	1,2218	0,8276*
1,5	1,7198	1,5853*	* 1,4139	1,0631
2,0	3,0087	1,9055*	* 2,7004	1,1121
2,5	* 2,3053	2,1172	* 1,6114	1,1709
3,0	* 2,5125	2,3609	* 1,9288	1,2850
3,5	* 2,7087	2,3942	* 2,1267	1,1358
4,0	* 2,8353	2,5226	* 2,5838	1,1869
4,5	* 3,4968	3,4777	* 2,8890	2,3194
5,0	* 3,5736	2,5981	* 3,7549	1,2512

TABLA 7

Parámetro a. Modalidad M_1. Tamaño muestral 50.

c_i	NO CONTAMINADAS		CONTAM. INFERIOR		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV	EDN	EMV
0,5	-0,0074	-0,0012*	*-0,5016	-0,5025	-0,0070	-0,0012*
1,0	-0,0182	-0,0072*	*-0,0860	-0,0984	-0,0100	-0,0072*
1,5	*-0,0056	0,0086	*-0,0264	-0,0401	0,0200	0,0151*
2,0	-0,0090	0,0043*	*-0,0165	-0,0217	* 0,0259	0,0445
2,5	*-0,0011	0,0020	*-0,0050	-0,0121	* 0,0180	0,0985
3,0	-0,0079	-0,0072*	*-0,0089	-0,0160	* 0,0050	0,1448
3,5	0,0058	0,0044*	0,0050	-0,0014*	* 0,0147	0,2198
4,0	0,0006	-0,0002*	*-0,0002	-0,0050	* 0,0099	0,2466
4,5	*-0,0010	-0,0021	*-0,0014	-0,0050	* 0,0031	0,3100
5,0	* 0,0002	0,0006	*-0,0003	-0,0018	* 0,0079	0,3520

TABLA 8

Parámetro a. Modalidad M_2. Tamaño muestral 50.

c_i	NO CONTAMINADAS EDN	EMV	CONTAM. INFERIOR EDN	EMV	CONTAM. SUPERIOR EDN	EMV
0,5	-0,0177	-0,0032*	*-0,4751	-0,4895	-0,0138	-0,0028*
1,0	-0,0217	-0,0174*	*-0,1015	-0,1094	-0,0126	-0,0124*
1,5	* 0,0011	0,0097	*-0,0485	-0,0495	* 0,0200	0,0245
2,0	* 0,0331	0,0378	*-0,0065	-0,0078	* 0,0545	0,0629
2,5	0,0886	0,0871*	0,0540	0,0478*	* 0,1124	0,1285
3,0	0,1474	0,1441*	0,1162	0,1087*	* 0,1561	0,1860
3,5	0,2005	0,1950*	0,1707	0,1617*	* 0,2159	0,2453
4,0	0,2215	0,2181*	0,1931	0,1867*	* 0,2309	0,2883
4,5	0,3248	0,3221*	0,2977	0,2912*	* 0,3346	0,3724
5,0	0,3297	0,3254*	0,3033	0,2963*	* 0,3345	0,3850

TABLA 9

Parámetro b. Modalidad M_1. Tamaño muestral 50.

c_i	NO CONTAMINADAS EDN	EMV	CONTAM. SUPERIOR EDN	EMV
0,5	1,0767	1,0725*	* 1,1671	1,4719
1,0	1,0439	1,0274*	* 1,1698	1,4481
1,5	1,0045	1,0031*	* 1,1406	1,4584
2,0	0,9918	0,9926*	* 1,0997	1,4828
2,5	* 0,9982	0,9972	* 1,0553	1,5252
3,0	0,9905	0,9934*	* 1,0199	1,5600
3,5	1,0078	1,0017*	* 1,0339	1,5989
4,0	* 1,0003	1,0015	* 1,0235	1,6343
4,5	* 0,9987	0,9955	* 1,0153	1,6495
5,0	1,0002	1,0001*	* 1,0104	1,6897

TABLA 10

Parámetro b. Modalidad M_2. Tamaño muestral 50.

c_i	NO CONTAMINADAS		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV
0,5	* 1,0448	1,3986	* 1,1544	2,1415
1,0	* 0,9706	1,0753	* 1,0626	1,6232
1,5	0,9634	1,0010*	* 1,0497	1,5117
2,0	0,9387	0,9748*	* 1,0075	1,5187
2,5	0,8844	0,9132*	* 0,9306	1,4849
3,0	0,8386	0,8481*	* 0,8767	1,3963
3,5	0,7575	0,7831*	* 0,7860	1,3621
4,0	0,7652	0,7806*	* 0,7856	1,3803
4,5	* 0,6605	0,6584	0,6816	1,1795*
5,0	0,6450	0,6553*	0,6560	1,2394*

TABLA 11

Parámetro c. Modalidad M_1. Tamaño muestral 50.

c_i	NO CONTAMINADAS		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV
0,5	* 0,4955	0,5047	* 0,4658	0,4382
1,0	0,9916	1,0012*	* 0,8999	0,7579
1,5	1,4812	1,4994*	* 1,2575	1,0011
2,0	1,9698	2,0169*	* 1,5717	1,2012
2,5	* 2,4861	2,5404	* 2,0341	1,3799
3,0	2,8728	3,0034*	* 2,2785	1,4614
3,5	3,6132	3,5834*	* 3,0248	1,5622
4,0	4,0298	4,0097*	* 3,4527	1,6847
4,5	4,7279	4,6491*	* 4,1285	1,7134
5,0	4,8620	5,0520*	* 4,6068	1,7653

TABLA 12

Parámetro c. Modalidad M_2. Tamaño muestral 50.

c_i	NO CONTAMINADAS		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV
0,5	0,5288	0,5158*	* 0,4948	0,4474
1,0	1,0565	1,0137*	* 0,9079	0,7634
1,5	* 1,4788	1,4221	* 1,1956	0,9677
2,0	* 1,9106	1,8227	* 1,4549	1,1347
2,5	* 2,3154	2,1947	* 1,6824	1,2349
3,0	* 2,6044	2,4735	* 2,0297	1,2876
3,5	* 2,6968	2,5150	* 2,0102	1,2648
4,0	* 3,1589	2,9781	* 2,6507	1,3378
4,5	* 3,0144	2,8649	* 2,2383	1,3015
5,0	* 3,4186	3,1123	* 2,6619	1,3456

4.2. Distribución semidiscreta por la derecha

La simulación de valores aleatorios de la v.a. SDD se realiza mediante la fórmula

$$x = n + 1 - \left[1 - \left(\frac{1 - r - b(n, \vartheta)}{b(n + 1, \vartheta) - b(n, \vartheta)} \right)^{1 + 6} \right]^{\frac{1}{1 + \vartheta}}$$

donde r es uniforme en $[0, 1]$ y n es tal que $b(n, \vartheta) \leqslant 1 - r \leqslant$ $\leqslant b(n + 1, \vartheta)$.

Consideramos dos tamaños muestrales $n_1 = 10$ y $n_2 = 50$. Con n_1 utilizamos los verdaderos valores de $\vartheta \in \{0,5; 0,75; 1; 1,5; 3; 4,5; 6; 7,5; 9; 10,5; 12; 13,5; 15; 20$ *y* $30\}$. Con n_2 utilizamos además los valores 0,25; 50 y 100 (no fue posible hacerlo con n_1 por problemas de desbordamiento numérico en el cálculo de los estimadores *EMV*). Realizamos el mismo tipo de contaminación que para la distribución Weibull. No se pudo obtener el estimador en las muestras contaminadas inferiormente.

Las tablas 13 y 14 recogen los valores medios obtenidos. La primera columna contiene el verdadero valor de ϑ.

TABLA 13

Distribución SDD. Tamaño muestral 10.

ϑ	NO CONTAMINADAS		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV
0,5	* 0,5741	0,5875	------	------
0,75	0,8279	0,7858*	* 0,7045	0,6954
1,0	1,0915	1,0506*	* 0,9514	0,9386
1,5	* 1,6102	1,6507	1,4346	1,5050*
3,0	3,2153	3,1768*	* 2,9478	2,8321
4,5	4,8422	4,5116*	* 4,5093	4,1859
6,0	6,4486	6,0804*	* 6,0239	5,6187
7,5	8,0581	7,8107*	* 7,5622	7,3466
9,0	9,6740	9,3310*	* 9,0802	8,6697
10,5	* 11,2851	11,4954	* 10,5780	10,7943
12,0	* 12,8963	13,5870	* 12,0819	12,8865
13,5	* 14,5046	15,2976	* 13,5882	14,4632
15,0	* 16,1128	27,1955	* 15,0887	27,8949
20,0	* 21,6653	46,9035	* 20,4931	45,1212
30,0	* 38,7146	114,3279	* 37,4585	110,1433

TABLA 14

Distribución SDD. Tamaño muestral 50.

ϑ	NO CONTAMINADAS		CONTAM. SUPERIOR	
	EDN	EMV	EDN	EMV
0,25	0,2633	0,2573*	0,1698	0,2389*
0,5	0,5051	0,5024*	0,3631	0,4640*
0,75	* 0,7545	0,7559	0,5898	0,6873*
1,0	* 1,0057	0,9786	0,8204	0,9041*
1,5	* 1,5080	1,5163	1,2803	1,3881*
3,0	3,0249	3,0147*	2,6664	2,7519*
4,5	* 4,5442	4,4057	4,0698	4,0733*
6,0	* 6,0532	5,9407	5,4525	5,5774*
7,5	* 7,5603	7,3982	6,8362	6,9976*
9,0	* 9,0655	9,1888	8,2089	8,7797*
10,5	* 10,5814	11,6850	9,5997	11,2081*
12,0	* 12,0931	15,2078	* 10,9845	14,6385
13,5	* 13,6140	18,1463	* 12,3704	17,4957
15,0	* 15,1339	23,7145	* 13,7627	23,6175
20,0	* 20,1949	44,8058	* 18,4362	43,9126
30,0	* 31,1422	104,9788	* 29,8669	102,0000
50,0	* 97,5433	267,4467	* 97,5433	259,3212
100,0	*429,2000	791,9275	*429,2000	782,5309

De las tablas 13 y 14 se derivan las siguientes consecuencias:

— Muestras no contaminadas. En tamaño muestral 10 se aprecia cierta igualdad entre ambos estimadores para valores del parámetro menores que 10,5 y una mayor robustez de los ED_n para valores mayores. En tamaño 50 se aprecia un mejor comportamiento general de los ED_n y, también, una mayor robustez para valores grandes del parámetro.

— Muestras contaminadas superiormente. En tamaño 10 son claramente mejores los ED_n, mientras que en tamaño 50 esta mejoría sólo ocurre para valores del parámetro mayores que 12.

5. CONCLUSIONES

En este trabajo se ha estudiado un procedimiento de estimación paramétrica consistente en obtener los estimadores que minimizan el estadístico D_n de Kolmogorov-Smirnov. La justificación teórica de este método es del mismo tipo que la del de máxima verosimilitud: en ambos casos se pretende obtener los estimadores que más se acercan a una propiedad razonable del ajuste.

La calidad de los estimadores ED_n es comparable a la de los *EMV* en los casos estudiados; incluso, a veces, es superior y da resultados aceptables en situaciones en las que los *EMV* no son fiables. La obtención computacional de los ED_n es algo más costosa que la de los *EMV*, ya que la evaluación de D_n requiere, en general, más operaciones que la de la función de verosimilitud y el número de evaluaciones de D_n en el algoritmo de optimización es algo mayor que el de la función de verosimilitud.

Los estimadores de mínimo D_n no deben considerarse como sustitutos de los de máxima verosimilitud, sino como una alternativa válida con fundamentos similares.

Queda abierto el estudio de propiedades generales de sesgo, varianza, asintóticas, etc., de estos estimadores y su comportamiento en distribuciones concretas de interés.

REFERENCIAS

BAZARAA, M. S., and SHETTY, C. M.: (1979) *Nonlinear Programming. Theory and Algorithms*, Wiley.

DANNENBRING, D. G. (1977): «Procedures for Estimating Optimal Solution Values for Large Combinatorial Problems», *Man. Sci.*, 23, 1273-1283.

PARR, W. C. (1981): «Minimum Distance Estimation. A Bibliography», *Commun. Statist. Theor. Meth.*, A10 (12), 1205-1224.

ROHATGI, V. K. (1976): *An Introduction to Probability Theory and Mathematical Statistic*», Wiley.

WOLFE, M. A. (1978): *Numerical Methods for Unconstrained Optimization*, Van Nostrad Reinhold Company.

WOLFOWITZ, J. (1957): «The Minimum Distance Method», *Annals of Mathematical Statistics*, 28, 75-88.