Revista de Matemática: Teoría y Aplicaciones
ISSN: 1409-2433
mta.cimpa@ucr.ac.cr
Universidad de Costa Rica
Costa Rica

Martínez-Camblor, Pablo; Carleos, Carlos; Corral, Norberto
Sobre el estadístico de Cramér-von Mises
Revista de Matemática: Teoría y Aplicaciones, vol. 19, núm. 1, 2012, pp. 89-101
Universidad de Costa Rica
San José, Costa Rica

Disponible en: http://www.redalyc.org/articulo.oa?id=45326925007

Revista de Matemática: Teoría y Aplicaciones 2012 **19**(1) : 89–101
cimpa – ucr issn: 1409-2433

SOBRE EL ESTADÍSTICO DE CRAMÉR–VON MISES

ON THE CRAMÉR–VON MISES STATISTIC

Pablo Martínez-Camblor* Carlos Carleos†
Norberto Corral‡

Received: 3-May-2010; Revised: 25-May-2011; Accepted: 2-Nov-2011

Resumen

Uno de los criterios más utilizados para comparar funciones es el introducido por los investigadores Harald Cramér y Richard Edler von Mises y conocido como criterio de Cramér–von Mises ($\mathcal{C}_M$) siendo aplicado a problemas que van desde la bondad de ajuste de una distribución hasta la comparación de la igualdad entre cópulas. En este trabajo, se aplican procesos empíricos para la obtención de la distribución asintótica de la generalización del estadístico $\mathcal{C}_M$ al problema de comparación de k-muestras independientes propuesta por Kiefer. Se estudia la calidad de esta aproximación y se indica como, dado un problema concreto, aproximar la significación final.

*Oficina de Investigación Biosanitaria del Principado de Asturias, C/Rosal 7 bis, 33009 Oviedo, España. E-mail: pmcamblor@hotmail.com

†Departamento de Estadística e Investigación Operativa y Didáctica de la Matemática, Facultad de Ciencias, Universidad de Oviedo – Campus de Llamaquique c/ Calvo Sotelo s/n 33007 Oviedo, España. E-mail: carleos@uniovi.es

‡Misma dirección que/*Same address as* C. Carleos. E-mail: norbert@uniovi.es

Palabras clave: Criterio de Cramér–von Mises, procesos empíricos, comparación de k-muestras

Abstract

Probably, one of the most useful criterions in order to compare distribution functions is the one introduced by the researchers Harald Cramér and Richard Edler von Mises which is known as Cramér-von Mises criterion ($\mathcal{C}_M$). It has been applied on a vast variety of problems. In this work, the theory of empirical processes is applied in order to obtain the asymptotic distribution for the generalization to the k-sample problem of $\mathcal{C}_M$ proposed by Kiefer. The quality of this approximation is also studied and some indications about how to obtain an approximation to the final P-value are also included.

Keywords: Cramér–von Mises criterion, empiric process, k-sample problem.

Mathematics Subject Classification: 60E05, 62G10.

1 Introducción

Desde que a finales de los años veinte (Cramér; 1928) el matemático sueco Harald Cramér (nacido en Estocolmo en 1893) y, de forma independiente, a principios de los años treinta (von Mises; 1931) el físico austrohúngaro Richard Edler von Mises (nacido en Lemberg en 1883) introdujeran el ahora conocido como criterio de Cramér-von Mises ($\mathcal{C}_M$) en el estudio de problemas de bondad de ajuste, que, dada una determinada distribución de probabilidad, F^*, y una función de distribución F^0, queda definido por

$$\mathcal{C}_M = \int (F^*(t) - F^0(t))^2 dF^0(t), \tag{1}$$

ha sido aplicado a una gran variedad de problemas siendo objeto de innumerables estudios y publicaciones científicas. Dada una muestra $X = \{x_1, \ldots, x_n\}$, sin más que sustituir en (1) $F^*(t)$ y $F^0(t)$ por la correspondiente función de distribución empírica (FDE), $F_n(X, t)$, y por una función de distribución teórica, respectivamente, se obtiene el test de Cramér-von Mises para bondad de ajuste. Csörgő y Faraway (1996) calculan la distribución exacta para este estadístico. Anderson (1962) estudia la aplicación del criterio de Cramér-von Mises al problema de comparación de dos muestras independientes. Calcula su esperanza y varianza (exactas y asintóticas) y deriva tablas para la aproximación de la significación en su implementación práctica. También se ha aplicado este criterio en la

construcción de tests de bondad de ajuste cuando los datos tienen censuras aleatorias (Koziol et al.; 1976), para comparar la simetría de una distribución (Viollaz et al.; 1996) o, más recientemente, en la comparación de la igualdad entre cópulas (Rémillar y Scaillet; 2009) o de curvas ROC (Martínez-Camblor et al.; 2011) entre otras muchas aplicaciones. Por supuesto, también han surgido innumerables versiones; multidimensionales (Deheuvels; 2005), ponderadas (Öztürk et al.; 1987) o de tipo L_1 (Schmid et al.; 1996) entre otras muchas.

En este trabajo, se retoma la generalización del estadístico de Cramér-von Mises al problema de comparación de k-muestras independientes propuesta por Kiefer (1959) y definida por

$$\mathcal{C}_M(k) = \sum_{i=1}^{k} n_i \int (\hat{F}_{n_i}(t) - \bar{F}_n(t))^2 d\bar{F}_n(t) \tag{2}$$

donde dadas k-muestras independientes de tamaños $n_1, \ldots, n_k$, $\hat{F}_{n_i}$ representa la FDE asociada a la i-ésima muestra ($1 \leq i \leq k$) y $\bar{F}_n = k^{-1} \sum_{i=1}^{k} \hat{F}_{n_i}$. Se utiliza la expansión de Karhunen-Loève de los correspondientes funcionales cuadráticos del proceso *Gaussiano* resultante para obtener su distribución asintótica. Se proponen distintas aproximaciones (basadas en la distribución asintótica) para el cálculo de la significación (P-valor) y, mediante el método de Monte Carlo, se estudia la calidad de las mismas. La aproximación mediante el método de permutaciones aleatorias es también considerada.

2 Distribución asintótica

El estadístico $\mathcal{C}_M(k)$ tiene distribución libre y su distribución muestral exacta puede ser calculada enumerando de forma exhaustiva las $n!/(n_1!n_2!\cdots n_k!)$ combinaciones distintas ($n = \sum_{i=1}^{k} n_i$) o, más usualmente, ser aproximada por una selección aleatoria de esas posibilidades. Este procedimiento ha sido considerado en diversos estudios. Por ejemplo, en Martínez-Camblor (2008), el estadístico de Cramér-von Mises es incluído en un estudio de simulación en el que se consideran otros seis estadísticos basados en la FDE y, uno más, basado en la estimación núcleo para función de densidad en problemas de tres muestras. Sobre doce modelos distintos (seis modelos simétricos y otros seis asimétricos), el estadístico $\mathcal{C}_M(k)$ obtuvo resultados competitivos en casi todos los modelos, si bien, en aquellos en los que las diferencias están localizadas principalmente en el parámetro de localización, tests específicos para este tipo de situaciones, como el de Kruskall-Wallis, son considerablemente más potentes.

La convergencia de FDE a la función de distribución real es, probablemente, uno de los tópicos más ampliamente estudiados en el campo de la estadística matemática. Desde el Teorema 3 del trabajo de Komlòs et al. (1975) se deriva, bajo condiciones muy generales, que existe un espacio probabilístico, de modo que se da la igualdad (esta igualdad es conocida como Proceso Húngaro),

$$\sqrt{n}(\hat{F}_n(X,t) - F(t)) = B\{F(t)\} + \mathcal{O}(\log n/\sqrt{n}) \qquad \text{c.s.} \tag{3}$$

donde $B\{t\}$ $(0 \leq t \leq 1)$ es un *Puente Browniano* usual, esto es, un proceso estocástico *Gaussiano*, de media nula, varianza $\mathbb{E}[B\{t\}^2] = t(1-t)$ y covarianza $\mathbb{E}[B\{t\}B\{s\}] = t \wedge s - st$ $(t \wedge s = \min\{s,t\})$. Por tanto, si se asume que $\lim_{n_i \wedge n_j \to \infty} n_i/n_j = \alpha_{ij}^2 < \infty$ $(1 \leq i,j \leq k)$ desde el Teorema de la Convergencia Dominada y, sin más que realizar un cambio de variable se tiene la convergencia

$$\begin{aligned} \mathcal{C}_M(k) &= \sum_{i=1}^{k} n_i \int (\hat{F}_{n_i}(t) - \bar{F}_n(t))^2 d\bar{F}_n(t) \\ &\xrightarrow{\mathcal{L}} \sum_{i=1}^{k} \int [B_i\{F(t)\} - \bar{B}_i\{F(t)\}]^2 dF(t) \end{aligned} \tag{4}$$

donde para cada $i \in 1,\ldots,k$, $B_i\{F(t)\}$ son puentes brownianos independientes y $\bar{B}_i\{F(t)\} = k^{-1}\sum_{j=1}^{k} \alpha_{ij} B_j\{F(t)\}$. Para cada $1 \leq i \leq k$, el proceso estocástico $\mathcal{X}_i\{t\} = B_i\{F(t)\} - \bar{B}_i\{F(t)\}$ es gaussiano y centrado, además, dado que para $i \neq j$, B_i y B_j son procesos independientes y teniendo en cuenta que $i = j \Rightarrow \alpha_{ij} = 1$,

$$\begin{aligned} \mathbb{E}[\mathcal{X}_i\{t\}^2] =& \mathbb{E}[(B_i\{F(t)\} - \bar{B}_i\{F(t)\})^2] \\ =& \frac{1}{k^2}\mathbb{E}\left[\left((k-1)B_i\{F(t)\} + \sum_{j\neq i}^{k} \alpha_{ij} B_j\{F(t)\}\right)^2\right] \\ =& \frac{(k-1)^2}{k^2}\mathbb{E}[B_i\{F(t)\}^2] - \frac{1}{k^2}\sum_{j\neq i}^{k} \alpha_{ij}^2 \mathbb{E}[B_j\{F(t)\}^2] \\ =& \left(\frac{(k-1)^2}{k^2} + \frac{1}{k^2}\sum_{j\neq i}^{k} \alpha_{ij}^2\right) F(t)(1-F(t)). \end{aligned}$$

Análogamente,

$$\begin{aligned} \mathbb{E}[\mathcal{X}_i\{t\}\mathcal{X}_i\{s\}] =& \mathbb{E}[(B_i\{F(t)\} - \bar{B}_i\{F(t)\})(B_i\{F(s)\} - \bar{B}_i\{F(s)\})] \\ =& \left(\frac{(k-1)^2}{k^2} + \frac{1}{k^2}\sum_{j\neq i}^{k} \alpha_{ij}^2\right) (F(t) \wedge F(s) - F(s)F(t)). \end{aligned}$$

Por tanto, para cada $1 \leq i \leq k$ se tiene que si

$$C_i = \left(\frac{(k-1)^2}{k^2} + \frac{1}{k^2} \sum_{j \neq i}^{k} \alpha_{ij}^2 \right)^{-1/2},$$

el proceso $B_i^*\{F(t)\} = C_i \mathcal{X}_i\{t\}$ es un puente browniano standard. Estudiar la distribución asintótica de $\mathcal{C}_M(k)$ es estudiar la distribución de

$$\sum_{i=1}^{k} \int \mathcal{X}_i\{t\}^2 dF(t) = \sum_{i=1}^{k} \int (C_i^{-1} B_i^*\{F(t)\})^2 dF(t), \tag{5}$$

donde para cada $1 \leq i \leq k$, $B_i^*\{t\}$, $(0 \leq t \leq 1)$ es un puente browniano standard. Notar que estos procesos no son los definidos anteriormente y, en particular, no son independientes, esto es $\mathbb{E}[B_i^*\{t\}B_j^*\{t\}] \neq 0$ para $1 \leq i, j \leq k$. Concretamente, realizando cálculos similares a los anteriores se tiene que

$$\mathbb{E}[B_i^*\{F(t)\}B_j^*\{F(t)\}] = -\frac{2(k-1)}{k^2 C_i C_j} F(t)(1 - F(t)).$$

Las propiedades generales de los procesos estocásticos nos dicen que, dado un proceso gaussiano centrado, $\mathcal{N}\{t\}$ $(0 \leq t \leq 1)$, satisfaciendo que $\int \mathbb{C}(s,t)^2 ds dt < \infty$, donde $\mathbb{C}(s,t) = \mathbb{E}[\mathcal{N}\{t\}\mathcal{N}\{s\}]$ se verifica que

$$\mathcal{N}\{t\} = \sum_{j \in \mathbb{N}} \sqrt{\lambda_j} e_j(t) Y_j, \tag{6}$$

donde $\{Y_j\}_{j \in \mathbb{N}}$ es una familia de variables aleatorias independientes con distribución normal de media cero y varianza uno, $\{e_j(\cdot)\}_{j \in \mathbb{N}}$ es una base ortonormal del espacio de Hilbert $L^2([0,1])$ y $\{\lambda_j\}_{j \in \mathbb{N}}$ es una sucesión de números reales no negativos verificando $\lambda_1 \geq \lambda_2 \geq \ldots$. Además, desde la ortonormalidad de la base $\{e_j(\cdot)\}_{j \in \mathbb{N}}$ se deduce que

$$\sum_{j \in \mathbb{N}} \lambda_j = \iint \mathbb{C}(s,t) ds ds < \infty.$$

La representación dada en (6), usualmente conocida como *descomposición de Karhunen-Loève* (ver, por ejemplo, Adler; 1990), permite directamente obtener la igualdad

$$\int \mathcal{N}\{t\}^2 dt = \sum_{j \in \mathbb{N}} \lambda_j Y_j^2. \tag{7}$$

La determinación explicita de la distribución anterior, exige conocer los valores de $\{\lambda_j\}_{j \in \mathbb{N}}$, para ello, se debe calcular la función $\mathbb{C}(s,t)$ (también

conocida como kernel). Esta descomposición es, en general, un cálculo complicado que suele involucrar, entre otras cosas, la resolución de una ecuación diferencial, funciones de Bessel, etc... (ver, por ejemplo, Abramowitz et al.; 1965). En el caso del puente browniano standard, $B\{t\}$ $(0 \leq t \leq 1)$ se tiene que (ver, entre otros, Anderson et al.; 1952 o Anderson; 1962)

$$\int B\{t\}^2 dt = \sum_{j\in\mathbb{N}} \frac{1}{j^2\pi^2} Y_j^2 = \sum_{j\in\mathbb{N}} \frac{1}{j^2\pi^2}(Y_j^2 - 1) + \frac{1}{6} \quad , \tag{8}$$

donde $\{Y_j\}_{j\in\mathbb{N}}$ es una familia de variables aleatorias independientes con distribución normal de media cero y varianza uno (al mismo resultado se llega considerando el correspondiente U-estadístico degenerado, ver por ejemplo Van der Vaart; 1998). Por otro lado, en Tolmatz (2002) se proponen algunas aproximaciones para la función de distribución de la variable aleatoria dada en (8).

Volviendo a la distribución de la variable aleatoria dada en (5) (objeto de este estudio), se tiene, sin más que aplicar la ecuación (6), que para cada $i \in 1, \ldots, k$,

$$B_i^*\{F(t)\} = \sum_{j\in\mathbb{N}} \sqrt{\lambda_j} e_{i,j}(t) Y_{i,j}, \tag{9}$$

donde $\{Y_{i,j}\}_{j\in\mathbb{N}}$ es una familia de variables aleatorias independientes con distribución normal de media cero y varianza uno, $\{e_{i,j}(\cdot)\}_{j\in\mathbb{N}}$ es una base ortonormal (definida anteriormente) y $\lambda_j = (j\pi)^{-2}$ $(j \in \mathbb{N})$. Desde la ecuación (8),

$$\begin{aligned} \sum_{i=1}^{k} \int \mathcal{X}_i\{t\}^2 dF(t) &= \sum_{i=1}^{k} C_i^{-2} \int B_i^*\{F(t)\}^2 dF(t) \\ &= \sum_{i=1}^{k} C_i^{-2} \left(\sum_{j\in\mathbb{N}} \frac{1}{j^2\pi^2}(Y_{i,j}^2 - 1) + \frac{1}{6} \right) \end{aligned} \tag{10}$$

donde $\{\boldsymbol{Y_j} = (Y_{1,j}, \ldots, Y_{k,j})\}_{j\in\mathbb{N}}$ es una familia de variables aleatorias con distribución k-dimensional, cuyas distribuciones marginales (normales) tienen media cero y varianza uno. Para demostrar la normalidad de $\boldsymbol{Y_j}$ (para cada $j \in \mathbb{N}$), se comprobará que cada combinación lineal de sus componentes, $\sum_{i=1}^{k} a_i Y_{i,j}$, $(a_i \in \mathbb{R} \quad \forall i \in 1, \ldots, k)$ sigue una distribución normal. Se tiene que

$$\boldsymbol{\mathcal{X}^*}\{t\} = (a_1 \mathcal{X}_1\{t\}\, e_{1,j}(t), \cdots, a_k \mathcal{X}_k\{t\}\, e_{k,j}(t))$$

es un proceso k-dimensional *Gaussiano* centrado. Desde la ecuación (9) y para cada $i \in 1, \ldots, k$

$$a_i \mathcal{X}_i\{t\}\, e_{i,j}(t) = a_i C_i^{-1} B_i^*\{F(t)\}\, e_{i,j}(t) a_i\, e_{i,j}(t) \sum_{l\in\mathbb{N}} \sqrt{\lambda_i}\, e_{i,l}(t)\, Y_{i,l},$$

por tanto

$$a_i \sum_{l\in\mathbb{N}} \sqrt{\lambda_l}\, Y_{i,l} \int e_{i,j}(t) e_{i,l}(t) dF(t) = a_i \int \mathcal{X}_i\{t\}\, e_{i,j}(t) dF(t).$$

Luego,

$$\sum_{i=1}^{k} a_i\, Y_{i,j} = \sum_{i=1}^{k} \frac{a_i}{\sqrt{\lambda_j}} \int \mathcal{X}_i\{t\}\, e_{i,j}(t) dF(t) = \int \sum_{i=1}^{k} \frac{a_i}{\sqrt{\lambda_i}} \mathcal{X}_i\{t\}\, e_{i,j}(t) dF(t)$$

sigue una distribución normal. Por otro lado, si para cada $i \in 1, \ldots, k$ se define $\mathcal{S}_{(i)} = C_i^{-2} \int \mathcal{X}_i\{t\}^2 dF(t)$, la covarianza entre los k sumandos involucrados en la expresión (5) viene determinada por

$$\begin{aligned}
\mathbb{C}ov[C_i^2 \mathcal{S}_{(i)}, C_j^2 \mathcal{S}_{(j)}] =& \mathbb{C}ov\left[\int \mathcal{X}_i\{t\}^2 dF(t), \int \mathcal{X}_j\{s\}^2 dF(s)\right] \\
=& \iint \mathbb{C}ov\left[\mathcal{X}_i\{t\}^2, \mathcal{X}_j\{s\}^2\right] dF(t) dF(s) \qquad (11) \\
=& \iint 2\, \mathbb{E}[\mathcal{X}_i\{t\}\mathcal{X}_j\{s\}]^2 dF(t) dF(s). \qquad (12)
\end{aligned}$$

Y, por tanto,

$$\mathbb{E}[\mathcal{X}_i\{t\}\mathcal{X}_j\{s\}] = \frac{1}{k^2}\left(\sum_{l=1}^{k} \alpha_{il}\alpha_{jl} - k(\alpha_{ij} + \alpha_{ji})\right)(F(t) \wedge F(s) - F(s)F(t)). \qquad (13)$$

Sustituyendo en (11) y realizando los correspondientes cálculos para $1 \leq i \neq j \leq k$ se obtiene que

$$\mathbb{C}ov[C_i^2 \mathcal{S}_{(i)}, C_j^2 \mathcal{S}_{(j)}] = \frac{1}{45\, k^4}\left(\sum_{l=1}^{k} \alpha_{il}\alpha_{jl} - k(\alpha_{ij} + \alpha_{ji})\right)^2. \qquad (14)$$

3 Aproximaciones a la distribución

Los cálculos realizados en la sección anterior, demuestran que la distribución límite del estadístico $\mathcal{C}_M$ es una suma ponderada de las componentes de infinitas variables aleatorias normales k-dimensionales (no necesariamente

independientes). Obviamente, en la práctica, no se podrá hacer esta suma infinita y deberemos conformarnos con alguna aproximación. A continuación, se proponen dos métodos cuyo objetivo es la obtención de una probabilidad final.

En la aproximación más simple, se considera únicamente el primer sumando. Así las cosas, se tendría

$$\mathcal{C}_M \sim A(1) = \sum_{i=1}^{k} C_i^{-2} \left(\frac{1}{\pi^2}(Y_i^2 - 1) + \frac{1}{6} \right), \tag{15}$$

donde $\boldsymbol{Y} = (Y_1, \ldots, Y_k)$ es una variable aleatoria k-dimensional cuyas marginales tienen media nula y varianza uno y cuyas covarianzas, se derivarán desde (14) resultando que, para $1 \leq i \neq j \leq k$,

$$\mathbb{E}[Y_i Y_j] = \sigma_{i,j} = \pm \frac{|\sum_{l=1}^{k} \alpha_{il}\alpha_{jl} - k(\alpha_{ij} + \alpha_{ji})|\pi^2}{\sqrt{90}k^2 C_i C_j}.$$

Notar que, si la matriz de varianzas y covarianzas de $\boldsymbol{Y}$ se calcula desde (14), los k sumandos de $A(1)$ tendrán la misma relación que los k sumandos de (5). Así las cosas, se propone la siguiente aproximación,

$$\mathcal{C}_M \sim A(l) = \sum_{i=1}^{k} C_i^{-2} \left(\frac{1}{\pi^2}(Y_i^2 - 1) + \sum_{j=2}^{l} \frac{1}{j^2\pi^2}(Y_{i,j} - 1) + \frac{1}{6} \right) \tag{16}$$

donde $\boldsymbol{Y} = (Y_1, \ldots, Y_k)$ es la variable aleatoria definida en (15) y para $i \in 1, \ldots, k$, $j \in 1, \ldots, l$, $Y_{i,j}$ son variables aleatorias independientes con distribución normal de media cero y varianza uno.

3.1 Calidad de las aproximaciones

Una de las grandes ventajas del test de Cramér-von Mises para k muestras independientes radica en que su distribución no depende de la distribución de procedencia de las muestras (esta propiedad es compartida por muchos otros estadísticos basados en este criterio). Por este motivo, para ilustrar la calidad de las aproximaciones propuestas, no es necesario considear distintos modelos. Nosotros nos limitaremos a un pequeño estudio de simulación de Monte Carlo. El caso considerado es un problema de comparación de tres muestras de tamaños 50 y 100 procedentes de distribuciones normales standarizadas.

Dadas las características de los modelos considerados, se tiene que para $1 \leq i \leq 3$, $C_i = \sqrt{3/2}$. Además, para $1 \leq i \neq j \leq 3$ se tiene que $\sigma_{i,1} \approx \pm 0.2311$. Por tanto, las aproximaciones $A(l)$ $(l \geq 1)$ se pueden aproximar

Tabla 1: *Percentiles 99 (P_{99}), 95 (P_{95}) y 90 (P_{90}) para la distribución real y para las distintas aproximaciones: Permutaciones (AP), $A(m)$ con $m = 1$ $(A(1))$ $m = 5$ $(A(5))$ y $m = 10$ $(A(10))$, y los distintos tamaños muestrales considerados ($n = n_1 = n_2 = n_3$).*

n		Real	AP	$A(1)$	$A(5)$	$A(10)$
50	P_{99}	1.0699	1.0758	0.9322	0.0970	1.0126
	P_{95}	0.7427	0.7523	0.6810	0.7195	0.7370
	P_{90}	0.6027	0.6179	0.5627	0.6033	0.6127
100	P_{99}	1.0545	1.0461	0.9322	0.0970	1.0126
	P_{95}	0.7536	0.7496	0.6810	0.7195	0.7370
	P_{90}	0.6133	0.6117	0.5627	0.6033	0.6127

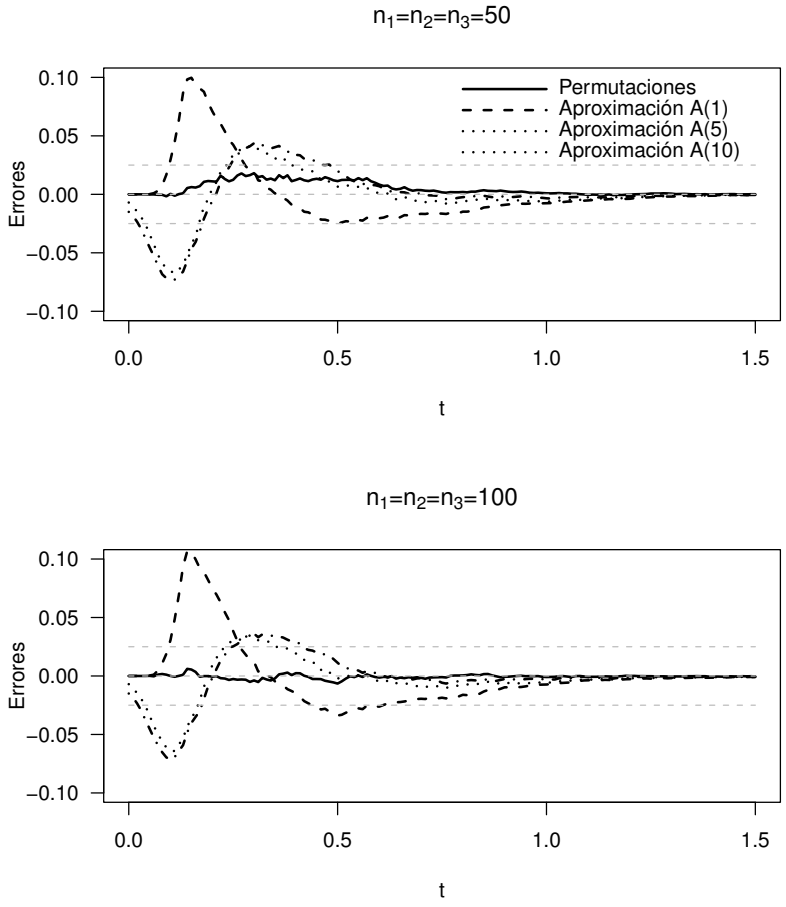
mediante el método de Monte Carlo (existen numerosos resultados sobre formas cuadráticas que también se podrían emplear; ver, por ejemplo, Alkarni and Siddiqui, 2001).

En la Figura 1 se muestran, para el problema anteriormente descrito, las diferencias entre la funciones de distribución del estadístico de Cramér-von Mises estimada por el método de Monte Carlo (10,000 iteraciones), la aproximación por permutaciones (10,000 remuestras), y la aproximación $A(m)$ (para $m = 1$, $m = 5$ y $m = 10$) cuando el tamaño muestral es 50 (arriba) y 100 (abajo). Se observa como los mayores errores son, lógicamente, para la aproximación $A(1)$ (la más simple) y están localizadas en los valores más bajos de t. Dado que, en la práctica, la zona más delicada y, por tanto, la parte que conviene aproximar bien, es la cola de la distribución y, como se puede observar en la Tabla 1, en esa parte de la curva, las diferencias son pequeñas entre todas las aproximaciones consideradas (en especial para $n = 100$). Se puede concluir que todas las aproximaciones estudiadas obtienen buenos resultados.

4 Conclusiones

Una gran parte de los artículos en los que se proponen tests clásicos (Cramér-von Mises, Kolmogorov-Smirnov, Anderson-Darling, etc...) están escritos en una terminología muy probabilística que, para aquellos investigadores que no están muy familiarizados con las técnicas usadas, hace complicada y difícil su lectura. Por otra parte, la técnicas de remuestreo actuales, hacen posible tabulaciones para los estadísticos mucho más sencillas que las utilizadas por los autores originales, lo que permite evitar ciertas cotas excesivamente complejas. En este trabajo, y haciendo nues-

Figura 1: *Diferencias entre la función de distribución real (estimada desde 10,000 réplicas de Monte Carlo) del estadístico de Cramér-von Mises y las distintas aproximaciones consideradas para* $n_1 = n_2 = n_3 = 50$ *(arriba) y* $n_1 = n_2 = n_3 = 100$ *(abajo).*



tra la cita del escritor (antaño físico) argentino Ernesto Sabato: "... *Una buena notación tiene tantas sutilezas y sugerencias que, en ocasiones, se asemeja a un maestro viviente...*", hemos retomado la generalización a k-muestras del estadístico de Cramér-von Mises propuesta por Kiefer (1956). Tratando de evitar pasos de una complejidad probabilística excesiva, pero sin denostar la valiosa carga teórica que estos resultados ofrecen, hemos usado una notación y un estilo que, pretende ser sencillo, para obtener su distribución asintótica.

En la Sección 3, mediante simulaciones de Monte Carlo, se estudia la calidad de las aproximaciones que se desprenden de la distribución asintótica y, del siempre socorrido método de las permutaciones y se comprueba como una aproximación simple ($A(1)$) obtiene buenos resultados estimando P-valores bajos (aproxima bien la cola de la distribución) necesitando utilizar aproximaciones más finas (valores de m elevados) si se desea precisión en la parte inicial de la distribución. El método propuesto, se aleja de las cálculos probabilísticos más complejos y propone técnicas que, desde la teoría clásica, aprovechan los métodos computacionales para tabular, con la precisión requerida, la distribución de los estadísticos clásicos, en particular, el de Cramér-von Mises para k-muestras independientes.

Referencias

[1] Abramowitz, M.; Stegun, I.A. (1965) *Handbook of Mathematical Integrals*. Dover, New York.

[2] Adler, R.J. (1990) *An introduction to continuity, extrema and related topics for general Gaussian processes*, IMS Lecture Notes-Monograph Series, **12**, Institute of Mathematical Statistics, Hayward, California.

[3] Alkarni, S.H.; Siddiqui, M.M. (2001) "An upper bound for the distribution function of a positive definite quadratic form", *Journal of Statistical Computation and Simulation* **69**(1): 51–56.

[4] Anderson, T.W.; Darling, D.A. (1952) "Asymptotic theory of certain 'Goodness of Fit' criteria based on stochastic processes", *Annals of Mathematical Statistics* **23**: 193–212.

[5] Anderson, T.W. (1962) "On the distribution of the two-sample Cramér-von Mises criterion", *Annals of Mathematical Statistics* **33**(3): 1148–1159.

[6] Cramér, H. (1928) "On the composition of elementary errors", *Skandinavisk Aktuarietidskrift* **11**: 141–180.

[7] Csörgő, S.; Faraway J.J. (1996) "The exact and asymptotic distributions of Cramér-von mises statistics", *Journal of the Royal Statistics Society B*, **58**(1), 1892–1903.

[8] Deheuvels, P. (2005) "Weighted multivariate Cramér-von Mises-type statistics", *Afrika Statistika* **1**(1): 1–14.

[9] Kiefer, J. (1959) "k-Samples analogues of the Kolmogorov-Smirnov, Cramér-von Mises tests", *Annals of Mathematical Statistics* **30**: 420–447.

[10] Komlòs, J.; Major, J.; Tusnády, G. (1975) "An approximation of partial sums of independent RV's, and the sample DF.I", *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **32**: 111–131.

[11] Koziol, J.A.; Green, S.B. (1976) "A Cramér-von Mises statistics for randomly censored data", *Biometrika* **63**: 465–474.

[12] Martínez-Camblor, P. (2008) "Tests de hipótesis para contrastar la igualdad entre k poblaciones", *Revista Colombiana de Estadística* **31**(1): 1–18.

[13] Martínez-Camblor, P.; Carleos, C.; Corral, N. (2011) "Powerful nonparametric statistics to compare k independent ROC curves", *Journal of Applied Statistics* **38**(7): 1317–1332.

[14] Öztürk, Ö.; Hettmansperger, T.P. (1997) "Generalised weighted Cramér-von Mises distance estimators", *Biometrika* **84**(2): 283–294.

[15] Rémillar, B.; Scaillet A.O. (2009) "Testing for equality between two copulas", *Journal of Multivariate Analysis* **100**(3): 377–386.

[16] Sabato, E. (2004) *España en los Diarios de mi Vejez*. Seix Barral, Barcelona.

[17] Schmid, F.; Trede, M. (1996) "An L_1–variant of the Cramér-von Mises test", *Statistics & Probability Letters* **26**(1): 91–96.

[18] Tomatz, L. (2002) "On the distribution of the square integral of the brownian bridge", *Annals of Probability* **30**(1): 253–269.

[19] Viollaz, A.J.; Rodríguez, J.C. (1996) "A Crámer-von Mises type goodness-of-fit test with asymmetric weight function. The Gaussian and exponential cases", *Communications in Statistics. Theory and Methods* **25**: 235–256.

[20] Van der Vaart, A.W. (1998) *Asymptotic Statistics*. Cambridge University Press, London.

[21] von Mises, R. (1931) *Wahrscheinlichkeitsrechnung*. Deuticke, Vienna.